

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Satellite Data and Supervised Learning to Prevent Impact of Drought on Crop Production: Meteorological Drought

*Leonardo Ornella, Gideon Kruseman and Jose Crossa*

## Abstract

Reiterated and extreme weather events pose challenges for the agricultural sector. The convergence of remote sensing and supervised learning (SL) can generate solutions for the problems arising from climate change. SL methods build from a training set a function that maps a set of variables to an output. This function can be used to predict new examples. Because they are nonparametric, these methods can mine large quantities of satellite data to capture the relationship between climate variables and crops, or successfully replace autoregressive integrated moving average (ARIMA) models to forecast the weather. Agricultural indices (AIs) reflecting the soil water conditions that influence crop conditions are costly to monitor in terms of time and resources. So, under certain circumstances, meteorological indices can be used as substitutes for AIs. We discuss meteorological indexes and review SL approaches that are suitable for predicting drought based on historical satellite data. We also include some illustrative case studies. Finally, we will survey rainfall products existing at the web and some alternatives to process the data: from high-performance computing systems able to process terabyte-scale datasets to open source software enabling the use of personal computers.

**Keywords:** remote sensing, supervised learning, meteorological index, wavelet

## 1. Introduction

Climate change is shifting the rainfall patterns and increasing the severity of droughts and floods around the Earth. Australia [1], Europe, and the rest of the continents have been affected by a number of major drought events [2]. In 2018, drought and heat waves reduced harvests up to 40–50% in some countries of northern and central Europe [3].

Drought is by far the Earth's most costly natural disaster and can have wide-spread impacts [4]. Globally, it is responsible for 22% of the economic damage caused by natural disasters and 33% of the damage in terms of the number of people affected [5]. Though average yields rose steadily between 1947 and 2008, there is no evidence that relative stress tolerance has improved [6, 7]. Therefore, until breeding programs develop adapted germplasm, drought forecasting will be

important to determine when to take contingency actions to prevent drought and mitigate its risk and impacts.

The practice of drought forecasting remains challenging and is subject to great uncertainty partly due to the instability of the components of the hydrologic cycle (e.g., rainfall, soil moisture, groundwater level, etc.); temporal variability involving trends, oscillating behavior, and sudden shifts that appear in hydroclimatic records, thus posing challenges to drought prediction [8, 9].

Although agricultural indices (AIs) better reflect the soil water situation that influences crop conditions, monitoring of soil moisture is costly in terms of time and resources [10]. On the other hand, some meteorological indices (e.g., the Standard Precipitation Index) can be calculated just knowing the precipitation (pp) data, and then the expert can give a very close condition of the vegetation [11].

A variety of methods has been developed to predict drought occurrence: statistical run theory [12], Markov chain [13], loglinear [14], renewal process [15], and Poisson process [16], among others.

A valuable alternative to the aforementioned methods is machine learning (ML), a branch of artificial intelligence that studies how to extract information from big data sets with minimal human intervention. ML has been successfully tested in very different areas such as bioinformatics [17], crop protection [18], and economics [19], among others. Therefore, its potential for predicting the climate seems far from being fully exploited.

The remainder of this chapter is organized as follows: in Section 2, we introduce some representative ML methods that have been proposed for drought forecasting; in Section 3, we present the concept of meteorological drought, in particular the standardized precipitation index that is considered a primary drought indicator. Section 4 describes some forecast examples using the abovementioned methods. In Sections 5 and 6, we review satellite precipitation products and how to access and process them; and finally, in Section 7 we present the conclusions of this work.

## **2. Machine learning**

ML is the science of algorithms and statistical models that computer systems use to progressively improve their performance of a specific task. They can be broadly categorized into supervised and unsupervised learning. In SL (classification or regression), the algorithm builds a function from a set of data relating the inputs to the outputs. In regression, the outputs are continuous, meaning they may have any value within a range (e.g., temperature and moisture), while in classification, the outputs are restricted to a limited set of values.

In unsupervised learning, the algorithm builds a mathematical model of a data set that contains inputs and no outputs. These unsupervised learning procedures are used to find structure in the data (e.g., cluster data) or reduce its dimensionality.

Examples of ML are land classification using remote sensing [20–22], amending satellite data assimilation [23], or decomposing the causes of climate change [24].

### **2.1 Support vector regression (SVR) and least squares support vector regression (LS-SVR)**

SVR is based on the Vapnik-Chervonenkis (VC) theory [25], which characterizes the properties of learning machines that enable them to generalize the unobserved data well.

Starting with the simplest example, that is, linear regression, the objective of both SVR [26] and LS-SVR [27] is to fit a linear relation  $y = \mathbf{w}^T \mathbf{x} + b$  between the  $\mathbf{x}$  regressors and the dependent variable  $y$  in the so-called feature space. In SVR, the problem is solved by minimizing

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (1)$$

under the constraints

$$\begin{aligned} y - \mathbf{w}^T \mathbf{x}_i - b &\leq \varepsilon + \xi_i \\ y - \mathbf{w}^T \mathbf{x}_i - b &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (2)$$

while for LS-SVR, the objective is to minimize

$$\|\mathbf{w}\|^2 + \gamma \sum_{i=1}^n e_i^2 \quad (3)$$

under the constraints

$$y_i - \mathbf{w}^T \mathbf{x}_i - b \quad (4)$$

Both methods are very similar, but in LS-SVR, the objective is to minimize the more usual sum of the squares of the errors, by replacing the  $\varepsilon$ -tube or  $\varepsilon$ -insensitive loss of SVR, that is, by ignoring all regression errors smaller than  $\varepsilon$  (**Figure 1**). Solving a nonlinear regression demands a “kernel trick” [26]. This trick uses kernel functions to transform the data of the input space into a higher dimensional feature space to make it possible to perform a linear regression. Common kernels are

$$\text{polynomial } k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d \quad (5)$$

$$\text{Gaussian radial basis function } k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right) \quad (6)$$

LS-SVR is an economic alternative to the original SVR model. It only relies on the cost function on a sum-of-squared-error (SSE) and equality constraints, instead of the computationally complex and time-consuming quadratic programming problem in SVR [28].

