One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques

Antonio Manuel Gómez-Orellana, David Guijo-Rubio, Jorge Pérez-Aracil, Pedro Antonio Gutiérrez, Sancho Salcedo-Sanz, César Hervás-Martínez

PII: S0169-8095(23)00005-4

DOI: https://doi.org/10.1016/j.atmosres.2023.106608

Reference: ATMOS 106608

To appear in: *Atmospheric Research*

Please cite this article as: A.M. Gómez-Orellana, D. Guijo-Rubio, J. Pérez-Aracil, et al., One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques, *Atmospheric Research* (2023), https://doi.org/10.1016/j.atmosres.2023.106608

Department of Computer Science and Numerical Analysis
Rabanales Campus.
14071 – Córdoba (Spain)

Atmospheric Research

Title: *"One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques"*

Authors: *A. M. Gómez-Orellana, D. Guijo-Rubio, J. Pérez-Aracil, P. A. Gutiérrez, S. Salcedo-Sanz, C. Hervás-Martínez*

Research Highlights

- Long-term air temperature prediction is carried out by a novel two-phased approach
- eXplainable Artificial Intelligence models are applied to better analyse the results
- The approach is applied to the Southern part of the Iberian Peninsula from ERA5 data
- A comprehensive comparison against state-of-the-art techniques has been carried out
- XAI models outperform the compared techniques and benefit from being interpretable

David Guijo Rubio

Department of Computer Science and Numerical Analysis, University of Córdoba,

e-mail address: dguijo@uco.es

One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques

Antonio Manuel Gómez-Orellana[a,1], David Guijo-Rubio[a,b,*,1], Jorge Pérez-Aracil[c], Pedro Antonio Gutiérrez[a], Sancho Salcedo-Sanz[d] and César Hervás-Martínez[a]

[a]Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, 14071, Spain
[b]School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, United Kingdom
[c]Department of Computer Systems Engineering, Universidad Politécnica de Madrid, Madrid, 28038, Spain
[d]Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Madrid, 28805, Spain

ARTICLE INFO

Keywords:
Air temperature
Long-term air temperature prediction
Climatology
XAI
Neural networks

ABSTRACT

In this paper we have tackled the problem of long-term air temperature prediction with eXplainable Artificial Intelligence (XAI) models. Specifically, we have evaluated the performance of an Artificial Neural Network (ANN) architecture with sigmoidal neurons in the hidden layer, trained by means of an evolutionary algorithm (Evolutionary ANNs, EANNs). This XAI model architecture (XAI-EANN) has been applied to the long-term air temperature prediction at different sub-regions of the South of the Iberian Peninsula. In this case, the average August air temperature has been predicted from ERA5 Reanalysis data variables, obtaining good predictions skills and explainable models in terms of the input climatological variables considered. A cluster analysis has been first carried out in terms of the average air temperature in the zone, in such a way that a number of sub-regions with different air temperature behaviour have been defined. The proposed XAI-EANN model architecture has been applied to each of the defined sub-regions, in order to find significant differences among them, which can be explained with the XAI-EANN models obtained. Finally, a comprehensive comparison against some state-of-the-art techniques has also been carried out, concluding that there are statistically significant differences in terms of accuracy in favour of the proposed XAI-EANN model, which also benefits from being an XAI model.

1. Introduction

Accurate long-term air temperature prediction is of undoubted interest (Nita et al., 2022), with different important applications in climate (Jacobs et al., 2013; You et al., 2013), energy (Bertini et al., 2010; Dombaycı and Gölcü, 2009), agriculture (Peng et al., 2017; Smith et al., 2006, 2009) or public medicine (Williams et al., 2012; Xu et al., 2014), among others. Specifically, the long-term prediction of extreme

*Corresponding author: David Guijo-Rubio. Department of Computer Science and Numerical Analysis, Universidad de Córdoba. 14071 Córdoba. Spain.
am.gomez@uco.es (A.M. Gómez-Orellana); dguijo@uco.es (D. Guijo-Rubio); jorge.perez.aracil@upm.es (J. Pérez-Aracil); pagutierrez@uco.es (P.A. Gutiérrez); sancho.salcedo@uah.es (S. Salcedo-Sanz); chervas@uco.es (C. Hervás-Martínez)
ORCID(s): 0000-0002-1929-2408 (A.M. Gómez-Orellana); 0000-0002-8035-4057 (D. Guijo-Rubio); 0000-0002-4456-9886 (J. Pérez-Aracil); 0000-0002-2657-776X (P.A. Gutiérrez); 0000-0002-4048-1676 (S. Salcedo-Sanz); 0000-0003-4564-1816 (C. Hervás-Martínez)
[1]These authors contributed equally to this work.

temperatures in summer is important, since it is related to an increase of deaths due to heat stroke, specially during heat-waves episodes (Khan et al., 2019; Díaz et al., 2002a,b). Also, one of the effects of climate change is warmer summers (Khan et al., 2019; Peña-Ortiz et al., 2015), which can be further studied by predicting average summer months air temperature at a long-term basis.

Different methods can be applied to predict long-term air temperature. In general, the methods presented in the literature for air temperature prediction can be divided into two categories: (1) weather-based, which relies on the study of physical phenomena to build a model (for instance, Numerical Weather Models, NWMs) (Kendzierski et al., 2018); and (2) time series-based, where both statistical (Ye et al., 2013; Chen and Hwang, 2000) or Artificial Intelligence/Machine Learning (AI/ML) methods are used to analyse historical temperature data series (Chevalier et al., 2011; Cifuentes et al., 2020; Salcedo-Sanz et al., 2020, 2022b). There are also hybrid approaches mixing NWMs with ML methods such as in Ortiz-García et al. (2012). NWMs have good prediction skills, specially on short-term prediction they are the state-of-the-art, without any doubt (Navascués et al., 2013). However, note that when the prediction time horizon is long-term, such as in monthly or seasonal air temperature prediction, numerical models may have much worse prediction skills, and ML approaches are meaningful, since they are able to exploit information in variables which numerical models are not able to process. Particularly, Artificial Neural Networks (ANNs) offer a highly compelling alternative in problems related to long-term prediction of air temperature, due to their ability to handle non-linear complex problems providing a robust solution and easy implementation. For example, different types of ANNs have been successfully applied to different temperature prediction scenarios and countries. In Chithra et al. (2015), ANNs 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 ANNs for long-term temperature prediction. In Ustaoglu et al. (2008), three different types of ANNs were applied to a problem of daily mean, maximum and minimum temperature time series in Turkey. In Abdel-Aal and Elhadidy (1995), different ANNs were applied to a problem of daily maximum temperature prediction in Dhahran, Saudi Arabia. Data from 18 weather parameters were considered as input variables, and the objective was to predict the maximum temperature on a given day, with different prediction time-horizons up to 3 days in advance. In De and Debnath (2009), a Multi-Layer Perceptron (MLP) is applied to the prediction of the maximum air temperature in the summer monsoon season in India. The mean temperature of previous months in the period of analysis is considered as input for the system.

Other ML approaches have also been applied to long-term prediction of air temperature. For example, in Paniagua-Tineo et al. (2011), a Support Vector Regression (SVR) algorithm was applied to a problem of daily maximum air temperature prediction, with a 24h prediction time-horizon. Input variables including previous 24h air temperature, precipitation, relative humidity, air pressure and synoptic situation were considered. Results in different European measurement stations were reported. In Mellit et al. (2013), a least squares SVR algorithm is applied to the prediction of time series temperature in Saudi Arabia. In Ahmed et al. (2020), different ML approaches are proposed to develop multi-model ensembles from global climate models. The objective is to obtain annual prediction of monsoon maximum and minimum temperature, among other variables, over Pakistan. In Peng et al. (2020), two ML algorithms (MLP and Natural Gradient Boosting, 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. In Oettli et al. (2022), a number of ML algorithms including ANNS, support vector machines, random forests, gradient boosting or regression trees are 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. Results in data from Tokyo (Japan) confirmed the good skill of the prediction. Recently, Deep Learning (DL) algorithms have also been successfully applied to air temperature prediction problems, e.g. Karevan and Suykens (2020), where a type of Long Short-Term Memory (LSTM) network (Transductive LSTM) is applied to a problem of temperature prediction in Belgium and the Netherlands.

In the last few years, eXplainable Artificial Intelligence (XAI) (Arrieta et al., 2020) has been acknowledged as a crucial feature of AI or ML algorithms, when applied to Science and Engineering problems. XAI is based on the fact that when developing a ML model to solve a given problem, the consideration of interpretability is an extremely important additional design driver, specially in some fields in which the physics of the problem plays a principal role (Tuia et al., 2021). This leads to the development of ML models in which the interpretation of the system is a primary objective, in such a way that the models provide either an understanding of the model mechanisms, or the reason why some predictions are obtained (Arrieta et al., 2020; Tuia et al., 2021). In response to this need, XAI has recently been applied to many different fields of atmospheric and climate science (Mamalakis et al., 2022b; Kolevatova et al., 2021; Mamalakis et al., 2022a; Mayer and Barnes, 2021), including problems related to long-term temperature and drought prediction (Dikshit and Pradhan, 2021; Labe and Barnes, 2022).

Following this trend, in this paper we consider a problem of long-term air temperature prediction with XAI techniques. Specifically, we consider the problem of predicting the mean summer temperature of August from reanalysis predictive variables in the previous month, by applying XAI Evolutionary Artificial Neural Networks (XAI-EANNs). We propose the use of an ANN with sigmoidal units as the basis functions in the hidden layer. The structure of the ANN is optimised by means of evolutionary computation, which facilitates its dynamic adaptation to the problem addressed and has been demonstrated to be accurate and efficient in other prediction studies. We show that this kind of XAI method is able to produce excellent prediction results, with explainable models, which can be interpreted in terms of the input variables considered. We show a case of application in the Iberian Peninsula, where there have been extremely hot summers in the last decades, related to extreme events due to climate change. A first clustering step is applied, in such a way that we divide the Southern part of the Iberian Peninsula into different sub-regions, each with special characteristics in terms of air temperature due to the orography and sea closeness. We show how the considered XAI techniques are able to accurately estimate the average temperature of August in the different sub-regions considered. A comparison with alternative ML techniques (non-XAI) is carried out to show the accuracy of the proposed method. In addition, the interpretation of the XAI models obtained in each sub-region allows a better understanding of the physical processes controlling the air temperature in the zone, which is key in alternative problems related to detection of extreme events related to temperature, and also in climate change attribution problems.

The rest of the paper is structured in the following way: next section presents the computational methods considered, including a hierarchical clustering to divide the zone under study into different sub-regions and the XAI-EANNs considered. Section 3 presents the experimental setting and evaluation metric of the paper. Section 4 shows and discusses the results obtained by applying the XAI techniques in the case study of the Iberian Peninsula. Finally, Section 5 closes the paper by giving some conclusions and remarks on the research carried out.

2. Methods: clustering analysis and XAI-EANN techniques

In this section, the hierarchical clustering applied to the zone under study prior to the prediction task is described. In addition, the architecture of the ANN model considered in this study is detailed, as well as the

evolutionary algorithm used for optimising the ANN models. The whole proposed methodology, which is detailed in the following subsections, is graphically summarised in the flowchart of Figure 1.

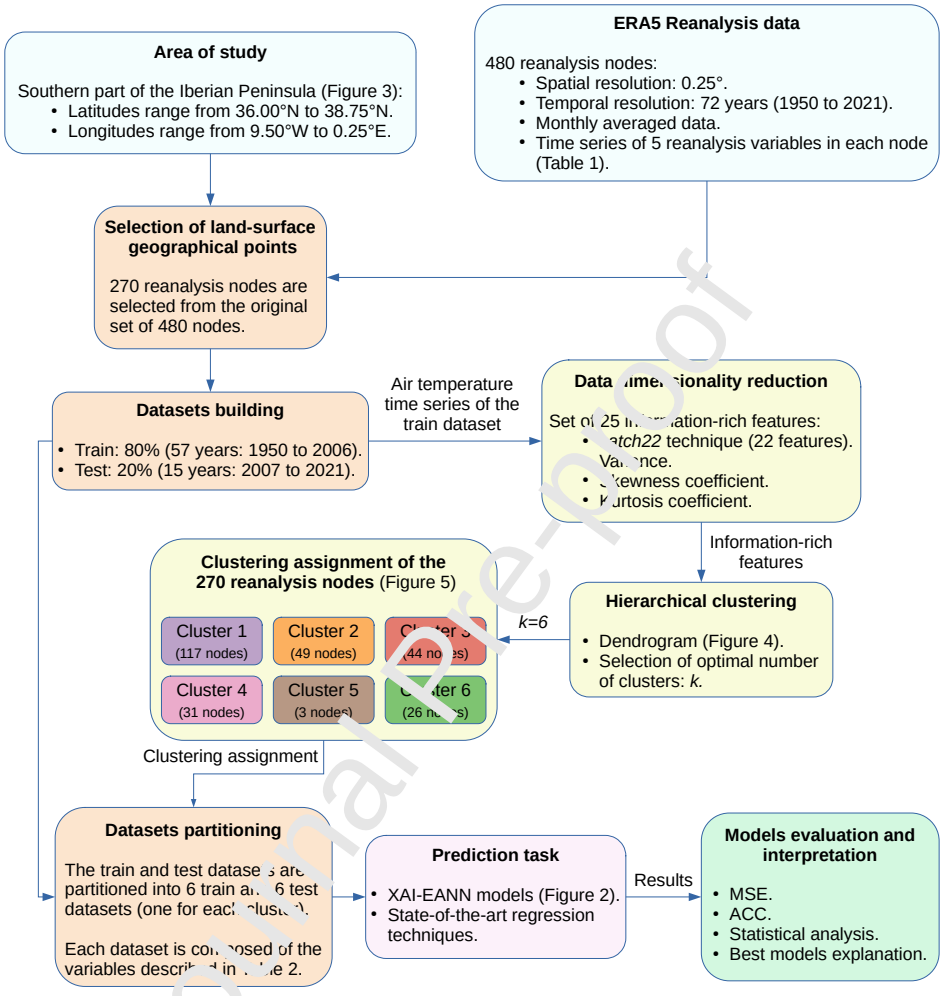


Figure 1: Flowchart of the proposed approach for one month in advance prediction of air temperature.