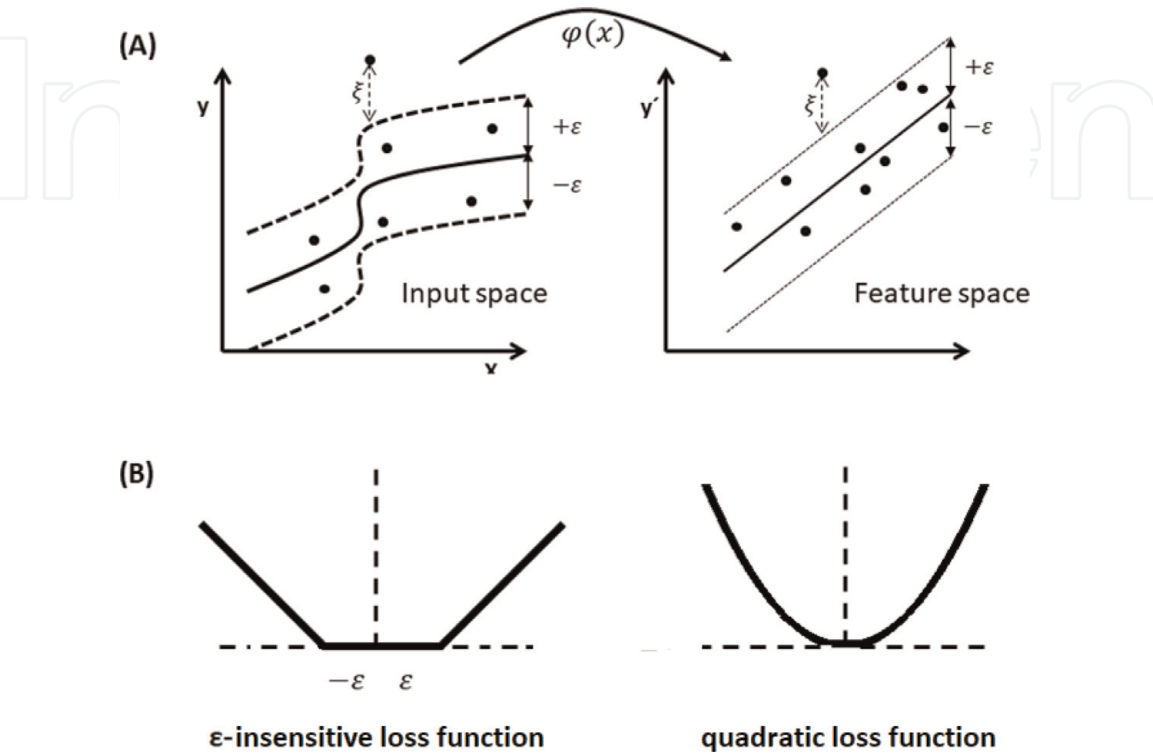
For optimal performance, parameter tuning is necessary [29]: for SVR,  $C$  and  $\varepsilon$  and the kernel-related parameters (e.g.,  $\sigma^2$  for the RBF kernel) and for LS-SVR,  $g$  (the regularization parameter determining the trade-off between the fitting error and the smoothness of the estimated function) and the kernel-related parameters. For further information about SVR in general, the reader should refer to [30].

## 2.2 Artificial neural network (ANN)

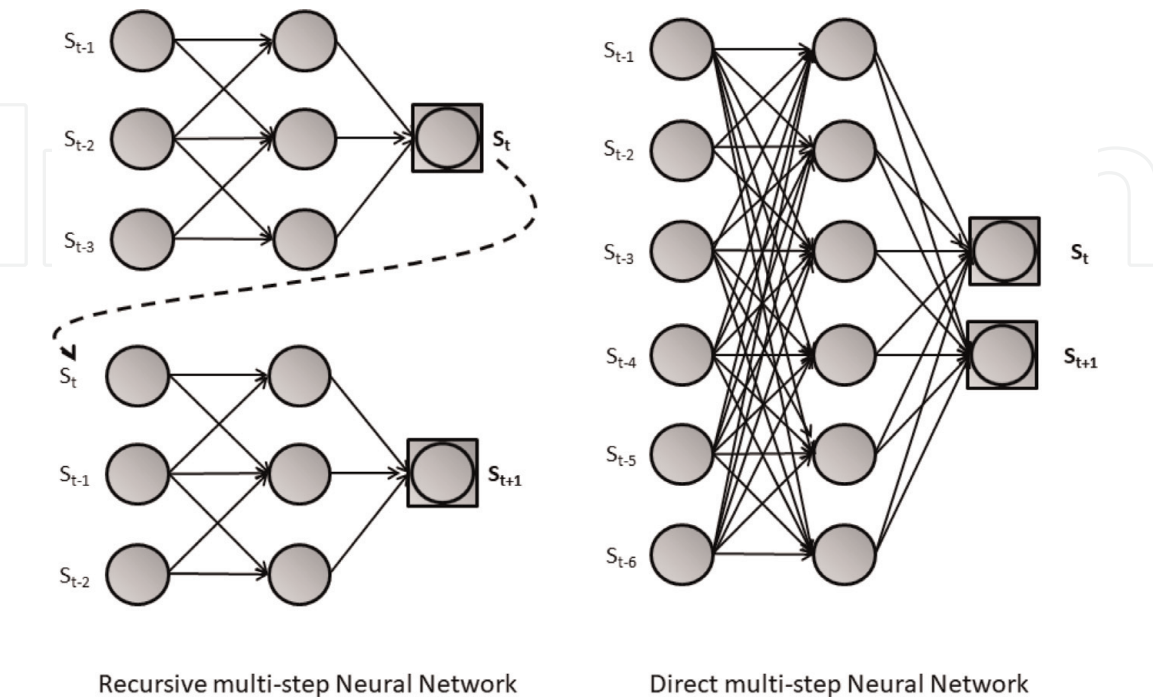
An ANN is a supervised learning model based on the operation of biological neurons. There are many architectures and training algorithms for ANN. The multilayer perceptron network (MLPN), the most common ANN architecture used for forecasting, consists of a feedforward neural network with at least three layers of neurons: an input layer, one or more hidden layers, and an output layer with a directed acyclic graph representation network (**Figure 2**). The input layer receives the data vector  $\mathbf{x}$ , while the output layer gives the output vector  $\mathbf{y}$ . An activation

function is applied to activate the neurons in the hidden layer. For a three-layer network system, the nonlinear mapping between input  $x$  and output  $y$  is given by the equation:

$$y = \frac{1}{2} \left[ \sum_{j=1}^h \left( w_j f_1 \sum_{i=0}^n w_{ji} x_i \right) \right] \tag{7}$$



**Figure 1.** (A) Kernel trick: mapping the data from the input space into a feature space. (B) Loss function used in support vector regression ( $\epsilon$ -insensitive loss) and least squares support regression (quadratic).



**Figure 2.** Architectures of forecasting artificial neural networks. Recursive multistep neural network versus direct multistep neural network.



An ANN is usually learned by adjusting the weights and biases in order to minimize a cost function, usually MSE using the error back-propagation algorithm.

Of the activation functions, we should mention the hyperbolic tangent  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  and the sigmoidal function:  $\sigma(x) = \frac{e^x}{1 + e^x}$ .

The number of hidden neurons is no less important, since a wrong number may cause either overfitting or underfitting problems. Normally it is selected via trial and error, but this is computationally costly. Several heuristics or formulas have been proposed to avoid this cumbersome work, and success depends on the type of data, the complexity of the network architecture, etc. [31].

Last but not least, ANN forecasting models can be separated into two broad groups, namely, the recursive multistep neural network (RMSNN) and the direct multistep neural network (DMSNN) (**Figure 2**). In RMSNN, the model forecast one time-step ahead, and the network is applied recursively, using previous predictions as inputs for subsequent forecasts—that is, a forecast horizon of 3 months will have, as inputs, the outputs of forecasts with lead times of 1 and 2 months.

Similar to the RMSNN model, the DMSNN approach has a single or multiple neurons in both the input and hidden layers. However, it can have several neurons in the output layer representing multiple-month lead time forecasts. Similar to the RMSNN model, the DMSNN model is designed to forecast drought conditions using the present index value and several months of past index values as inputs.

### 2.3 Deep belief networks (DBN)

ANNs are suitable for complex time series forecasting but have several weaknesses: (1) selection of the initial values of the weights (normally at random) can affect the learning process, leading to slower convergence or to different forecast results for each training process and (2) the training process may get stuck at local optima, especially in networks with several hidden layers. Hinton et al. [32] proposed a probabilistic generative model with multiple hidden layers that uses layer-wise unsupervised learning to pre-train the initial weights of the network and then fine-tune the whole network using standard supervised methods such as the back-propagation algorithm.