2.1. Hierarchical clustering

As it was aforementioned, the first step of the methodology proposed in this work is the use of a clustering technique (Saxena et al., 2017). By means of this task, similar data is organised into related or homogeneous groups without considering specific knowledge of the group definitions. Clustering is usually a preprocessing task prior to other data mining techniques, such as prediction (Guijo-Rubio et al., 2020a). Thus, in this work, a clustering analysis is performed on the area under study to identify those zones or sub-regions with similar behaviour in air temperature. Hence, obtaining robust sub-regions or clusters is of great importance, as the

subsequent task is to accurately predict the air temperature for a given cluster. It is well known that closer locations may have similar weather conditions, unless there is a significant geographical feature, such as a valley in a mountainous area, or, conversely, a mountain range in a valley area. In this way, the prediction for a specific zone will be far more accurate than using a global prediction model, which is not capable of exploiting the different behaviours that exist in the whole zone. For instance, the atmospheric conditions in coastal areas are often very different from those in inland areas. Moreover, it is common for some areas to suffer the same, or at least similar, effects that other areas have previously suffered. This relationship does not necessarily have to be due to their geographical proximity, but may be due to their similarity in weather conditions. Concretely, this work deals with one month in advance prediction of air temperature, therefore, sub-regions are identified according to the similarity in air temperature and without considering geographical localisation.

Among the most widely used clustering techniques, the partitional clustering (Kutbay et al., 2018) and the hierarchical clustering (Kaufman and Rousseeuw, 2005) stand out, mainly due to the wide variety of applications to which they have been successfully applied. Regarding the partitional clustering, the most common algorithms are k-means (Niennattrakul and Ratanamahatana, 2007) or k-medoids (Hautamaki et al., 2008; Vuori and Laaksonen, 2002). On the other hand, the hierarchical clustering is based on an agglomerative or a divisive algorithm.

As this work considers the use of XAI techniques, the hierarchical clustering based on an agglomerative algorithm is chosen over the other approaches for the following reasons: (1) as opposite to the partitional clustering, it is deterministic i.e. there is not any random initialisation nor any stochastic step; (2) results are typically represented using a dendrogram, which is a tree diagram showing the arrangement of the obtained clusters, being this a key point, as it facilitates the choice of the optimal number of clusters for the problem under consideration; (3) the simplicity of this technique is one of its strengths, which makes it very attractive to experts for analysing, using the dendrogram, the way the clusters are made up; and finally, (4) although in some cases the time complexity could be greater than desired, it does not become a drawback, as some preprocessing is carried out to extract information-rich features from the data, significantly reducing the original data.

As mentioned, these clustering techniques may be time intensive when the dimensionality of the data is huge. In this sense, instead of applying the hierarchical clustering straightforwardly to the original data, the

well-known 22 CAnonical Time-series CHaracteristics technique (catch22) (Lubba et al., 2019) is previously applied as a preprocessing step to reduce the data dimensionality. Specifically, this technique obtains a reduced set of 22 features, which are able to exhibit strong performance and being minimally redundant.

The set of features obtained as a result of applying the catch22 technique is diverse, including linear and non-linear autocorrelation, value distributions and outliers, successive differences, fluctuation analysis and other simple temporal statistics. A detailed list of all these 22 features can be found in (Lubba et al., 2019). In addition to such 22 features, 3 more are considered: the variance (measure of the dispersion of the data), the skewness coefficient (measure of the asymmetry of the probability distribution) and the kurtosis coefficient (measure of the tailedness of the probability distribution). The reason for adding these three features to the previous ones is that they also provide useful information. Therefore, the original air temperature time series are transformed into information-rich features vectors of length 25, significantly reducing the dimensionality of the original air temperature time series while avoiding the information loss.

Note that the catch22 is applied only to the training set, as it is the one used for the clustering stage, and that the test set is not considered in this clustering stage, to avoid including bias in the posterior prediction task.

After applying the catch22 technique, the hierarchical clustering procedure based on an agglomerative algorithm is carried out. This technique returns different cluster assignments depending on the number of clusters chosen. Hence, this decision is not only made according to the dendrogram, but also considering expert knowledge in the field and the objective of maximising the difference in air temperature between clusters, i. e. to minimise the intra-cluster distance and maximise the inter-cluster distance.

2.2. XAI-EANNs

Once the clustering procedure has been conducted, the second step of the proposed methodology concerns the prediction task, which is performed on each of the clusters (sub-regions) previously obtained. For this purpose, XAI-EANNs are considered, which are described next.

ANNs (Bishop, 1995) are non-linear models which use basis functions to apply non-linear transformations to input data. Given their features and flexibility, they have been extensively used in a large variety of classification and regression problems, obtaining excellent performance. Multilayer Perceptron (MLP) (Bishop, 1995) can be considered the most widely used model of feed-forward ANN, and its fully connected architecture consists of one input layer, one or more hidden layers and one output layer.

As mentioned above, this work considers the problem of predicting the mean air temperature of August one month in advance, which is defined as follows:

$$D = \left\{ \left(\mathbf{x}_i, y_i\right); i = 1, 2, \ldots, N \right\}, \tag{1}$$

where D is the dataset corresponding to the cluster being analysed, $\mathbf{x}_i$ is the vector of size d that contains the input variables, y_i is the mean air temperature of August, and N stands for the number of instances in the dataset. Given that the variables related to the air temperature of the remaining clusters are considered as inputs, the size of $\mathbf{x}_i$, i.e. d, will depend on the number of clusters resulting of the clustering analysis.

In this way, an ANN regression model is used to address the problem under study. Specifically, the input layer of the model has d neurons (one for each input variable), whereas the output layer has one linear neuron to predict the mean air temperature of August. Therefore, and without considering the hidden neurons, the output of the model is formulated as follows:

$$f(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^{m} \beta_j B_j(\mathbf{x}, \mathbf{w}_j). \tag{2}$$

where $B_j(\mathbf{x}, \mathbf{w}_j)$ represents each basis function (hidden neuron) that performs the non-linear transformations on the input vector $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_d)$, being d the size of the vector; $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_m)$ are the weights of the connections from hidden layer to output layer, including β_0 as bias; $\mathbf{w}_j^T = (w_{j0}, w_{j1}, \ldots, w_{jd})$ are the weights of the connections from input layer to hidden layer, being w_{j0} the bias; and m is the number of hidden neurons.

With respect to the hidden neurons, Sigmoidal Unit (SU) (Lippmann, 1989) is considered as the basis function. SUs represent an additive projection model and can accurately approximate any given continuous function. In fact, they have obtained excellent results when addressing regression problems (Guijo-Rubio et al., 2020b; Gómez-Orellana et al., 2022). Following the notation described in Equation (2), SU basis functions are defined as follows:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \frac{1}{1 + e^{-\left(w_{j0} + \sum_{i=1}^{d} w_{ji} x_i\right)}}, \quad j = 1, ..., m, \tag{3}$$

where w_{j0} stands for the bias.

In consequence, the architecture of the Artificial Neural Network regression model defined for this work is represented in Figure 2.

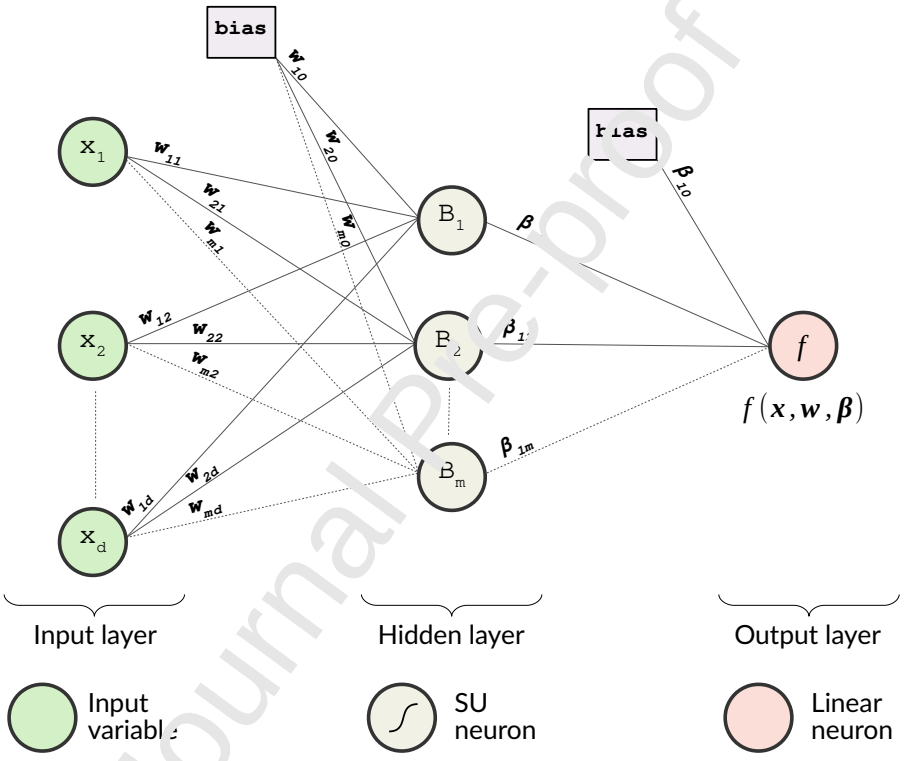


Figure 2: Architecture of the ANN regression model.

The training of ANNs is a complex task due to their convoluted error surface (Yao, 1999), with gradient-based optimisation algorithms being the most commonly ones used for this task. However, these types of algorithms are often trapped at local minima. Moreover, most of them use a fixed structure defined in advance, but it is difficult to know in advance which is the most appropriate structure to tackle a particular problem, so the optimisation of the architecture is usually based on trial and error. As alternative, Evolutionary Algorithms (EAs) represent a valuable technique to train ANNs (Ser et al., 2019), not requiring a fixed architecture. Furthermore, they conduct a global optimisation with the objective of performing a

more efficient exploration of the search space, trying to escape from local minima, which enables the EA to discover better performance solutions.

Therefore, an EA is used in this work to optimise both the structure (number of hidden neurons and connections) and the synaptic weights of the ANN regression model represented in Figure 2.

2.2.1. Evolutionary Algorithms

EAs are search and optimisation metaheuristics that mimic biological evolution to address real-world problems resolution. Specifically, they apply distinct nature-inspired mechanisms to individuals (population) to simulate the evolutionary process (i.e. search space exploration). A fitness function enables the EA to evaluate the performance of the individuals, guiding their optimisation throughout the evolution. Since the evolutionary process is based on randomly generated decisions, which are used by the mechanisms to optimise the population, the evolved individual that achieves the best fitness (final solution found) will vary according to randomness.

The pseudocode of the EA used in this study is detailed in Algorithm 1, which was presented in Martínez-Estudillo et al. (2008).

Algorithm 1 Evolutionary Algorithm

generate the individuals: initial random population of ANNs with the structure defined in Equation (2).
repeat
  compute the fitness of the individuals
  sort the individuals by their fitness
  replace the worst 10% of the population by a copy of the best 10%
  mutate the population:
  parametrically to the best 10%
  structurally to the remaining 90%
until *stopping criterion is fulfilled*
return *the best fitness individual*

Firstly, the EA begins by randomly generating the individuals (initial population of ANNs). Next, the individuals are optimised by the evolutionary process, which is iterative and involves applying the following steps: fitness calculation to sort the individuals, replacement of the worst 10% of the individuals by a copy of the best 10% of them, and mutation of the individuals (parametrically and structurally). Once the stopping criterion is fulfilled, the evolutionary process finishes and the individual that achieves the best fitness is selected and returned as the final solution found. The crossover operator is not used because it might be inefficient when optimising ANNs (Yao and Liu, 1997).

As detailed in Algorithm 1, the population of ANNs is mutated in two different ways, which are explained below. The parametric mutation modifies the synaptic weights of the connections of each individual, and it is applied by means of Gaussian noise whose variance is gradually lessened as the evolution advances. Conversely, the objective of the structural mutation is to optimise the structure of each individual (i.e. its number of both connections and hidden neurons) by means of the following sorts of mutations: add connection, delete connection, add neuron, delete neuron and neuron fusion. Thus, the structural mutation enables the EA to avoid being trapped in possible local optima and to maintain a diverse ANN population. In addition, both the add connection and the delete connection mutations (when applied to connections linking the input and hidden layers) can be considered as a kind of feature selection, favouring simpler and more explainable models.

The performance of the individuals is evaluated using the Mean Squared Error (MSE), defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2, \tag{4}$$

where $\hat{y}_i$ is the estimated value of the mean air temperature of August, in this case, it is equivalent to $f(\mathbf{x}_i, \mathbf{w}, \boldsymbol{\beta})$, which stands for the neural network output defined in Equation (2), and N is the number of instances in dataset as described in Equation (1).

Since the EA maximises the performance of the individuals throughout the evolution, the fitness function used to evaluate them is expressed as a decreasing transformation of the MSE into the range [0, 1]:

$$A = \frac{1}{1 + MSE}. \tag{5}$$

Thus, the individuals are optimised by minimising their MSE (i.e. maximising their fitness).

3. Experimental settings and evaluation

This section outlines the experimental design used for conducting the hierarchical clustering and for training and evaluating the performance of the proposed XAI-EANNs, in addition to the main state-of-the-art regression approaches considered for comparison purposes.

Table 1
Reanalysis variables from the ERA5 Reanalysis model used in this study.

Acronym	Variable description	Units
s	Volumetric soil water layer 1	$[m^3 \cdot m^{-3}]$
p	Mean sea level pressure	[Pa]
v	v-wind component	[m/s]
u	u-wind component	[m/s]
t	Air temperature	[Celsius]

3.1. Train and test datasets based on ERA5 reanalysis data

As mentioned, this study is focused on predicting the average air temperature of August in the Southern part of the Iberian Peninsula. For this, monthly averaged outputs from the ERA5 Reanalysis model (Hersbach et al., 2020) are considered, covering the period from 1950 to 2021, i.e. a total of 72 years. Concretely, 4 reanalysis variables related to the air temperature are identified: the volumetric soil water layer 1, the mean sea level pressure, and the u and v components of the wind. These reanalysis variables are selected as inputs given they present a great representation of the air temperature. In addition to these 4 reanalysis variables, the air temperature is also considered, which is the variable to be predicted. The description of the 5 reanalysis variables as well as their acronyms and measure units is detailed in Table 1.

Regarding the area of study, it lies within the latitude range from 36.00°N to 38.75°N and longitude range from 9.50°W to 0.25°E, which corresponds with the Southern part of the Iberian Peninsula. This area is represented in Figure 3 by plotting the geographical points corresponding to the location of each of the reanalysis nodes considered. As can be checked, the spatial resolution of the reanalysis data is 0.25°. Note that each reanalysis node includes the data of the 5 reanalysis variables detailed in Table 1, which describe each of geographical points of the zone under study. However, from the whole set of reanalysis nodes (or geographical points), those located on land-surface are selected, given that the goal of this work is to study the air temperature prediction on land-surface. Therefore, from the original set of 480 reanalysis nodes considered, 270 are kept.