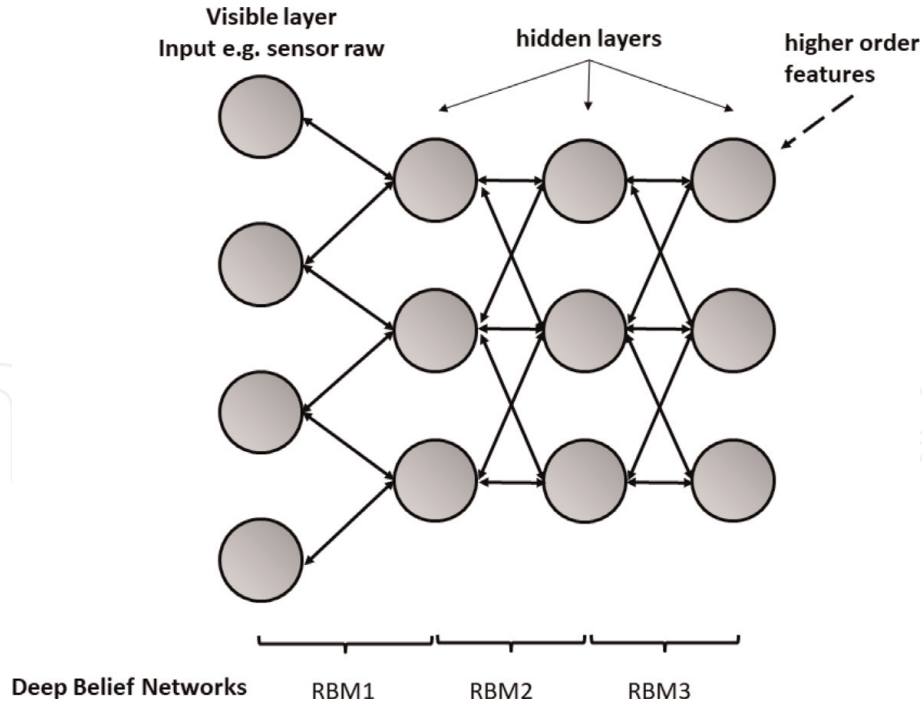
Classically, a DBN is constructed by stacking multiple restricted Boltzmann machines (RBMs) on top of each other (**Figure 3**). The layers are trained by using the feature activations of one layer as the training data for the next layer. Better initial values of weights in all layers are obtained by greedy layer-wise unsupervised training, and the entire network is fine-tuned using an SL algorithm. Pre-training can be done with principal component analysis or nonlinear generalization [33].

An RBM [34] is a neural network model used for unsupervised learning. Typically, it consists of a single layer of hidden units (the outputs) with undirected and symmetrical connections to a layer of visible units (the data) (**Figure 3**). The configuration (bipartite graph) defines the state of each unit. Only connections between a hidden unit and a visible unit are permitted—that is, no connections between two visible units or between two hidden units are allowed. An RBM is a special type of generative energy-based model that is defined in terms of the energies of the configurations between visible and hidden units.

The standard type of RBM has binary-valued (Boolean/Bernoulli) hidden and visible units.

### 2.4 Bagging

Bootstrap aggregating, or bagging, is an ML ensemble meta-algorithm designed to increase the stability and accuracy of unstable procedures, for example, artificial



**Figure 3.**  
Basic deep belief network (DBN) structure with three hidden layers.

neural networks or decision trees [35]. Given a standard training set  $T$  of size  $n$ , the algorithm sample is taken from  $T$  uniformly and with replacement  $m$  new training sets,  $T'$ , each of size  $n'$  (some observations may be repeated in each  $D$ ). This process is known as a bootstrap sampling [36]. The basic idea is that the samples are de-correlated, and this reduces the expected error as  $m$  increases.

The  $m$  models are fitted using the above  $m$  bootstrap samples, and results of an unknown instance are obtained by averaging the output (for regression) or by voting (for classification) (**Figure 4**).

This method may slightly degrade the performance of stable algorithms (e.g.,  $k$ -nearest neighbor) because smaller training sets are used to train each algorithm.

Bagging does not necessarily improve forecast accuracy in all cases. Nevertheless, this method and its derivatives tend to outperform traditional forecasting procedures [37].

## 2.5 Random forest regression (RFR)

A random forest (RF) [38] is a collection of  $K$  binary recursive partitioning trees, where each tree is grown on a subset of  $n$  instances extracted with replacement from the original training data. It is an instance of bagging where the individual learners are de-correlated trees. Each tree is grown in a top-down recursive manner, from the root node to terminal nodes or leaves (**Figure 5**). In each node, a random sample of  $m$  ( $m \approx p/3$ ) predictors is chosen as candidates from the full set of  $p$  predictors. The data are partitioned into the two descendant branches by choosing the variable that minimized:

$$RSS = \sum_{left} (y_i - y_L^*)^2 - \sum_{right} (y_i - y_R^*)^2 \quad (8)$$

The advantage of selecting a random subset of predictors is that two trees generated on same training data will be de-correlated (independent of each other) because randomly different variables were selected at each split. Each internal (non-leaf) node is signed with a predictor determined by the  $RSS$  test, and each one

of the two possible subsets of this variable labels the arcs connecting to the subordinate decision node. Each tree extends as much as possible until all the terminal nodes are maximally homogeneous (a minimum of five examples in each leaf is recommended).

Once the random forest is generated, the output of new data is obtained by averaging the predictions of the K trees.

The number of trees influences the error of prediction; it decreases as the number of trees (ntree) grows, but there is a threshold beyond which there is no significant gain [38, 39]. In general,  $ntree \approx 500$  gives good results [40]. RF can