To build the train and test datasets, the data (i.e. the time series of each of the 5 reanalysis variables included in each of the 270 reanalysis nodes) is split in such a way that the first 80% of the samples, i.e. monthly averaged data from the first 57 years, 1950 to 2006, belongs to the train dataset, whereas the remaining 20% of the samples, i.e. monthly averaged data from the last 15 years, 2007 to 2021, belongs

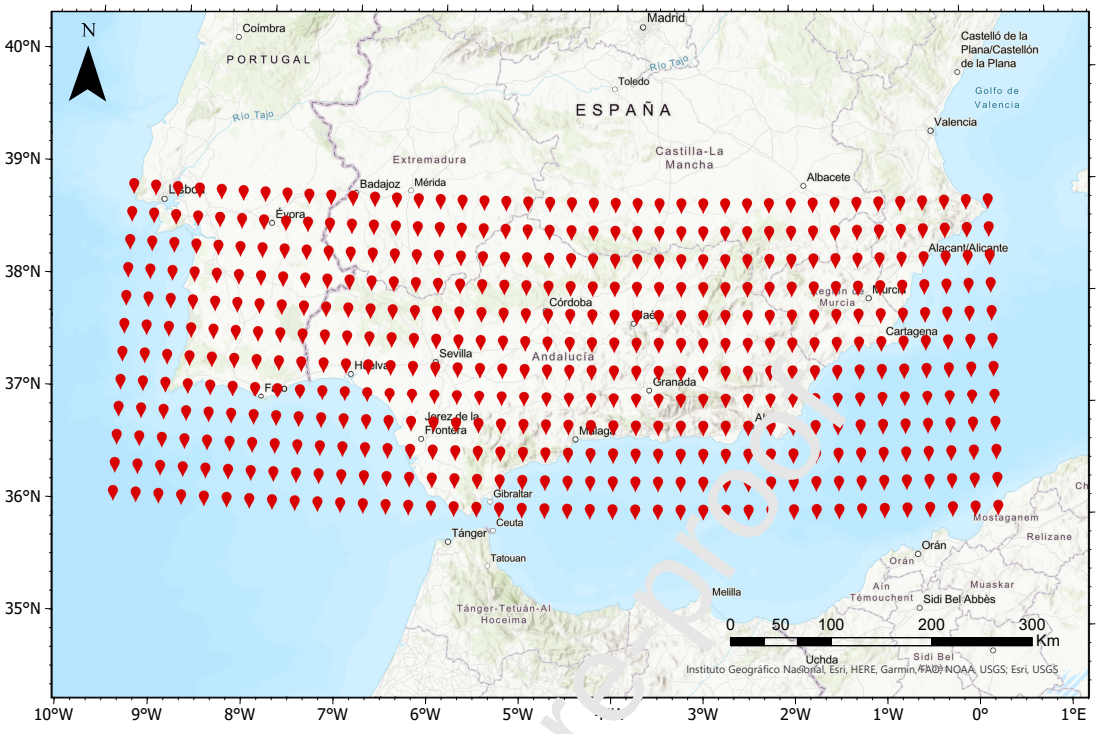


Figure 3: Geographical points considered from the Southern part of the Iberian Peninsula.

to the test dataset. Hence, given that the temporal resolution of the data is monthly, the number of patterns (monthly averaged measurements) for each of the reanalysis variables considered is 684 and 180 for the train and the test datasets, respectively.

Once the train and test datasets are built, the clustering stage is performed (as described in Section 2.1), resulting in k train and k test datasets (where k is the number of clusters), i.e. the Southern part of the Iberian Peninsula is divided into k clusters (sub-regions) to individually perform the prediction task in each of them. Specifically, each of the k train and each of the k test datasets includes the mean temperature in August (the variable to be predicted) and three AutoRegressive parts, AR(p), which are described next: the first one considers the 5 variables included in Table 1 with order $p = 1$ (i.e. the previous month of the same year, July); the second one considers the air temperature with order $p = 12$ (i.e. the air temperature in August of last year); and the third one considers the air temperature of the centroids of the remaining clusters (not that of itself) with order $p = 1$ (i.e. the air temperature of the previous month of the same year, July), hence, $k - 1$ variables are obtained from this third AR model, being k the number of clusters. Therefore, to perform

Table 2
Variables description for the prediction task.

	AR model	Acronym	Variable description	Units
Inputs	First	t	Air temperature in July	[Celsius]
		s	Volumetric soil water layer 1 in July	[$m^3 \cdot m^{-3}$]
		p	Mean sea level pressure in July	[Pa]
		v	v-wind component in July	[m/s]
		u	u-wind component in July	[m/s]
	Second	$t_{a_{-1}}$	Air temperature in August of last year	[Celsius]
	Third	t_{c_1}	Cluster's 1 centroid air temperature in July	[Celsius]
		t_{c_2}	Cluster's 2 centroid air temperature in July	[Celsius]
		t_{c_3}	Cluster's 3 centroid air temperature in July	[Celsius]
		$t_{c_{...}}$	Cluster's $\cdots$ centroid air temperature in July	[Celsius]
		t_{c_k}	Cluster's k centroid air temperature in July	[Celsius]
Output	-	t_a	Air temperature in August	[Celsius]

the prediction task, each of the k train and each of the k test datasets is composed of a total of $6 + k - 1$ input variables, and the air temperature in August as the output. Table 2 summarises the description of the variables considered for the prediction task. In this sense, we have chosen some of the most representative predictive variables for long-term temperature prediction (obtained from ERA5 Reanalysis) previously used in different related works (Cifuentes et al., 2020), plus those describing the temperature of the clusters. Note that the centroid is the most representative point of the cluster, and it is chosen as the closest geographical point to the virtual centre of the cluster (the mean of all the points).

3.2. Model training

For the optimisation of the XAI-EANNs proposed in this work, Table 3 shows the considered values for the most important parameters of the EA.

In order to provide a fair comparison, besides XAI-EANNs, the following state-of-the-art techniques are also considered: Linear Regression (LinearReg) (Bishop, 2006), Ridge Regression (RidgeReg) (Bishop, 2006), Lasso Regression (LassoReg) (Friedman et al., 2010), ElasticNet (Zou and Hastie, 2005), Support Vector Regressor (SVR) (Vapnik, 2013), MultiLayer Perceptron (MLP) (Bishop, 1995), Random Forest (Breiman, 2001) and eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016). As can be seen, a variety of techniques have been selected, including simple linear regressors and complex ML techniques.

Table 3
Considered values for the most important parameters of the Evolutionary Algorithm.

Parameter description	Value
Maximum number of generations (stopping criterion)	1000
Population size of ANNs	1000
Number of hidden neurons (population initialisation)	$[1, 3]$
Maximum number of hidden neurons (evolutionary process)	4
Weights for connections between input and hidden layers (initialisation & mutation)	$[-1, 1]$
Weights for connections between hidden and output layers (initialisation & mutation)	$[-5, 5]$
Number of hidden neurons to add or delete (mutation)	$[1, 4]$
Number of connections to add or delete (mutation)	$[1, 7]$
Input variables scale	$[0.1, 0.9]$
Output variable scale	$[-10, 10]$

Table 4
Considered range of values for tuning the parameters of the state-of-the-art techniques.

Technique	Parameter description	Range of values
RidgeReg	Regularisation	$\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$
LassoReg	Regularisation	$\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$
	Maximum number of iterations	$[1000, 2000, 3000, 4000, 5000]$
ElasticNet	Regularisation	$\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$
	Maximum number of iterations	$[1000, 2000, 3000, 4000, 5000]$
	Ratio of the L1 penalisation weight	$[0.10, 0.50, 0.70, 0.90, 0.95, 0.99, 1.00]$
SVR	Kernel width	$\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$
MLP	Number of hidden neurons	$[10, 25, 50, 100]$
	Regularisation	$\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$
	Initial learning rate	$[10^{-5}, 10^{-4}, 10^{-3}]$
	Number of iterations	$[1000, 1500]$
Random Forest	Number of trees in the forest	$[100, 500, 1000]$
	Maximum depth of the trees	$[3, 4, 5, 6]$
	Maximum depth of the trees: when all leaves are pure or contain less than samples than the minimum	$[2, 5]$
XGBoost	Number of gradient boosted trees	$[100, 500, 1000]$
	Maximum depth of the trees	$[4, 6, 8]$
	Learning rate	$[0.1, 0.3, 0.5, 0.7]$
	Subsample ratio of the training instances for the tree booster	$[0.5, 0.75, 1.00]$

The first five techniques are deterministic, thus, they are run once, whereas the remaining three techniques are non-deterministic, and, hence, they are run 40 times.

The parameter values of these techniques are tuned using a 10-fold cross-validation scheme over the training set, selecting the best value for each parameter (the one that achieves the lowest MSE) from the range of values shown in Table 4. Note that the LinearReg technique has no parameters to tune.

In addition to these state-of-the-art techniques, two models are also considered, namely Persistence (Salcedo-Sanz et al., 2022a) and Climatology. Both of them are highly used as baseline methods, as they are simple models achieving, generally, good solutions. The Persistence model predicts the current value to be equal to the previous observed value. On the other hand, the Climatology model predicts the current value to be equal to the average of all the previous observed values available.

3.3. Model evaluation

Two different metrics are used to analyse the performance of the models, the Anomaly Correlation Coefficient (ACC) and the Mean Squared Error (MSE, see Equation (4)). The ACC measures the correlation between anomalies of predictions and observed values with the reference values. It is defined as follows:

$$ACC = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - c_i\right)\cdot\left(y_i - c_i\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - c_i\right)^2 \cdot \frac{1}{N}\sum_{i=1}^{N}\left(y_i - c_i\right)^2}}, \quad (-1 \leq ACC \leq 1), \tag{6}$$

where c_i is the value of the reference model.

Specifically, two reference models have been considered, as they represent different objectives: the Persistence and the Climatology. Therefore, the ACC results are calculated separately depending on the reference model used. High correlations between forecasts and observations may be due to seasonal variations. Hence, the ACC is used to remove the reference model predictions from both forecasts and observations, verifying in this way the anomalies. Note that high ACC values mean that the evaluated model can better predict the anomalies with respect to the reference model prediction.

As aforementioned, XAI-EANN, MLP, Random Forest and XGBoost are stochastic techniques. Hence, their performance is also assessed in terms of stability, computing the Standard Deviation (SD) of the MSE results obtained in the 40 runs performed.

4. Results

This section firstly introduces the result of the hierarchical clustering procedure. Secondly, regarding the prediction task and from a quantitative perspective, the performance of the proposed XAI-EANNs is reported and compared to the considered state-of-the-art techniques. Then a statistical analysis of the performance

of the different techniques is presented. Finally, the best model obtained in each cluster is explained and discussed from a quantitative and qualitative point of view, respectively.

4.1. Clustering

The result of the hierarchical clustering is shown in the dendrogram represented in Figure 4, which illustrates the arrangement of the clusters obtained using a tree diagram. The vertical axis represents the distance inter-cluster (the greater the distance the lesser the similarity) and the horizontal axis reports the cluster assignment to each geographical point analysed.

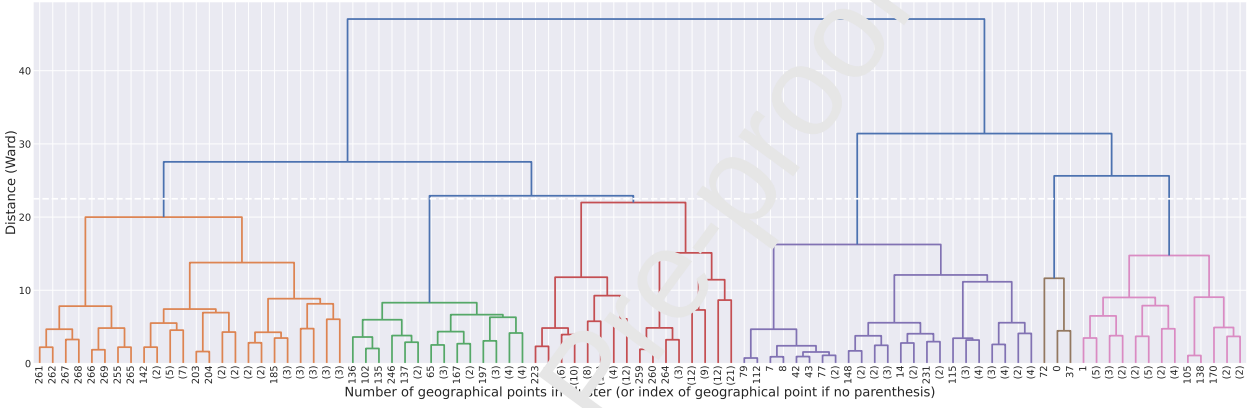


Figure 4: Dendrogram corresponding to the hierarchical clustering.

As can be seen, and considering expert knowledge, the selected optimal number of clusters is six, which are plotted in distinct colour. Hence, the region under study (composed of 270 geographical points) is divided into six sub-regions (clusters) according to their similarity in mean air temperature. This division is carried out prior to the prediction task, which is performed separately in each sub-region. In this way, the geographical representation of the six sub-regions obtained as a result of the clustering assignment is shown in Figure 5.

As can be checked, the clusters are clearly differentiated, that is, the points belonging to the same cluster are geographically located in the same area, and not dispersed. This makes sense, since nearby locations are expected to have similar air temperature behaviour. However, the points located at latitude 37.00°N and longitude 5.75°W, and at latitude 37.50°N and longitude 3.00°W (belonging to the clusters 3 and 6, respectively) are somewhat distant from the remaining points of their corresponding cluster. This circumstance could be due to the influence exerted on the air temperature by the orographic characteristics

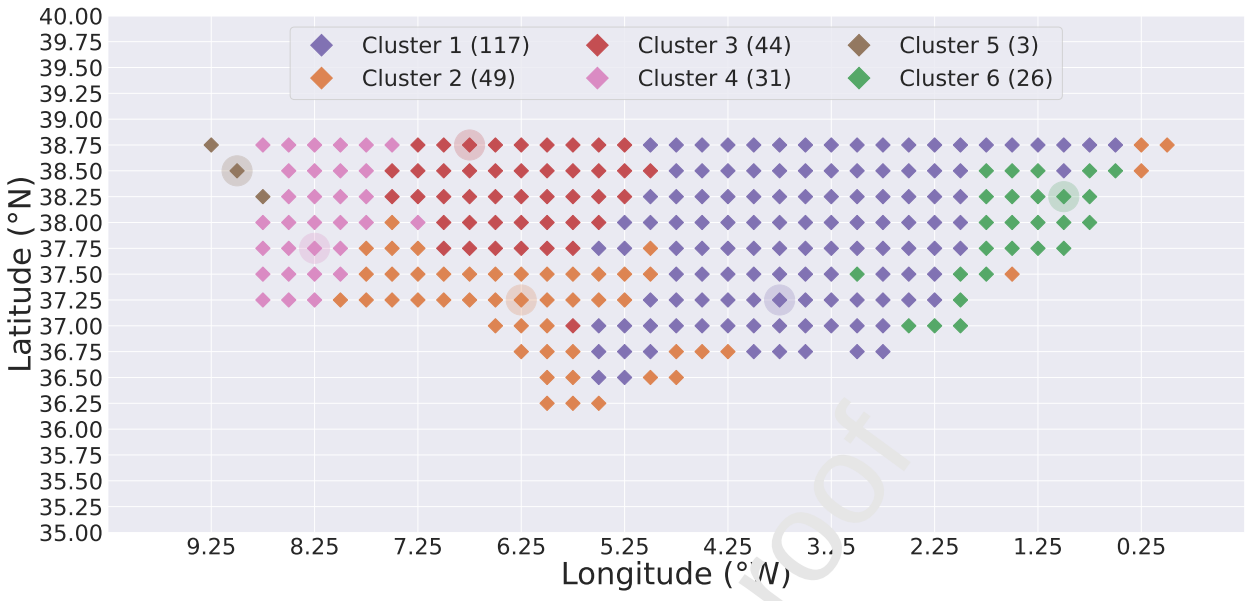


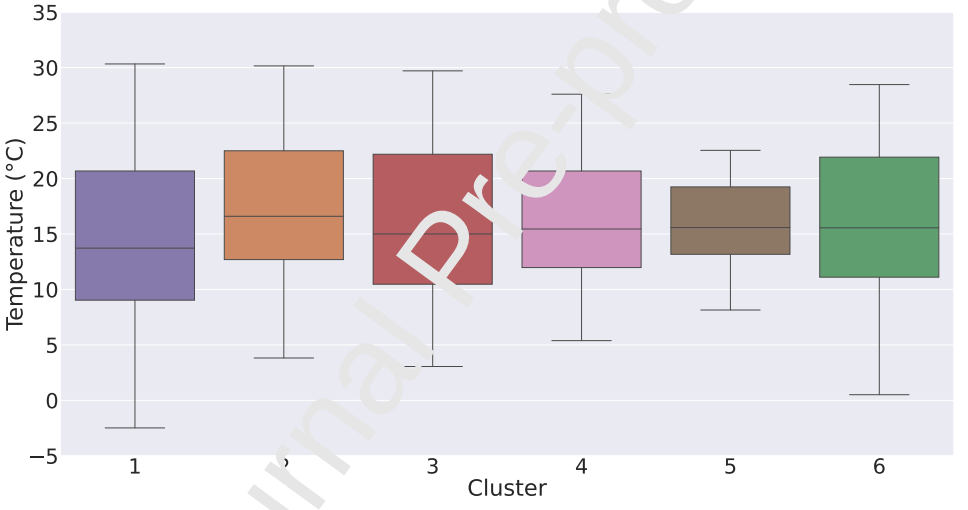
Figure 5: Geographical representation of the six clusters (sub-regions). The centroid of each cluster is represented by the point circularly shaded.

of both points compared to other nearby points. Something similar occurs in cluster 2, whose points are mostly concentrated in longitudes 8.00°W to 4.25°W, although some of them are located in longitudes 1.50°W, 0.25°W and 0.00°W. Therefore, both remarks also make sense since, and as mentioned above, the geographical location of each point (i.e. latitude and longitude) has not been used in the clustering procedure, given that the purpose of the clustering was to group the points according to their air temperature behaviour, not their location. Finally, it is worth of mentioning that cluster 5 only has three points, which highlights the particular air temperature conditions of these points in comparison to the remaining ones. Note that the points circularly shaded represent the centroid of each cluster, that is, the most representative point in terms of air temperature. As mentioned, the centroid is the closest geographical point to the virtual centre of all the points belonging to the cluster.

Further visual analysis to evaluate the result of the hierarchical clustering is done in Figure 6, where the box-plots corresponding to the distributions of the mean air temperature are represented considering the centroid of each cluster (upper part), or all the points of each cluster (lower part).

As can be observed, the mean air temperature distribution of the centroid of each cluster (upper part) is practically the same as that of all the points of the cluster (lower part), as expected since the centroid is the

(a) Box-plot of the centroid of each cluster.



(b) Box-plot of all the points of each cluster.

Figure 6: Box-plots corresponding to the mean air temperature considering: (a) the centroid of each cluster, (b) all the points of each cluster.

most representative point of the cluster. Hence, the differences between the six clusters are discussed next focusing on the box-plots that consider all the points of each cluster.

Analysing the lower part of Figure 6, it can be appreciated that cluster 5 shows the lowest variability in temperature (followed by cluster 4), since its mean temperature ranges approximately between 13 °C and 19 °C, in addition to having the lowest maximum mean temperature value and the highest minimum mean temperature value. The most probable reason for this small variability of temperature in cluster 5 is that this cluster detects a subregion near Lisbon area, where the temperature variation is further controlled by

the closeness to the sea and the Tagus river mouth, which makes that area significantly different from the rest of clusters. On the contrary, cluster 2 represents the hottest region, which has the highest median value (approximately 17 °C), whereas its maximum mean temperature value exceeds 30 °C and its minimum value is close to 4 °C. Although clusters 3 and 6 seem to be similar, some differences are observed. For example, the mean temperature value of quartile 1 is higher in cluster 6 than in cluster 3. In addition, cluster 6 has the minimum mean temperature value of both clusters, while cluster 3 reaches the maximum value. Regarding cluster 1, it is noteworthy that it is the only cluster that reaches a minimum mean temperature value below 0 °C, as well as having the lowest median value and being the only cluster with the quartile 1 value below 10 °C. Hence, this cluster groups the coolest geographical points compared to the other sub-regions.

Therefore, it can be said that the region under study has been divided into six sub-regions, which in terms of air temperature behaviour are heterogeneous among themselves and, at the same time, homogeneous considering the geographical points grouped in each of them. That is, each sub-region represents a zone with a similar air temperature behaviour, but different from the remaining sub-regions.

4.2. Models performance

To accurately predict the air temperature in a time-horizon of one month, 6 models are built, one for each cluster. Thus, each model is specifically zone-trained, as clusters represent geographical zones, in such a way that local information could be exploited as much as possible. For the sake of clarity, the stochastic techniques are firstly compared, expressing their results as the mean and Standard Deviation (SD) of the 40 runs carried out: $Mean_{SD}$. Specifically, the results report the MSE and the ACC (the latter computed twice, using the Persistence and the Climatology as reference models, respectively) achieved for the different clusters considering all the points belonging to each cluster, and whose amount of points differs from one cluster to another as shown in Figure 5. In this way, Table 5 shows the results achieved by the stochastic techniques: XAI-EANNs, MLP, Random Forest and XGBoost. In terms of MSE, XAI-EANNs achieve excellent results: best technique for clusters 1, 2 and 6, and second best for clusters 3 and 4, i.e. the proposed XAI-EANNs achieve outstanding average results except for cluster 5 (the one with the fewest points), whose performance is very close to the second best result. In addition, the small SD values achieved for the XAI-EANN models demonstrate the stability and robustness of the proposed methodology. On the other hand, XGBoost also achieves competitive results, being the best for clusters 3, 4 and 5 and the second best for clusters 1 and 2. The ACC using the Persistence as reference model indicates that XAI-EANNs achieve

Table 5
Comparison between the stochastic techniques applied in terms of MSE and ACC (with Persistence and Climatology) for the different clusters. The results are expressed as their mean and Standard Deviation (SD): $Mean_{SD}$

Technique applied	MSE					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
XAI-EANNs	$\mathbf{0.694_{0.056}}$	$\mathbf{0.908_{0.054}}$	$\mathit{1.428_{0.098}}$	$\mathit{1.405_{0.117}}$	$0.475_{0.051}$	$\mathbf{0.495_{0.065}}$
MLP	$0.963_{0.140}$	$1.075_{0.205}$	$2.358_{0.651}$	$1.637_{0.288}$	$0.729_{0.116}$	$\mathit{0.672_{0.106}}$
Random Forest	$0.978_{0.005}$	$1.010_{0.005}$	$2.235_{0.021}$	$1.512_{0.014}$	$\mathit{0.470_{0.009}}$	$0.675_{0.011}$
XGBoost	$\mathit{0.797_{0.038}}$	$\mathit{0.949_{0.042}}$	$\mathbf{1.369_{0.097}}$	$\mathbf{1.267_{0.074}}$	$\mathbf{0.422_{0.023}}$	$0.733_{0.045}$
Technique applied	ACC with Persistence					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
XAI-EANNs	$\mathbf{0.689_{0.013}}$	$\mathbf{0.704_{0.013}}$	$\mathbf{0.618_{0.019}}$	$0.719_{0.020}$	$0.768_{0.023}$	$\mathbf{0.729_{0.037}}$
MLP	$\mathit{0.682_{0.025}}$	$\mathit{0.696_{0.042}}$	$0.519_{0.092}$	$\mathbf{0.743_{0.044}}$	$\mathit{0.774_{0.024}}$	$0.643_{0.040}$
Random Forest	$0.618_{0.001}$	$0.691_{0.001}$	$0.408_{0.004}$	$0.710_{0.002}$	$0.765_{0.005}$	$\mathit{0.660_{0.005}}$
XGBoost	$0.634_{0.013}$	$\mathbf{0.704_{0.009}}$	$\mathit{0.577_{0.030}}$	$\mathit{0.735_{0.017}}$	$\mathbf{0.794_{0.013}}$	$0.626_{0.022}$
Technique applied	ACC with Climatology					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
XAI-EANNs	$\mathbf{0.906_{0.008}}$	$\mathbf{0.711_{0.020}}$	$\mathit{0.509_{0.034}}$	$0.566_{0.029}$	$0.550_{0.031}$	$\mathbf{0.854_{0.022}}$
MLP	$\mathit{0.892_{0.011}}$	$0.663_{0.071}$	$0.238_{0.190}$	$0.561_{0.051}$	$\mathbf{0.593_{0.018}}$	$\mathit{0.850_{0.011}}$
Random Forest	$0.865_{0.001}$	$0.680_{0.002}$	$0.393_{0.004}$	$\mathbf{0.618_{0.002}}$	$0.543_{0.008}$	$0.821_{0.002}$
XGBoost	$0.891_{0.005}$	$\mathit{0.698_{0.014}}$	$\mathbf{0.502_{0.027}}$	$\mathit{0.614_{0.024}}$	$\mathit{0.568_{0.027}}$	$0.810_{0.011}$

The best results are highlighted in **bold**, whereas the second best are in *italics*.

the best results in 4 out of the 6 clusters, whereas the other two best results are obtained by MLP in the case of cluster 4 and by XGBoost in the case of cluster 5. As can be observed, the ACC results computed using the Persistence as reference model are always over 0.6 in the case of XAI-EANNs, which means that a high correlation is achieved between the observed and predicted values when the information provided by the reference model is subtracted. On the other hand, the results achieved by ACC using Climatology are similar to those achieved by MSE in what respects to XAI-EANNs: best results in 3 out of the 6 clusters. Nevertheless, XGBoost, despite being a good technique, only gets the best results in 1 out of the 6 clusters. In this case, it is worth mentioning that results achieved by XAI-EANNs for clusters 1, 2, and 6 are over 0.7.

Next, with the goal of identifying one best model for each cluster, the deterministic techniques are compared against the best model obtained by each of the stochastic techniques (i.e. the one that achieves the lowest MSE of the 40 runs). As well as for the previous Table 5, results are reported considering all the points belonging to each cluster. In this way, Table 6 shows the results achieved by the deterministic techniques: LinearReg, SVR and ElasticNet, and the best XAI-EANN, MLP, Random Forest and XGBoost models.

Furthermore, for comparison purposes, the results achieved in terms of MSE by the two aforementioned baseline models, the Persistence and the Climatology, are also included in this Table 6. Computing the corresponding ACC results for these two baseline models is meaningless, since each one is used as the reference model in their own ACC computation. Concerning the MSE results, it can be observed that the proposed XAI-EANNs achieve the best results as they obtain excellent results in 5 out of the 6 clusters (for 3 of them is the best technique and the second best in the other 2 groups). The second best technique is XGBoost as it obtains the best results in 3 out of the 6 clusters. MLP and SVR achieve the second best results in 3 and 1 out of the 6 clusters, respectively. Finally, Persistence and Climatology are always outperformed by the ML techniques, with a few exceptions. A similar behaviour can be observed regarding the ACC results using both Persistence and Climatology as reference models. Except for Cluster 5, XAI-EANNs achieve the best results in 5 out of the 6 clusters (3 being the first and 2 being the second one) for both reference models. It is worth mentioning that XGBoost and MLP are the second and third best techniques, respectively. Moreover, when it comes to clusters with a high number of points, the proposed XAI-EANNs achieve the best results (see those for Clusters 1 and 2, with 117 and 49 points, respectively).

Another interesting analysis consists in evaluating the performance of the models when predicting the air temperature at a specific point, instead of considering all the points belonging to the cluster. For this purpose, the performance of the same models included in Table 6 are now compared in Table 7 reporting the results considering only the most representative point of each cluster, i.e. the centroid. From this Table 7, and concerning the MSE results, it can be observed that the proposed XAI-EANNs achieve the best results in 5 out of the 6 clusters, obtaining excellent results in comparison with the rest of the techniques: SVR achieves the best result for the remaining cluster, whereas MLP, XGBoost and RidgeReg obtain the 3, 2 and 1 second best results, respectively. A similar performance can be checked regarding the ACC results computed with the Climatology as reference model, except that the second best results are divided between MLP (4), Random Forest (1), and XGBoost shared with linear methods (1). Nevertheless, for the ACC computed using the Persistence, despite XAI-EANNs achieving excellent results, the number of best results is reduced to 3, and achieving one second best result. The remaining three best results are for XGBoost, MLP, and SVR.

Additionally, to give a visual comparison of the results shown in Table 7, the air temperature predictions made by the best and second best techniques for the centroid of each cluster are represented along with the observed values in Figure 7, for clusters 1 to 3, and in Figure 8, for clusters 4 to 6. The technique achieving the

Table 6
Comparison between the deterministic techniques and the best model obtained by each stochastic technique applied in terms of MSE and ACC (with Persistence and Climatology) for the different clusters.