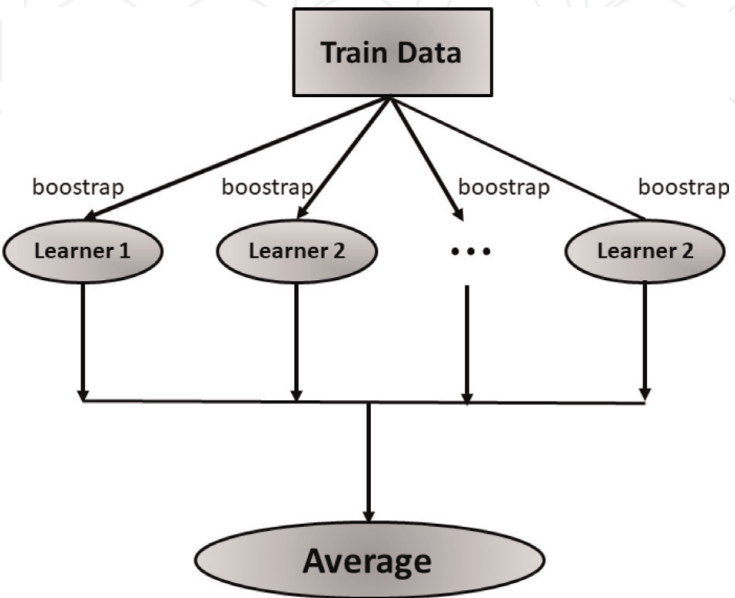


Figure 4.  
Structure of bootstrap aggregating, or bagging.

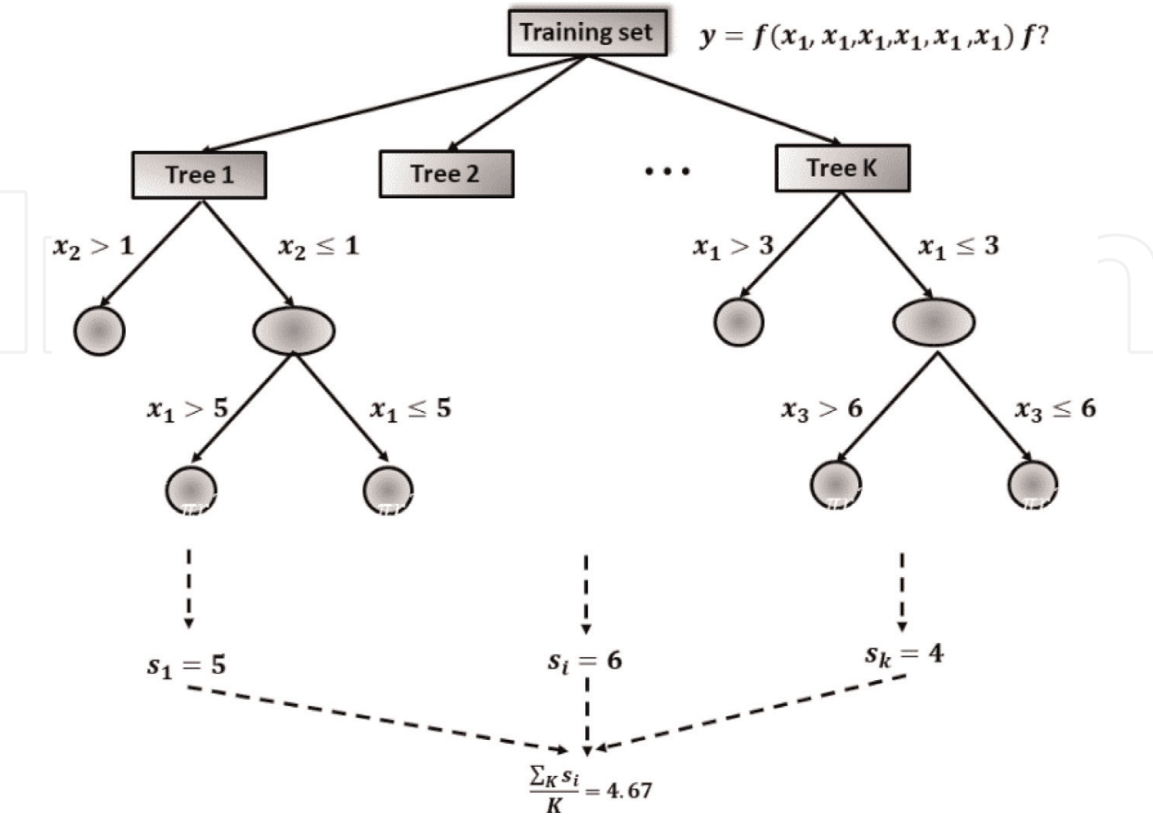


Figure 5.  
Architecture of the random forest model.



successfully handle high dimensionality and multicollinearity, because it is both fast and insensitive to overfitting. It is, however, sensitive to the sampling design.

## 2.6 Adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS)

ANFIS is a hybrid learning procedure which employs the linguistic concept of fuzzy systems (human knowledge) and the training power of the ANN to solve a regression problem [41]. All ANFIS works reported here are based on the Takagi-Sugeno fuzzy inference system [42], where the fuzzy rule applied has the form: if  $x$  is  $A$  and  $y$  is  $B$  then  $z = f(x, y)$ . Other fuzzy methods are Mamdani-type or Tsukamoto-type [42].

**Figure 6** depicts a typical ANFIS architecture. Square nodes (adaptive nodes) have parameters, while circle nodes (fixed nodes) do not. The first and the fourth layers contain the parameters that can be modified over time. A particular learning method was required to update these parameters.

In layer 1, every node is adaptive and associated with an appropriate continuous and piecewise differentiable function such as Gaussian, generalized bell-shaped, trapezoidal-shaped, and triangular-shaped functions.