Technique applied	MSE					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Persistence	0.997	1.411	1.853	2.696	1.120	1.028
Climatology	3.865	1.821	2.008	1.940	0.602	1.799
LinearReg	0.782	0.892	1.592	1.657	0.721	0.667
RidgeReg	0.776	0.892	1.592	1.642	0.715	0.667
LassoReg	0.793	0.911	1.552	1.573	0.645	0.661
ElasticNet	0.793	0.908	1.552	1.573	0.643	0.661
SVR	1.645	1.847	1.347	1.692	*0.381*	1.570
XAI-EANNs	**0.581**	**0.753**	*1.229*	*1.146*	0.389	**0.383**
MLP	*0.704*	*0.775*	1.374	1.200	0.579	*0.481*
Random Forest	0.967	0.997	2.197	1.481	0.452	0.653
XGBoost	0.735	0.848	**1.176**	**1.060**	**0.358**	0.634
Technique applied	ACC with Persistence					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
LinearReg	0.679	0.723	0.621	0.698	0.703	0.632
RidgeReg	0.679	0.723	0.621	0.700	0.705	0.632
LassoReg	0.676	0.717	0.626	0.706	0.724	0.634
ElasticNet	0.676	0.717	0.626	0.706	0.721	0.634
SVR	0.524	0.540	0.621	0.660	*0.813*	0.361
XAI-EANNs	**0.719**	*0.752*	**0.670**	*0.767*	0.808	**0.797**
MLP	*0.702*	**0.744**	0.600	0.749	0.789	*0.732*
Random Forest	0.620	0.693	0.414	0.715	0.774	0.669
XGBoost	0.659	0.725	*0.632*	**0.781**	**0.826**	0.678
Technique applied	ACC with Climatology					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
LinearReg	0.894	0.720	0.562	0.548	0.542	0.812
RidgeReg	0.893	0.720	0.562	0.549	0.542	0.812
LassoReg	0.892	0.712	0.567	0.554	0.548	0.811
ElasticNet	0.892	0.713	0.567	0.554	0.545	0.811
SVR	0.759	0.345	0.574	0.387	0.634	0.426
XAI-EANNs	**0.922**	**0.766**	*0.639*	*0.645*	0.595	**0.892**
MLP	*0.906*	*0.760*	0.596	0.629	*0.597*	*0.866*
Random Forest	0.867	0.684	0.397	0.623	0.548	0.822
XGBoost	0.900	0.732	**0.655**	**0.683**	**0.647**	0.836

The best results are highlighted in **bold**, whereas the second best are in *italics*.

best results is denoted by means of the superscript [1], whereas the second best is denoted using the superscript [2].

From these Figures 7 and 8, it can be seen that the predictions made are highly accurate, as most of the predictions made by the best technique (generally, XAI-EANNs) are close to the observed values. Also, it is worthy of mention that predicting the extreme peaks above-average of the time series is a very difficult

Table 7
Comparison between the deterministic techniques and the best model obtained by each stochastic technique applied in terms of MSE and ACC (with Persistence and Climatology) for the centroid of each cluster.

Technique applied	Centroid of cluster 1	Centroid of cluster 2	Centroid of cluster 3	Centroid of cluster 4	Centroid of cluster 5	Centroid of cluster 6
			MSE			
Persistence	0.517	1.207	2.964	2.702	1.375	0.919
Climatology	4.225	1.971	2.099	1.879	0.659	1.802
LinearReg	0.460	0.999	1.808	1.654	0.689	0.624
RidgeReg	*0.455*	0.999	1.808	1.642	0.681	0.624
LassoReg	0.466	1.038	1.753	1.574	0.609	0.620
ElasticNet	0.466	1.031	1.753	1.574	0.596	0.620
SVR	0.887	2.753	1.549	1.312	**0.357**	1.682
XAI-EANNs	**0.303**	**0.829**	**1.259**	**1.127**	0.425	**0.317**
MLP	0.510	*0.871*	*1.343*	1.250	0.428	*0.445*
Random Forest	0.786	1.373	2.984	1.637	0.464	0.539
XGBoost	0.476	1.076	1.366	*1.136*	*0.399*	0.599
			ACC with Persistence			
Technique applied	Centroid of cluster 1	Centroid of cluster 2	Centroid of cluster 3	Centroid of cluster 4	Centroid of cluster 5	Centroid of cluster 6
LinearReg	*0.750*	*0.655*	0.659	0.679	0.720	0.598
RidgeReg	*0.750*	*0.655*	0.659	0.680	0.722	0.598
LassoReg	0.749	0.647	0.666	0.686	0.742	0.599
ElasticNet	0.749	0.647	0.666	0.686	0.743	0.600
SVR	0.517	0.463	0.676	0.688	**0.837**	0.255
XAI-EANNs	**0.766**	0.647	**0.739**	*0.752*	0.792	**0.814**
MLP	0.733	0.654	*0.723*	**0.764**	**0.837**	*0.725*
Random Forest	0.594	0.605	0.394	0.690	0.770	0.677
XGBoost	0.627	**0.670**	0.710	*0.752*	*0.807*	0.636
			ACC with Climatology			
Technique applied	Centroid of cluster 1	Centroid of cluster 2	Centroid of cluster 3	Centroid of cluster 4	Centroid of cluster 5	Centroid of cluster 6
LinearReg	*0.982*	0.675	0.460	0.441	0.568	0.802
RidgeReg	*0.982*	0.675	0.460	0.443	0.569	0.802
LassoReg	*0.982*	0.659	0.471	0.447	0.575	0.801
ElasticNet	*0.982*	0.662	0.471	0.447	0.572	0.801
SVR	0.970	−0.091	0.513	0.406	**0.670**	0.316
XAI-EANNs	**0.988**	**0.749**	**0.618**	**0.548**	0.604	**0.901**
MLP	0.980	*0.725*	*0.596*	0.461	*0.646*	*0.857*
Random Forest	0.971	0.555	0.137	*0.547*	0.544	0.839
XGBoost	*0.982*	0.646	0.583	0.552	0.616	0.833

The best results are highlighted in **bold**, whereas the second best are in *italics*.

task, as they may be caused due to heat waves that have occurred during a short period of the month of August. Furthermore, as it can be observed, the predictions are smoother than the observed values given

Table 8
Comparison between the stochastic techniques applied in terms of the number of connections for the different clusters. The results are expressed as their mean and Standard Deviation (SD): $Mean_{SD}$

Technique applied	# Connections					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
XAI-EANNs	$\mathbf{33.2_{3.8}}$	$\mathbf{32.4_{3.3}}$	$\mathbf{23.9_{3.9}}$	$\mathbf{26.1_{4.9}}$	$\mathbf{21.5_{6.8}}$	$\mathbf{29.1_{4.1}}$
MLP	$\mathit{915.0_{193.3}}$	$\mathit{981.0_{226.3}}$	$\mathit{949.3_{206.6}}$	$\mathit{993.7_{192.5}}$	$\mathit{1000.7_{175.5}}$	$\mathit{976.9_{221.9}}$
Random Forest	$7.2 \cdot 10^6_{1.8 \cdot 10^6}$	$2.9 \cdot 10^6_{8.0 \cdot 10^5}$	$2.3 \cdot 10^6_{8.4 \cdot 10^5}$	$1.9 \cdot 10^6_{5.3 \cdot 10^5}$	$1.1 \cdot 10^5_{7.8 \cdot 10^4}$	$1.4 \cdot 10^6_{4.9 \cdot 10^5}$
XGBoost	$1.2 \cdot 10^5_{6.6 \cdot 10^4}$	$1.0 \cdot 10^5_{3.3 \cdot 10^4}$	$2.8 \cdot 10^4_{1.0 \cdot 10^2}$	$3.7 \cdot 10^4_{2.0 \cdot 10^4}$	$9.0 \cdot 10^3_{1.2 \cdot 10^3}$	$2.8 \cdot 10^4_{1.9 \cdot 10^2}$

The best results are highlighted in **bold**, whereas the second best are in *italics*.

that the models have been trained using all the points belonging to each cluster, i.e. models have not been specifically trained for each point of the cluster.

Finally, the last part of the analysis of the results is focused on the complexity of the models. Note that complexity does not refer to computational cost, but to how explainable the models are, i.e. the fewer number of connections, the more simple and explainable the model is. In this sense, Table 8 shows the $Mean_{SD}$ number of connections of the models obtained in the 40 runs carried out for each stochastic technique in each of the different clusters. As can be seen, XAI-EANNs are the simplest stochastic models for all the clusters (the number of connections ranges from 21.5 for cluster 5 to 33.2 for cluster 1), whereas Random Forest is the most complex technique. Besides, Table 9 shows the number of connections of the deterministic models and the best model obtained by each stochastic technique applied (i.e. the one that achieved the lowest MSE of the 40 runs). In this case, linear models have no competitors as they always obtain the simplest models. Nevertheless, beyond linear models, the simplest technique is XAI-EANNs, which only double the number of connections of the linear models, or, in the worst case, triple the number of connections of the linear models. This indicates that even XAI-EANNs are considered a complex ML technique, they can achieve outstanding results without considerably increasing their complexity, that is, the proposed XAI-EANNs allow for developing competitive performance and explainable models.

4.3. Statistical analysis

In this section, a statistical analysis is carried out to derive some conclusions about the results achieved by the developed models. More specifically, the intention is to determine whether there are significant differences between such results. For this purpose, the MSE results obtained by the stochastic techniques for

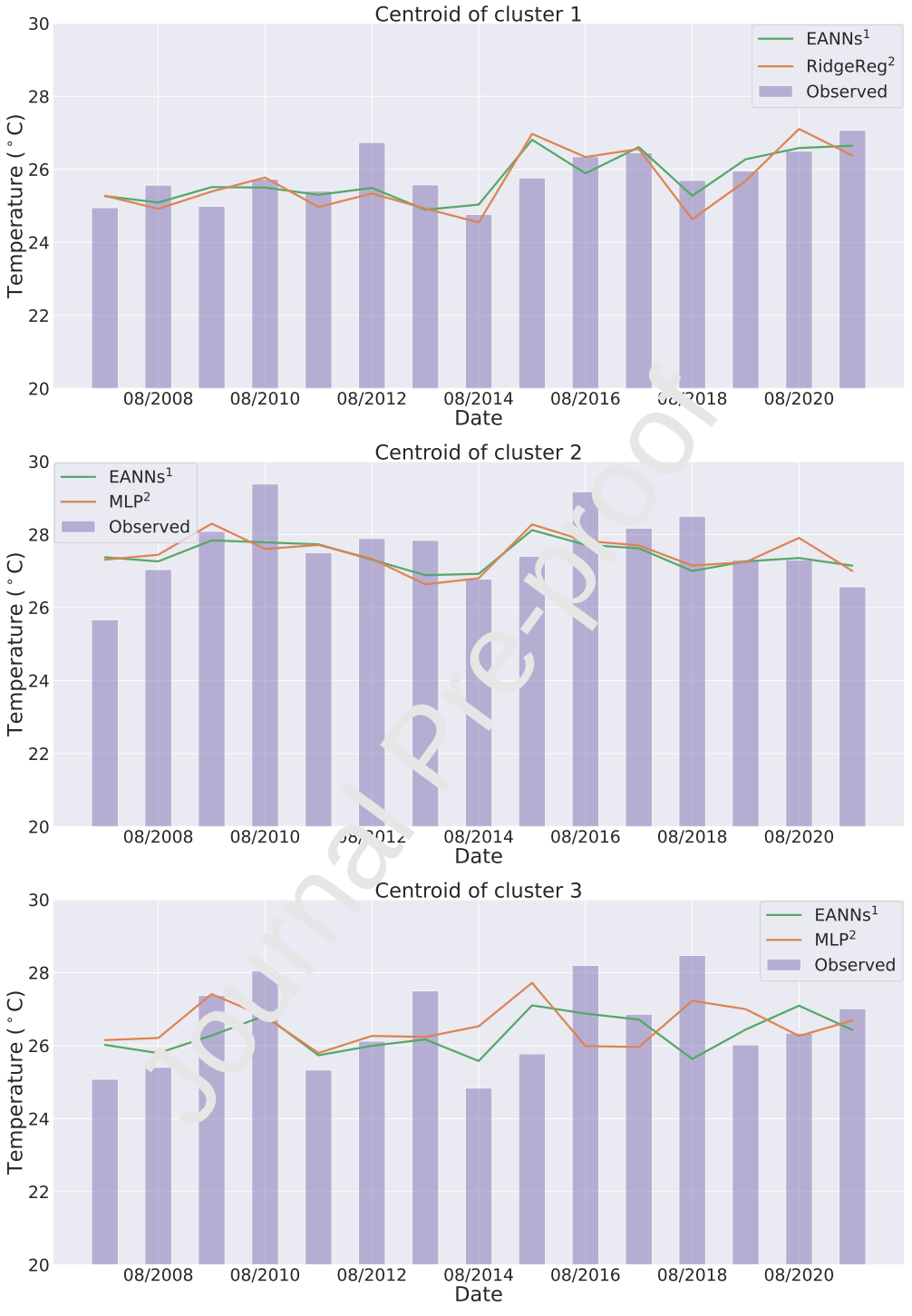


Figure 7: Graphical representation of the observed air temperature values (purple) and the predictions made by the best (green) and the second best (orange) techniques for the centroid of clusters 1 (top), 2 (middle) and 3 (bottom). The predictions shown correspond to the whole test dataset of the centroid of each cluster (15 years, 2007 to 2021).

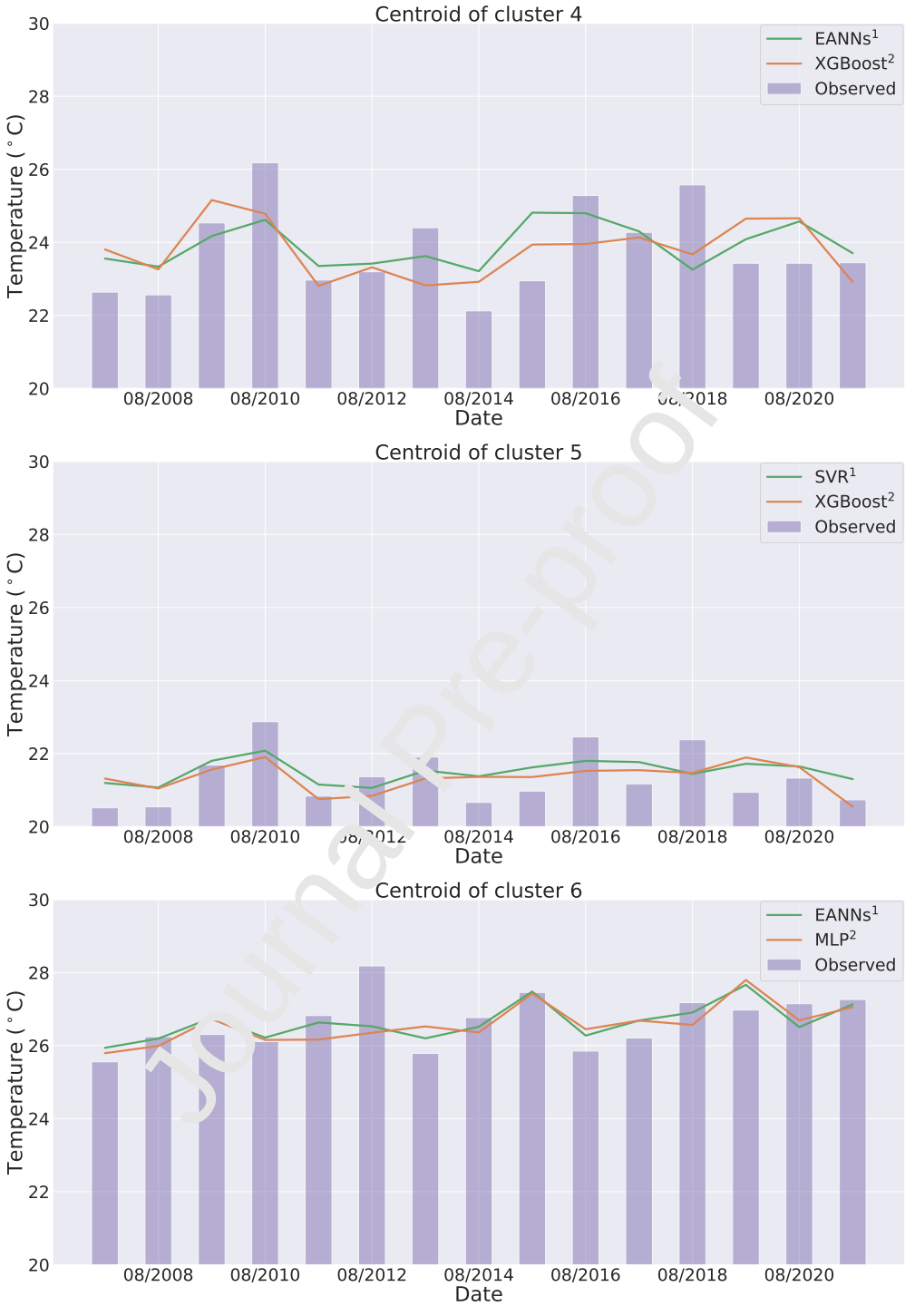


Figure 8: Graphical representation of the observed air temperature values (purple) and the predictions made by the best (green) and the second best (orange) techniques for the centroid of clusters 4 (top), 5 (middle) and 6 (bottom). The predictions shown correspond to the whole test dataset of the centroid of each cluster (15 years, 2007 to 2021).