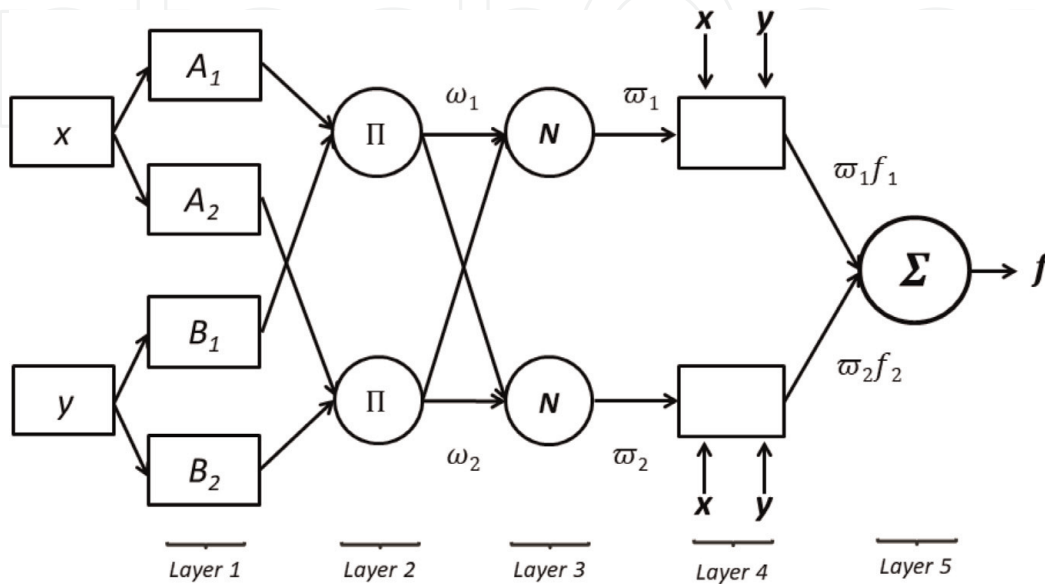
In layer 2, every node is fixed and represents the firing strength of each rule. This is calculated by the fuzzy and connective method of the “product” of the incoming signals, that is,  $O_i^2 = w_i = \mu_{A_i}(x) * \mu_{B_i}(y)$ ,  $i=1, 2$ .

In layer 3, every node is also fixed, showing the normalized firing strength of each rule. The  $i$ th node calculates the ratio of the  $i$ th rule’s firing strength to the summation of two rules’ firing strengths.

In every adaptive node of layer 4 (consequent nodes) is a function indicating the contribution of the  $i$ th rule to the overall output:  $O_{4,i} = \bar{w}_i f = \bar{w}_i (p_i + q_i + r_i)$ , where  $w_i$  is the output of layer 3 and  $p_i, q_i, r_i$  is the parameter set. Finally, layer 5 (output node) is a single node that computes the overall output of the ANFIS as:

$$O_{5,1} = \sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}.$$

One of the most important steps in developing a satisfactory forecasting model is the selection of the input variables. These variables determine the structure of the forecasting model and affect the weighted coefficients and the results of the model



**Figure 6.**  
Architecture of an adaptive network-based fuzzy inference system (ANFIS).

function in layer 2. As the number of parameters increases with the fuzzy rule increment, the model structure becomes more complicated. A very good description of ANFIS is presented in [43, 44].

## 2.7 Boosting

Boosting attempts to increase the performance of a given learning algorithm by iteratively adjusting the weight of an observation based on the last training/testing process. In other words, the meta-algorithm produces a sequence of models by adaptive reweighting of the training set [45].

AdaBoost, the first boosting algorithm, is definitely beaten by noisy data; its performance is highly affected by outliers, as the algorithm tries to fit every point perfectly. Friedman [46] extended the concept to present gradient boosting, which constructs additive regression models by sequentially fitting a simple parameterized function (base learner) to current “pseudo”-residuals by least squares at each iteration. The pseudo-residuals are the gradient of the loss function being minimized with respect to the model values at each training data point evaluated in the current step. This reduces the loss of the loss function. We iteratively added each model and computed the loss. The loss represents the difference between the actual value and the predicted value (the error residual), and using this loss value, the predictions are updated to minimize these residuals.

A regularization method that penalizes various parts of the boosting algorithm is necessary to avoid overfitting. This generally improves the performance of the algorithm by reducing overfitting.

## 2.8 Hybrid models

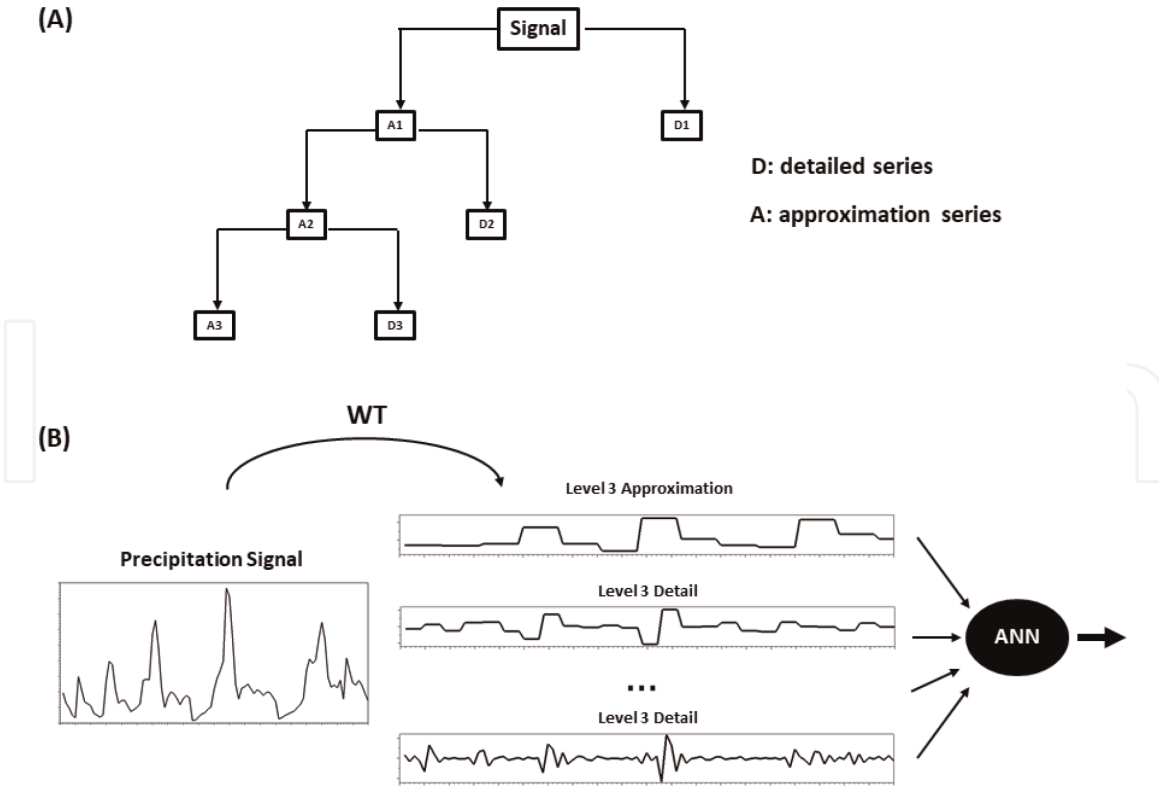
The time series that characterize the evolution of meteorological events (drought, precipitation) in the temporal domain have localized high- and low-frequency components with dynamic nonlinearity and non-stationary features. MM models have not always proven to be good at capturing the behavior of the time series. Hybrid models can perform superbly when forecasting hydrological and climatological time series. Different combination techniques have been proposed in order to overcome the deficiencies of single models and improve forecasting performance [47]. Many combined models have been introduced in the literature, for example, ANN-ARIMA [48], SVR-ARIMA [49], etc.

Here we will only focus on WT-ML hybrids, where ML is a machine learning method (e.g., ANN or SVR) and WT is a discrete wavelet transform [50].

### 2.8.1 Wavelet transform (WT)

WT is a time-dependent spectral analysis that decomposes time series in the time-frequency space and provides a timescale illustration of processes and their relationships. In this method, the data series are broken down by transforming them into “wavelets,” which are scaled and shifted versions of a mother wavelet [50]. This allows the use of long time intervals for low-frequency information and shorter intervals for high-frequency information and can reveal aspects of data such as tendencies, breakdown points, and discontinuities that other signal analysis techniques might miss, for example, Fourier transform.

There are two main alternatives for WT: discrete wavelet transform (DWT) and continuous wavelet transform (CWT). For DWT, the WT is applied using a discrete set of the wavelet scaling and shifting, whereas in the case of CWT, this scaling and shifting is continuous—that is, CWT is computationally expensive



**Figure 7.** Time series wavelet-ANN conjunction model. (A) Three-level wavelet decomposition tree (DWT). (B) Example of the decomposition of a precipitation signal.

and most researchers use DWT. For more information about CWT, the reader should refer to [51].

DWT operates two sets of functions (scaling and wavelets) viewed as high-pass (HPF) and low-pass (LPF) filters. The signal is convolved with the pair of HPF and LPF followed by subband downsampling producing two components. The first component, which is obtained by passing the signal through the low-pass filter, is called an approximation component (or series), and the other component (fast events) is called a detailed component (**Figure 7**). This process is iterated  $n$  times with successive approximation series being decomposed in turn, so that the original time series is broken down into the minimum number of components needed to reflect the time series according to the mother wavelet.

The filterbank implementation of wavelets can be interpreted as computing the wavelet coefficients of a discrete set of child wavelets for a given mother. This mother wavelet function was defined at scale  $a$  and location  $b$  as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (9)$$

$\psi_{0,0}(t)$  is a mother wavelet prototype and  $a, b$  are scaling and shifting parameters, respectively.

Several wavelet families have proven useful for forecasting various hydrological time series. As an example, we can mention Haar, which is also known as daubechies1 or db1 [50]. It is defined as

$$\begin{cases} 1 & \text{if } 0 < t < 0.5 \\ -1 & \text{if } 0.5 < t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

A full description of DWT can be found in [50, 52].

### 3. Meteorological indices

Drought can be defined as a period of unusually arid conditions (usually due to rainfall deficiency) that have lasted long enough to cause non-balance in a region's hydrological situation. Based on its intensity and persistence, drought can be classified into four categories [53]: (1) meteorological drought, which occurs when precipitation is less than usual, is characterized by changes in weather patterns; (2) agricultural (vegetation) drought refers to water deficits in plants; it occurs after meteorological drought and before hydrological drought; (3) hydrological drought ensues when the level of surface water and the groundwater table are less than the long-term average; and finally, (4) socioeconomic drought materializes when water resources required for industrial, agricultural, and household consumption are less than required and thus cause socioeconomic anomalies.