Table 9
Comparison between the deterministic techniques and the best model obtained by each stochastic technique applied in terms of the number of connections for the different clusters.

Technique applied	# Connections					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
LinearReg	*12*	*12*	*12*	**12**	**12**	**12**
RidgeReg	*12*	*12*	*12*	**12**	**12**	**12**
LassoReg	**11**	**10**	**11**	**12**	**12**	**12**
ElasticNet	**11**	**10**	**11**	**12**	**12**	**12**
SVR	$4.9 \cdot 10^4$	$2.1 \cdot 10^4$	$1.4 \cdot 10^4$	$1.2 \cdot 10^4$	$1.8 \cdot 10^3$	$1.2 \cdot 10^4$
XAI-EANNs	36	31	26	*24*	*23*	*31*
MLP	1061	526	1081	1157	1021	544
Random Forest	$8.3 \cdot 10^6$	$3.5 \cdot 10^6$	$1.6 \cdot 10^6$	$2.2 \cdot 10^6$	$1.4 \cdot 10^4$	$9.1 \cdot 10^5$
XGBoost	$9.7 \cdot 10^4$	$8.4 \cdot 10^4$	$2.8 \cdot 10^4$	$7.8 \cdot 0^4$	$1.0 \cdot 10^4$	$2.8 \cdot 10^4$

The best results are highlighted in **bold**, whereas the second best are in *italics*.

each of the six clusters (results included in Table 5) are analysed, and the 40 runs performed are considered sufficient for statistical testing.

Firstly, the normality of the distributions of the MSE results is evaluated. To proceed with, the non-parametric Kolmogórov-Smirnov (Frank, 1951) test (K-S test) of adjustment to a normal distribution is performed. The p-values obtained, which represent the exact bilateral significance, are greater than $\alpha = 0.05$ in 23 out of the 24 tests. Hence, the hypothesis of normality is accepted. In this way, taking into account these hypotheses of normality of the samples, the ANalysis Of the VAriance (ANOVA) (Fisher, 1925, 1939) statistical test is used to check the influence of two main factors (ANOVA II (Miller, 1997)) in the MSE results: (1) technique applied (XAI-EANNs, MLP, Random Forest and XGBoost), and (2) cluster under study. This test, which analyses the mean variance, establishes whether or not such influence is significant regarding the average MSE results.

In order to perform this test, a linear model based on the best MSE value is proposed to determine the accuracy of the predictions (i.e. the significance of the influence), defined as follows:

$$\begin{aligned} MSE_{ijk} &= \mu + T_i + C_j + TC_{ij} + \varepsilon_{ijk}, \\ i \in \{1,2,3,4\},\ \ j &\in \{1,\dots,6\},\ \ k \in \{1,\dots,40\}, \end{aligned} \tag{7}$$

where μ represents the overall mean of the model (fixed influence common to all the populations), T_i analyses the influence on the average MSE for the i-th level of the factor "technique applied", with levels $i = 1$ for

Table 10
Results of the ANOVA II test

Source	Sum of squares	DF	Mean square	F ratio	p-value
Corrected model	247.288	23	10.752	387.438	$<$ **0.001**
Intercept	1063.155	1	1063.155	38311.048	$<$ **0.001**
T_i	20.025	3	6.675	240.535	$<$ **0.001**
C_j	205.273	5	41.055	1479.410	$<$ **0.001**
TC_{ij}	21.990	15	1.466	52.828	$<$ **0.001**
Error	25.975	936	0.028		
Total	1336.417	960			
Corrected total	273.262	959			

T_i: technique applied, C_j: cluster under study, TC_{ij}: interaction between both factors, DF: degrees of Freedom. Significant differences are highlighted in **bold**.

XAI-EANNs, $i = 2$ for MLP, $i = 3$ for Random Forest and $i = 4$ for XGBoost; C_j analyses the influence on the average MSE for the j-th level of the factor "cluster under study", with levels $j = 1, \ldots, 6$ for cluster $= 1, \ldots, 6$; TC_{ij} is an interaction term representing the joint influence of the presence of level i of the first factor (technique) and level j of the second one (cluster), finally, the term ε_{ijk} is the effect on the MSE of everything that could not be controlled or of random factors. Hence, the variation of the experimental results undergone by the MSE is explained through the effects produced by the four and six different levels of the two fixed factors (T_i and C_j, respectively) and by their interaction (TC_{ij}). In this way, using the linear model defined in Equation (7) and the MSE values, the ANOVA II test is performed, and the results are shown in Table 10.

From Table 10, significant differences in average MSE are observed both as a function of the technique applied (T_i) and the cluster under study (C_j), given that their p-values are both less than 0.001 and the significance level considered is $\alpha = 0.05$. Besides, the interaction between both factors (TC_{ij}) is also statistically significant. Hence, the null hypothesis that there are not significant differences between the average MSE values for the four techniques applied is rejected. For the same reason, the null hypothesis that there are not significant differences between the average MSE values for the six clusters analysed is rejected. Since there are significant differences in average MSE, the post-hoc HSD Tukey's (Tukey, 1949) test is considered to determine whether or not there are significant differences among the distinct levels of a factor. The objective is to find the level of each factor whose average MSE is significantly better than the average MSE of the remaining levels.

Table 11
Results of the HSD Tukey's test for the factor technique applied (T_i).

Technique applied	Average MSE Subset 1	Subset 2	Subset 3	*p*-value
XAI-EANNs	0.9008			0.463
XGBoost	0.9230			
Random Forest		1.1468		1.000
MLP			1.2389	1.000

HSD Tukey$^{N=240,\ \alpha=0.05}$

Table 12
Results of the HSD Tukey's test for the factor cluster under study (C_i).

Cluster	Average MSE Subset 1	Subset 2	Subset 3	Subset 4	Subset 5	Subset 6	*p*-value
Cluster 5	0.5239						1.000
Cluster 6		0.6439					1.000
Cluster 1			0.8581				1.000
Cluster 2				0.9855			1.000
Cluster 4					1.4553		1.000
Cluster 3						1.8475	1.000

HSD Tukey$^{N=160,\ \alpha=0.05}$

On the one hand, the results of the Tukey's test corresponding to the first factor (technique applied) are shown in Table 11, where the distinct levels of such factor are grouped in homogeneous subsets.

As can be observed, the results show that there are no significant differences between the average MSE values obtained by the XAI-EANNs and XGBoost techniques (i.e. both techniques are grouped in the same subset), although the average results achieved by XAI-EANNs are better. However, there are significant differences in the average values for these two techniques with respect to Random Forest, and also between the three techniques and MLP. Therefore, it can be concluded that the best results are obtained by the proposed XAI-EANNs, whereas the worst results are obtained by MLP.

On the other hand, the results of the Tukey's test corresponding to the second factor (cluster under study) are shown in Table 12, grouping in homogeneous subsets the distinct levels of such factor.

As can be seen, the results indicate that there are significant differences between the average MSE values for all the clusters (i.e. each cluster belongs to a different subset), the prediction in cluster 5 being significantly the most accurate, whereas the prediction in cluster 3 is significantly the worst.

4.4. Best models explanation

In the last few years, eXplainable Artificial Intelligence (XAI) has emerged as a novel field, in which both performance and explainability are given the same importance. Hence, the explanation of the best models achieved is paramount to its proper understanding and implementation in current processes.

From the previous statistical analysis it was depicted that XAI-EANN models were those achieving the best results in terms of MSE, as can be observed in Table 11. Furthermore, regarding the number of connections reported in Table 8, the best XAI-EANN model obtained for each cluster is also the simplest with respect to the techniques achieving competitive results. Therefore, the best XAI-EANN models obtained for the different clusters are explained next and their mathematical expressions are presented in Table 13. Note that the * indicates that the variable should be scaled before using the model.

As can be seen, each model is composed of the bias (the first term of the model), a different number of hidden neurons (which are denoted by the letter B) and their corresponding synaptic weights, which indicate how each hidden neuron contributes to the model prediction. The term $\hat{t}_a^*$ denotes the air temperature of August predicted by the model, and the function $\sigma(f(x))$ defines each hidden neuron in such a way that:

$$\sigma(f(x)) = \frac{1}{1 + e^{-f(x)}}, \tag{8}$$

where $f(x)$ is the function including the bias, the input variables and their weights.

Analysing, for example, the best model for cluster 1, it can be checked that the bias is -0.49, and the synaptic weights for the 4 neurons (B_1 to B_4) are -8.95, -6.42, 5.72 and 4.22, respectively. Concerning its first hidden neuron B_1, the bias is -1.52, and the neuron interacts with the input variables t^*, $t^*_{a_{-1}}$, v^*, $t^*_{c_6}$ and $t^*_{c_3}$ by means of the weights 4.79, 3.97, -1.10, -0.95 and -0.04, respectively.

As can be observed, each model may have a different number of neurons, depending on the cluster it has been trained for. Specifically, the best models obtained for clusters 1, 2 and 6 have 4 neurons, whereas for the remaining clusters 3, 4 and 5 they have 3 neurons. As for the weights of the input variables, the higher the absolute value of the weight, the greater the importance of the input variable. In addition, a positive weight means that increments in the input variable result in increments of t_a^*. Conversely, negative weights result in decrements of t_a^* as the input variable increases.

Table 13
Best XAI-EANN models obtained for each cluster.

Cluster 1	$\hat{t}_a^* = -0.49 - 8.95B_1 - 6.42B_2 + 5.72B_3 + 4.22B_4$ $B_1 = \sigma(-1.52 + \mathbf{4.79}t^* + \mathbf{3.98}t_{a_{-1}^*} - 1.10v^* - 0.95t_{c_6}^* - 0.04t_{c_3}^*)$ $B_2 = \sigma(-2.38 + \mathbf{4.96}t^* - \mathbf{3.27}p^* + 1.24t_{c_6}^* + 0.74v^* + 0.68t_{c_5}^* - 0.66u^* - 0.02t_{a_{-1}}^*)$ $B_3 = \sigma(3.76 - \mathbf{5.00}t_{c_2}^* - \mathbf{5.00}t_{c_5}^* - \mathbf{3.87}t_{c_4}^* - \mathbf{3.76}t^* - 1.25u^* + 0.69v^* - 0.32p^*)$ $B_4 = \sigma(5.00 + \mathbf{4.70}t_{c_3}^* - \mathbf{4.44}t^* - \mathbf{3.73}u^* - \mathbf{3.58}t_{a_{-1}}^* - 1.10t_{c_4}^* + 0.89t_{c_2}^* - 0.85p^* + 0.48s^*)$
Cluster 2	$\hat{t}_a^* = 4.04 - 15.14B_1 - 7.52B_2 + 6.77B_3 + 3.89B_4$ $B_1 = \sigma(-1.83 + \mathbf{5.00}t^* - \mathbf{1.86}t_{c_3}^* + \mathbf{1.02}t_{a_{-1}}^* + 0.30t_{c_6}^* - 0.22t_{c_4}^* - 0.20p^*)$ $B_2 = \sigma(-1.03 + \mathbf{1.58}t_{c_5}^* + \mathbf{1.05}t_{c_1}^* + \mathbf{0.96}t^* - 0.28s^*)$ $B_3 = \sigma(3.28 - \mathbf{5.00}t_{c_1}^* - \mathbf{5.00}t_{c_3}^* - \mathbf{3.91}t_{c_4}^* + \mathbf{1.62}t_{c_5}^* - 0.08p^*)$ $B_4 = \sigma(0.93 + \mathbf{4.96}t_{c_4}^* - \mathbf{4.74}u^* + \mathbf{4.11}t^* + \mathbf{3.07}t_{c_3}^* - \mathbf{2.79}t_{a_{-1}}^* - 1.64s^* - 0.75v^*)$
Cluster 3	$\hat{t}_a^* = 1.25 - 6.28B_1 + 5.58B_2 - 4.33B_3$ $B_1 = \sigma(-3.60 + \mathbf{2.61}t^* + \mathbf{2.23}t_{c_6}^* + 1.41t_{c_4}^* + 1.06u^* + 0.64t_{c_1}^* - 0.37t_{a_{-1}}^* + 0.31t_{c_2}^*)$ $B_2 = \sigma(2.56 - \mathbf{3.86}t^* - \mathbf{2.35}t_{c_1}^* + 0.27t_{c_4}^* + 0.22u^* - 0.09t_{a_{-1}}^*)$ $B_3 = \sigma(-1.14 + \mathbf{1.45}t_{c_1}^* - \mathbf{1.35}u^* + \mathbf{1.26}t_{a_{-1}}^* + \mathbf{1.06}t_{c_6}^* - 0.86t_{c_4}^* + 0.70t_{c_5}^* + 0.32v^*)$
Cluster 4	$\hat{t}_a^* = -1.96 + 13.19B_1 - 7.95B_2 - 6.82B_3$ $B_1 = \sigma(2.26 - \mathbf{5.00}t^* + \mathbf{3.29}t_{c_3}^* - \mathbf{2.61}t_{c_6}^* + 0.93p^* - 0.58t_{c_1}^*)$ $B_2 = \sigma(-3.58 + \mathbf{3.85}s^* + \mathbf{3.70}t_{c_1}^* + \mathbf{3.27}t_{c_3}^* + 1.22u^* - 0.92v^* + 0.33p^*)$ $B_3 = \sigma(-0.94 + \mathbf{4.86}t^* - \mathbf{1.67}t_{c_1}^* + \mathbf{1.58}p^* - 0.96t_{c_2}^* - 0.49u^* + 0.28s^*)$
Cluster 5	$\hat{t}_a^* = 4.19 - 9.20B_1 - 7.67B_2 + 2.07B_3$ $B_1 = \sigma(-1.90 + \mathbf{3.88}t^* + \mathbf{1.35}u^* - 0.48p^*)$ $B_2 = \sigma(0.16 + \mathbf{5.00}t_{c_6}^* - \mathbf{3.43}v^* + \mathbf{2.40}t^* - 1.67t_{c_3}^* + 1.41t_{c_2}^* - 1.32t_{a_{-1}}^* + 0.82t_{c_4}^* - 0.57p^* - 0.43t_{c_1}^*)$ $B_3 = \sigma(2.28 - \mathbf{3.03}u^* - \mathbf{1.79}t_{c_4}^* - 1.45v^* + 0.57t_{a_{-1}}^*)$
Cluster 6	$\hat{t}_a^* = -1.96 + 10.97B_1 - 9.88B_2 - 7.92B_3 + 4.10B_4$ $B_1 = \sigma(-0.02 - \mathbf{5.00}t_{c_4}^* - \mathbf{5.00}t_{c_5}^* - \mathbf{4.04}t_{c_1}^* + \mathbf{3.53}v^* + \mathbf{2.69}t_{c_3}^* - \mathbf{1.42}t^* + 0.65p^* - 0.43u^*)$ $B_2 = \sigma(-0.76 + \mathbf{4.68}t^* - \mathbf{1.36}p^* - 1.03t_{c_3}^* - 0.77u^*)$ $B_3 = \sigma(-2.07 + \mathbf{5.00}t_{a_{-1}}^* - 0.83t^* - 0.81t_{c_3}^* + 0.29t_{c_1}^* - 0.002t_{c_5}^*)$ $B_4 = \sigma(0.52 + \mathbf{1.73}t_{c_2}^* - \mathbf{1.01}t_{c_1}^* - 0.73t_{c_4}^* + 0.69t^* - 0.33u^*)$

To highlight the most significant variables, those with a high absolute value of the weight with respect to the input variables of the same neuron are shown in bold. For instance, regarding the first neuron B_1 of the model for cluster 1, the input variables t^* and $t_{a_{-1}}^*$ are the most important with respect to the remaining input variables of that neuron. Furthermore, both variables have a positive weight meaning that increments in either one increase the predicted value of t_a^*. In general, the input variables related to air temperature (t^*, $t_{a_{-1}}^*$, $t_{c_{1\ldots6}}^*$) have a highest importance across the 6 clusters. Nevertheless, the remaining input variables (u^*, v^*, s^* and p^*) also play a crucial role in accurately predicting the air temperature. Therefore, it can be said that the relative importance of the input variables depends on each cluster.

4.5. Discussion

In this section we will discuss in detail the results obtained, particularising the explanation for a cluster, in this case cluster 1, which has been used previously to illustrate the explanation of the XAI-EANN model obtained. Besides, the remaining clusters are also interpreted. As can be seen in Table 13, the XAI-EANN model for long-term air temperature prediction in cluster 1 is the following:

$$\hat{t}_a^* = -0.49 - 8.95B_1 - 6.42B_2 + 5.72B_3 + 4.22B_4, \tag{9}$$

where:

$$\begin{aligned}
B_1 =& \sigma(-1.52 + \mathbf{4.79t}^* + \mathbf{3.98t}^*_{\mathbf{a_{-1}}} - 1.10v^* - 0.95t^*_{c_6} - \\
& - 0.04t^*_{c_3}), \\
B_2 =& \sigma(-2.38 + \mathbf{4.96}t^* - \mathbf{3.27}p^* + 1.24t^*_{c_6} + 0.74[illegible] \\
& + 0.68t^*_{c_5} - 0.66u^* - 0.02t^*_{a_{-1}}), \\
B_3 =& \sigma(3.76 - \mathbf{5.00}t^*_{c_2} - \mathbf{5.00}t^*_{c_5} - \mathbf{3.87}t^*_{c_4} - [illegible]\mathbf{.76}t^* - \\
& - 1.25u^* + 0.69v^* - 0.32p^*), \\
B_4 =& \sigma(5.00 + \mathbf{4.70}t^*_{c_3} - \mathbf{4.44}t^* - [illegible] - \mathbf{3.58}t^*_{a_{-1}} - \\
& - 1.10t^*_{c_4} + 0.89t^*_{c_2} - 0.[illegible]5p^* + 0.48s^*).
\end{aligned} \tag{10}$$

It is of course a highly non-linear model, but it is interpretable (explainable). As can be seen, the most important neuron is B_1 with a negative weight of -8.95. A closer analysis of this neuron shows that it mainly takes into account air temperature variables t^* and $t^*_{a_{-1}}$, and with a smaller importance wind and other clusters temperatures v^*, $t^*_{c_6}$ and $t^*_{c_3}$. Note that in this case, the greater t^* and $t^*_{a_{-1}}$ the higher B_1, and therefore the lower t^*_a (cluster 1 air temperature, target), and the greater v^*, $t^*_{c_6}$ and $t^*_{c_3}$, the lower B_1, and therefore the greater t^*_a. This can be interpreted as follows: the years in which July average air temperature in cluster 1 (t^*) tends to be high, the August air temperature in cluster 1 tends to be lower. Similarly, the August air temperature of one year tends to be lower if the August air temperature in the last year was high. Regarding the other variables, which affect to a lesser extent to cluster 1 air temperature prediction, it seems

that if v-wind, air temperature in cluster 3 and air temperature in cluster 6 are high in July, the air temperature in cluster 1 in August tends to be also high.

Neuron B_2 takes into account the importance of other variables such as pressure and also average air temperature in cluster 1, in addition to other variables less important to the prediction. In this case, the analysis of the model indicates again that the larger the average air temperature in July, the lower in cluster 1, and the higher the pressure in July, the larger the air temperature of cluster 1. Other variables such as v-wind speed, temperatures of neighbour clusters and air temperature of August last year have an inverse effect in cluster 1 August air temperature. The results in this neuron are consistent to that of neuron B_1.

Regarding the neuron B_3, note that the most important predictive variables included in this neuron are related to air temperature. In this case, the larger $t^*_{c_2}$, $t^*_{c_5}$, $t^*_{c_4}$ and t^* in July, the lower the August air temperature of cluster 1.

Finally, neuron B_4 relies on four main variables: $t^*_{c_3}$, t^*, v (July) and $t^*_{a_{-1}}$ (August air temperature last year), with an inverse relation with August air temperature in cluster 1, i.e., the larger these variables, the lower August air temperature in cluster 1.

We can summarise this prediction model for August air temperature prediction in cluster 1 from July variables as follows: in general, when the July average air temperature in cluster 1 is high in July, the August air temperature in cluster 1 tends to be lower. High pressures in July promote higher temperatures in cluster 1. In general, high August air temperature in the last year is associated with lower August air temperature next year, and a high air temperature in July in neighbour clusters (3 and 6) implies a high air temperature in August in cluster 1. Wind components in July have direct or inverse effect in August air temperature in cluster 1, depending on the neuron considered. In general, a high zonal component u-wind is associated with lower temperatures, whereas the v-wind component effect depends on the neuron where it is involved, so it is not clear the real effect of the variable v-wind in the general behaviour of the model. There are other less important variables which serve to adjust the model, which analysis may be important to complement the principal variables discussed.

The analysis of the rest of the clusters is very similar to that of cluster 1, with some peculiarities and specific variables (see Table 13 for detail on the best models obtained for each cluster). For example, note that cluster 2 model is very similar to cluster 1. There are again 4 neurons in cluster 2, with coefficients pretty similar to that of cluster 1. Note, for example, that neuron $B1$ in cluster 2 also involves t^* and $t^*_{a_{-1}}$,

in such a way that the greater t^* and $t^*_{a_{-1}}$ the higher B_1, and the lower is the air temperature in August. In general, a high value of t^* in July is associated with a lower value of air temperature in August. This pattern is common and consistent in the models for all the clusters in the region. In a similar line, it is possible to see that, in the models for all sub-regions, a high value of persistence $t^*_{a_{-1}}$ (August air temperature last year) is also associated with a low value of the target for the present year. Wind components variables are also involved in many models, for the zonal component u^*, it is consistent that increasing values of u^* in July are associated with a lower air temperature in August, whereas increasing values of v^* in July seem to be related to increasing air temperature in August for all the sub-regions considered. Pressure variable is usually associated with increasing of August temperatures when high pressures occur in July, as in cluster 1 and cluster 6. However, note that in cluster 3, this pattern is different and here high values of pressure in July are associated with lower temperatures in August, though the coefficient of p^* in cluster 3 is lower than in the other two sub-regions where this variable is involved. The rest of important variables for the prediction in all the models are mainly related to temperatures in other sub-regions, which help the ANN obtain an accurate prediction of August air temperature for all the sub-regions considered.

5. Conclusions

In this paper we have proposed the use of eXplainable Artificial Intelligence (XAI) models in a problem of long-term air temperature prediction. The XAI model is obtained from the output of an Artificial Neural Network (ANN) with sigmoidal neurons, optimised by means of an evolutionary algorithm (EA), known as XAI-EANNs. The proposed method has been applied to a specific problem of long-term air temperature prediction (average air temperature in August) from ERA5 reanalysis (Hersbach et al., 2020) data in July. A cluster analysis has been first carried out in terms of the average air temperature in the zone under study (the Southern part of the Iberian Peninsula), so different sub-regions are obtained and analysed with the proposed approach. The proposed XAI-EANN approach has shown to be a very good method for this problem of long-term air temperature prediction, obtaining accurate prediction of air temperature for all sub-regions, with explainable models. The results obtained by the XAI-EANN approach achieves statistically significant differences against several state-of-the-art machine learning techniques, such as Multi-Layer Perceptron (MLP), Random Forest or eXtreme Gradient Boosting (XGBoost). Furthermore, note that, not only does

the proposed XAI-EANN model obtain the best results in terms of mean squared error, but also benefits from being a XAI technique, which is one of the main interest of this work.

In all the models obtained for the sub-regions considered, the average air temperature in July seemed to have a positive effect over the air temperature of August, consistent in all models obtained for all the sub-regions. The same effect occurred with the persistence variable: high values of August air temperature for last year, seemed to be associated with lower values for the August air temperature of the next year. In general, wind components have a different effect in the models obtained: whereas increasing values of u-wind component in July seemed to be associated with lower values of August air temperature, the v-wind component seemed to have a contrary effect in the models, the higher the v-wind in July, the higher the air temperature in August, which may be associated with hot winds coming from Africa. Finally, in the majority of the models, increasing values of pressure in July are associated with increasing temperatures in August, as expected, though in some of the models for specific sub-regions (cluster 4, South-West of the Iberian Peninsula and Atlantic coast of Portugal), high pressure values in July are associated with lower temperatures in August. Finally, the models obtained were fine tuned using different July air temperature variables from other sub-regions, which means a clear inter-connection or relationship among sub-regions, as expected.

As future research lines, we propose the extension of these XAI models to other climatological variables of interest or different topics such as wind speed for renewable energy studies. The application of XAI models in attribution problems could also be explored, by running the models over different datasets, with and without forcing by anthropogenic factors, and comparing the final models obtained in each case. Also, the application of long-term prediction of extreme temperature values in summer (heatwaves) is another future research line to be explored. It is an extremely difficult problem due to the scarce number of such events in the reanalysis period, which limits the application of Machine Learning or Deep Learning algorithms. In this topic, the application of XAI methods can also be of importance in order to characterise the drivers of heatwaves.

Acknowledgements

This research has been partially supported by the European Union, through H2020 Project "CLIMATE INTELLIGENCE Extreme events detection, attribution and adaptation design using machine learning

(CLINT)", Ref: 101003876-CLINT (Sancho Salcedo-Sanz). This research has also been partially supported by the projects PID2020-115454GB-C21 and PID2020-115454GB-C22 of the Spanish Ministry of Science and Innovation (MICINN). This work was also partially supported by the "Consejería de Salud y Familia (Junta de Andalucía)" (grant reference: PS-2020-780) and the "Consejería de Transformación Económica, Industria, Conocimiento y Universidades (Junta de Andalucía) y Programa Operativo FEDER 2014-2020" (grant reference: PY20_00074). Antonio M. Gómez-Orellana's research has been supported by "Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía" (Grant Ref. PREDOC-00489). David Guijo-Rubio's research is supported by the University of Córdoba through grants to Public Universities for the requalification of the Spanish university system of the Ministry of Universities, financed by the European Union - NextGenerationEU (grant reference: UCOR01MS).

CRediT authorship contribution statement

Antonio Manuel Gómez-Orellana: Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Writing - Review & Editing. **David Guijo-Rubio:** Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Writing - Review & Editing. **Jorge Pérez-Aracil:** Formal analysis, Conceptualization, Methodology, Validation, Data Curation. **Pedro Antonio Gutiérrez:** Software, Investigation, Writing - Review & Editing, Project administration, Funding acquisition. **Sancho Salcedo-Sanz:** Resources, Writing - Original Draft, Supervision, Project administration, Funding acquisition. **César Hervás-Martínez:** Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

References

Abdel-Aal, R., Elhadidy, M., 1995. Modeling and forecasting the daily maximum temperature using abductive machine learning. Weather and Forecasting 10, 310–325. doi:10.1175/1520-0434(1995)010<0310:maftdm>2.0.co;2.

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. Atmospheric Research 236, 104806. doi:10.1016/j.atmosres.2019.104806.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information fusion 58, 82–115. doi:10.1016/j.inffus.2019.12.012.

Bertini, I., Ceravolo, F., Citterio, M., De Felice, M., Di Pietra, B., Margiotta, F., Pizzuti, S., Puglisi, G., 2010. Ambient temperature modelling with soft computing techniques. Solar Energy 84, 1264–1272. doi:10.1016/j.solener.2010.04.003.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York. URL: https://dl.acm.org/doi/10.5555/235248.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer. URL: https://link.springer.com/book/9780387310732.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32. doi:/10.1023/A:1010933404324.

Chen, S.M., Hwang, J.R., 2000. Temperature prediction using fuzzy time series. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 30, 263–275. doi:10.1109/3477.836375.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Krishnapuram, B., Shah, M. (Eds.), Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM. ACM. pp. 785–794. URL: https://doi.org/10.1145/2939672.2939785.

Chevalier, R.F., Hoogenboom, G., McClendon, R.W., Paz, J.A., 2011. Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. Neural Computing and Applications 20, 151–159. doi:10.1007/s00521-010-0363-y.

Chithra, N., Thampi, S.G., Surapaneni, S., Nannapaneni, R., Reddy, A., Kumar, J.D., 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. Theoretical and Applied Climatology 121, 581–590. doi:10.1007/s00704-014-1257-1.

Cifuentes, J., Marulanda, G., Bello, A., Reneses, J., 2020. Air temperature forecasting using machine learning techniques: a review. Energies 13, 4215. doi:10.3390/en13164215.

De, S., Debnath, A., 2009. Artificial neural network based prediction of maximum and minimum temperature in the summer monsoon months over india. Applied Physics Research 1, 37. doi:10.5539/apr.v1n2p37.

Díaz, J., Garcia, R., De Castro, F.V., Hernández, E., López, C., Otero, A., 2002a. Effects of extremely hot days on people older than 65 years in seville (spain) from 1986 to 1997. International Journal of Biometeorology 46, 145–149. doi:10.1007/s00484-002-0129-z.

Díaz, J., Jordán, A., García, R., López, C., Alberdi, J., Hernández, E., Otero, A., 2002b. Heat waves in madrid 1986–1997: effects on the health of the elderly. International archives of occupational and environmental health 75, 163–170. doi:10.1007/s00420-001-0290-4.

Dikshit, A., Pradhan, B., 2021. Interpretable and explainable AI (XAI) model for spatial drought prediction. Science of the Total Environment 801, 149797. doi:10.1016/j.scitotenv.2021.149797.

Dombaycı, Ö.A., Gölcü, M., 2009. Daily means ambient temperature prediction using artificial neural network method: A case study of turkey. Renewable Energy 34, 1158–1161. doi:10.1016/j.renene.2008.07.007.

Fisher, R., 1939. The comparison of samples with possibly unequal variances. Annals of Eugenics 9, 174–180. doi:10.1111/j.1469-1809.1939.tb02205.x.

Fisher, R.A., 1925. Theory of statistical estimation. Mathematical Proceedings of the Cambridge Philosophical Society , 700–725doi:10.1017/S0305004100009580.

Frank, J.M., 1951. The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association 46, 68–78. doi:10.1080/01621459.1951.10500769.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, Articles 33, 1–22. doi:10.18637/jss.v033.i01.

Guijo-Rubio, D., Durán-Rosal, A.M., Gutiérrez, P.A., Troncoso, A., Hervás-Martínez, C., 2020a. Time-series clustering based on the characterization of segment typologies. IEEE transactions on cybernetics 51, 5409–5422. doi:10.1109/TCYB.2019.2962584.

Guijo-Rubio, D., Gómez-Orellana, A.M., Gutiérrez, P.A., Hervás-Martínez, C., 2020b. Short- and long-term energy flux prediction using multi-task evolutionary artificial neural networks. Ocean Engineering 216, 108089. doi:10.1016/j.oceaneng.2020.108089.

Gómez-Orellana, A., Guijo-Rubio, D., Gutiérrez, P., Hervás-Martínez, C., 2022. Simultaneous short-term significant wave height and energy flux prediction using zonal multi-task evolutionary artificial neural networks. Renewable Energy 184, 975–989. doi:10.1016/j.renene.2021.11.122.

Hautamaki, V., Nykanen, P., Franti, P., 2008. Time-series clustering by approximate prototypes, in: Ejiri, M., Kasturi, R., Sanniti di Baja, G. (Eds.), Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE. IEEE. pp. 1–4. URL: https://doi.org/10.1109/ICPR.2008.4761105.

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. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. doi:10.1002/qj.3803.

Jacobs, S.J., Pezza, A.B., Barras, V., Bye, J., Vihma, T., 2013. An analysis of the meteorological variables leading to apparent temperature in australia: Present climate, trends, and global warming simulations. Global and Planetary Change 107, 145–156. doi:10.1016/j.gloplacha.2013.05.009.

Karevan, Z., Suykens, J.A., 2020. Transductive lstm for time-series prediction: An application to weather forecasting. Neural Networks 125, 1–9. doi:10.1016/j.neunet.2019.12.030.

Kaufman, L., Rousseeuw, P.J., 2005. Finding groups in data: an introduction to cluster analysis. volume 344. John Wiley & Sons. URL: https://doi.org/10.1002/9780470316801.

Kendzierski, S., Czernecki, B., Kolendowicz, L., Jaczewski, A., 2018. Air temperature forecasts' accuracy of selected short-term and long-term numerical weather prediction models over poland. Geofizika 35, 19–37. doi:10.15233/gfz.2018.35.5.

Khan, N., Shahid, S., Juneng, L., Ahmed, K., Ismail, T., Nawaz, N., 2019. Prediction of heat waves in pakistan using quantile regression forests. Atmospheric Research 221, 1–11. doi:10.1016/j.atmosres.2019.01.024.

Kolevatova, A., Riegler, M.A., Cherubini, F., Hu, X., Hammer, H.L., 2021. Unraveling the impact of land cover changes on climate using machine learning and explainable artificial intelligence. Big Data and Cognitive Computing 5, 55. doi:10.3390/bdcc5040055.

Kutbay, U., et al., 2018. Partitional clustering. Recent Applications in Data Clustering 10. doi:10.5772/intechopen.75836.

Labe, Z.M., Barnes, E.A., 2022. Predicting slowdowns in decadal climate warming trends with explainable neural networks. Geophysical Research Letters , e2022GL098173doi:10.1002/essoar.10508874.3.

Lippmann, R.P., 1989. Pattern classification using neural networks. IEEE Communications Magazine 27, 47–50. doi:10.1109/35.41401.

Lubba, C.H., Sethi, S.S., Knaute, P., Schultz, S.R., Fulcher, B.D., Jones, N.S., 2019. catch22: Canonical time-series characteristics. Data Mining and Knowledge Discovery 33, 1821–1852. doi:10.1007/s10618-019-00647-x.

Mamalakis, A., Barnes, E.A., Ebert-Uphoff, I., 2022a. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. Artificial Intelligence for the Earth Systems 1, e220012. doi:10.1175/aies-d-22-0012.1.

Mamalakis, A., Ebert-Uphoff, I., Barnes, E.A., 2022b. Explainable artificial intelligence in meteorology and climate science: Model fine-tuning, calibrating trust and learning new science, in: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K., Samek, W. (Eds.), International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers, Springer. Springer. pp. 315–339. URL: https://doi.org/10.1007/978-3-031-04083-2_16.

Martínez-Estudillo, F., Hervás-Martínez, C., Gutiérrez, P., Martínez-Estudillo, A., 2008. Evolutionary product-unit neural networks classifiers. Neurocomputing 72, 548–561. doi:10.1016/j.neucom.2007.11.019.

Mayer, K.J., Barnes, E.A., 2021. Subseasonal forecasts of opportunity identified by an explainable neural network. Geophysical Research Letters 48, e2020GL092092. doi:10.1029/2020GL092092.

Mellit, A., Pavan, A.M., Benghanem, M., 2013. Least squares support vector machine for short-term prediction of meteorological time series. Theoretical and applied climatology 111, 297–307. doi:10.1007/s00704-012-0661-7.

Miller, R.G., 1997. Beyond ANOVA: Basics of Applied Statistics. Chapman and Hall/CRC, New York. URL: https://doi.org/10.1201/b15236.

Navascués, B., Calvo, J., Morales, G., Santos, C., Callado, A., Cansado, A., Cuxart, J., Díez, M., del Río, P., Escribà, P., et al., 2013. Long-term verification of hirlam and ecmwf forecasts over southern europe: History and perspectives of numerical weather prediction at aemet. Atmospheric Research 125, 20–33. doi:10.1016/j.atmosres.2013.01.010.

Niennattrakul, V., Ratanamahatana, C., 2007. Inaccuracies of shape averaging method using dynamic time warping for time series data. Computational Science–ICCS 2007 , 513–520doi:10.1007/978-3-540-72584-8_68.

Nita, I.A., Sfîcă, L., Voiculescu, M., Birsan, M.V., Micheu, M.M., 2022. Changes in the global mean air temperature over land since 1980. Atmospheric Research 279, 106392. doi:10.1016/j.atmosres.2022.106392.

Oettli, P., Nonaka, M., Richter, I., Koshiba, H., Tokiya, Y., Hoshino, I., Behera, S.K., 2022. Combining dynamical and statistical modeling to improve the prediction of surface air temperatures 2 months in advance: A hybrid approach. Frontiers in Climate 4. doi:10.3389/fclim.2022.862707.

Ortiz-García, E., Salcedo-Sanz, S., Casanova-Mateo, C., Paniagua-Tineo, A., Portilla-Figueras, J., 2012. Accurate local very short-term temperature prediction based on synoptic situation support vector regression banks. Atmospheric Research 107, 1–8. doi:10.1016/j.atmosres.2011.10.013.

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. Renewable Energy 36, 3054–3060. doi:10.1016/j.renene.2011.03.030.

Peña-Ortiz, C., Barriopedro, D., García-Herrera, R., 2015. Multidecadal variability of the summer length in europe. Journal of Climate 28, 5375–5388. doi:10.1175/jcli-d-14-00429.1.

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, 823. doi:10.3390/atmos11080823.

Peng, X., She, Q., Long, L., Liu, M., Xu, Q., Zhang, J., Xiang, W., 2017. Long-term trend in ground-based air temperature and its responses to atmospheric circulation and anthropogenic activity in the yangtze river delta, china. Atmospheric Research 195, 20–30. doi:10.1016/j.atmosres.2017.05.013.

Salcedo-Sanz, S., Casillas-Pérez, D., Del Ser, J., Casanova-Mateo, C., Cuadra, L., Piles, M., Camps-Valls, G., 2022a. Persistence in complex systems. Physics Reports 957, 1–73. doi:10.1016/j.physrep.2022.02.002.

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. Information Fusion 63, 256–272. doi:10.1016/j.inffus.2020.07.004.

Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G., Del Ser, J., Casillas-Pérez, D., Kadow, C., Fister, D., Barriopedro, D., García-Herrera, R., Restelli, M., et al., 2022b. Analysis, characterization, prediction and attribution of extreme atmospheric events with machine learning: a review. arXiv preprint arXiv:2207.07580 URL: https://arxiv.org/abs/2207.07580.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.T., 2017. A review of clustering techniques and developments. Neurocomputing 267, 664–681. doi:10.1016/j.neucom.2017.06.053.

Ser, J.D., Osaba, E., Molina, D., Yang, X.S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P.N., Coello, C.A.C., Herrera, F., 2019. Bio-inspired computation: Where we stand and what's next. Swarm and Evolutionary Computation 48, 220–250. doi:10.1016/j.swevo.2019.04.008.

Smith, B.A., Hoogenboom, G., McClendon, R.W., 2009. Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture 68, 52–61. doi:10.1016/j.compag.2009.04.003.

Smith, B.A., McClendon, R.W., Hoogenboom, G., 2006. Improving air temperature prediction with artificial neural networks. International Journal of Computational Intelligence 3, 179–186. doi:doi.org/10.5281/zenodo.1075076.

Tuia, D., Roscher, R., Wegner, J.D., Jacobs, N., Zhu, X., Camps-Valls, G., 2021. Toward a collective agenda on ai for earth science data analysis. IEEE Geoscience and Remote Sensing Magazine 9, 88–104. doi:10.1109/MGRS.2020.3043504.

Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114. doi:10.2307/3001913.

Ustaoglu, B., Cigizoglu, H., Karaca, M., 2008. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling 15, 431–445. doi:10.1002/met.83.

Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media. URL: https://doi.org/10.1007/978-1-4757-2440-0.

Vuori, V., Laaksonen, J., 2002. A comparison of techniques for automatic clustering of handwritten characters, in: Kasturi, R., Laurendeau, D., Suen, C. (Eds.), Pattern Recognition, 2002. Proceedings. 16th International Conference on IEEE. IEEE. pp. 168–171. URL: https://doi.org/10.1109/ICPR.2002.1047821.

Williams, S., Nitschke, M., Sullivan, T., Tucker, G.R., Weinstein, P., Pisaniello, D.L., Parton, K.A., Bi, P., 2012. Heat and health in adelaide, south australia: assessment of heat thresholds and temperature relationships. Science of the Total Environment 414, 126–133. doi:10.1016/j.scitotenv.2011.11.038.

Xu, Z., Liu, Y., Ma, Z., Li, S., Hu, W., Tong, S., 2014. Impact of temperature on childhood pneumonia estimated from satellite remote sensing. Environmental research 132, 334–341. doi:10.1016/j.envres.2014.04.021.

Yao, X., 1999. Evolving artificial neural networks. Proceedings of the IEEE 87, 1423–1447. doi:10.1109/5.784219.

Yao, X., Liu, Y., 1997. A new evolutionary system for evolving artificial neural networks. IEEE Transactions on Neural Networks , 694–713doi:10.1109/72.572107.

Ye, L., Yang, G., Van Ranst, E., Tang, H., 2013. Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. Advances in Atmospheric Sciences 30, 382–396. doi:10.1007/s00376-012-1252-3.

You, Q., Fraedrich, K., Min, J., Kang, S., Zhu, X., Ren, G., Meng, X., 2013. Can temperature extremes in china be calculated from reanalysis? Global and Planetary Change 111, 268–279. doi:10.1016/j.gloplacha.2013.10.003.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320. doi:10.1111/j.1467-9868.2005.00503.x.

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Department of Computer Science and Numerical Analysis
Rabanales Campus.
14071 – Córdoba (Spain)

Atmospheric Research

Title: *"One month in advance prediction of air temperature from Reanalysis data with eXplainable Artificial Intelligence techniques"*

Authors: *A. M. Gómez-Orellana, D. Guijo-Rubio, J. Pérez-Aracil, P. A. Gutiérrez, S. Salcedo-Sanz, C. Hervás-Martínez*

Author Statement

Antonio Manuel Gómez-Orellana: Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Writing - Review & Editing.
David Guijo-Rubio: Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Writing - Review & Editing.
Jorge Pérez-Aracil: Formal analysis, Conceptualization, Methodology, Validation, Data Curation.
Pedro Antonio Gutiérrez: Software, Investigation, Writing - Review & Editing, Project administration, Funding acquisition.
Sancho Salcedo-Sanz: Resources, Writing - Original Draft, Supervision, Project administration, Funding acquisition.
César Hervás-Martínez: Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

David Guijo Rubio

Department of Computer Science and Numerical Analysis, University of Córdoba,

e-mail address: dguijo@uco.es