A drought index is an indicator or measure derived from a series of observations that reveals some of the cumulative effects of a prolonged and abnormal water deficit. It integrates pertinent meteorological and/or hydrological parameters (accumulated precipitation, temperature, and evapotranspiration) into a single numerical value or formula and gives a comprehensive picture of the situation [53]. Such an index is more readily usable and comprehensible than the raw data and, if presented as a numerical value, makes it easier for planners and policymakers to make decisions. Authorities and public and private committees evaluate the impact of drought using these indices and take measures to prevent its effects [54].

More than 100 drought indices have so far been proposed, and each one has been formulated for a specific condition [55]. The reclamation drought index (RDI), for example, was developed in the USA to activate drought emergency relief funds associated with public lands affected by drought; the crop moisture index (CMI) was designed to show the effects of water conditions on growing crops in the short term and is not a good instrument for displaying long-term conditions. Here we will only describe the standardized precipitation index, which those indices used in case studies.

#### 3.1 Standardized precipitation index (SPI)

Most of the forecasting works reviewed here are based on SPI [56]. It is perhaps the most popular index for forecasting meteorological drought and has been recommended by the World Meteorological Organization [57]. It can be defined as the number of standard deviations that the observed cumulative rainfall at a given time scale (1,3,6 month) would deviate from the long-term mean for that same time scale over the entire length of the record (z-score).

More specifically, SPI is calculated by building a frequency distribution from historical precipitation data (at least 30 years) at a specific location for the precipitation accumulated during a specified period, for example, 1 month (SPI1), 3 months, (SPI3), 24 months (SPI24), and so on. A theoretical probability density function (usually the gamma distribution) is fitted to the empirical distribution for the selected time scale.

SPI1 to SPI6 are considered indices for short-term or seasonal variation (soil moisture), whereas SPI12 is considered a long-term drought index (groundwater and reservoir storage).

The "drought" part of the SPI range is arbitrarily split into "near normal" ( $0.99 > \text{SPI} > -0.99$ ), "moderately dry" ( $-1.0 > \text{SPI} > -1.49$ ), "severely dry"



( $-1.5 > \text{SPI} > -1.99$ ), and “extremely dry” ( $\text{SPI} < -2.0$ ) conditions [56]. A drought event starts when SPI becomes negative and ends when it becomes positive again.

SPI is easy to calculate (using precipitation only) and can characterize drought or abnormal wetness on different time scales. Its standardization ensures independence from geographical position, and it is thus more comparable across regions with different climates. The index can be computed using several packages of the R project [58], for example, the SPEI package [59] or the SPI package [60]. Limitations of SPI include the following: (1) it does not account for evapotranspiration; (2) it is sensitive to the quantity and reliability of the data used to fit the distribution; and (3) it does not consider the intensity of precipitation and its potential impacts on runoff, streamflow, and water availability within the system. A more detailed explanation of how SPI is calculated can be found at [43].

### 3.2 Other indices

Other indices including only precipitation data are EDI [61], SIAP [62], deciles index (DI), percent of normal (PN), standard precipitation index (SPI), China-Z index (CZI), modified CZI (MCZI), and z-score [55].

## 4. Forecasting meteorological drought

Forecasting meteorological drought using historical data is not a trivial task. The time series that characterize the evolution of meteorological events (drought, precipitation) in the temporal domain have localized high- and low-frequency components with dynamic nonlinearity and non-stationary features. Several statistical indicators have been proposed to evaluate the success of prediction. Most of these metrics are not independent; for example, MSE can be decomposed in many ways to link it with the bias and the correlation coefficient [63]. A standard practice of model corroboration is to compute a common set of performance metrics, typically more than three. Most important is that at least three critical components, that is, one dimensionless statistic, one absolute error index statistic, and one graphical technique, should be represented in the corroboration [64].

Regarding the dimensionless statistic, we must mention:

- Pearson’s correlation coefficient (R) is used to evaluate how well the estimates correspond to the observed values. Due to the standardization of many indices, the robustness of R can be limited [64].

$$R = \frac{\sum_{i=1}^n (p_i - \bar{p})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2} \sqrt{\sum_{i=1}^n (o_i - \bar{o})^2}} \quad (11)$$

- Coefficient of determination (R<sup>2</sup>) measures the degree of association among the observed ( $o_i$ ) and predicted values ( $p_i$ ).

$$R^2 = \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (12)$$



- Nash-Sutcliffe efficiency (NSE) or MSE skill [65].

$$NSE = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (13)$$

- Willmott's index (WI) represents the ratio of the mean square error and the potential error [66].

$$WI = 1 - \left[ \frac{\sum_{i=1}^n (p_i - o_i)^2}{(\sum_{i=1}^n |p_i - o_i|)^2} \right] \quad (14)$$

Among one absolute error index statistic most used are

- Mean Error(MSE) estimates the average estimate error.

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \quad (15)$$

- Mean absolute error (MAE).

$$MAE = \frac{1}{N} \sum_{i=1}^n |p_i - o_i| \quad (16)$$

In all the formulas presented above,  $o_i, p_i$  represent the observed and estimated values,  $n$  is the number of records, and  $\bar{o}, \bar{p}$  indicate the means of the observed and predicted values, respectively.

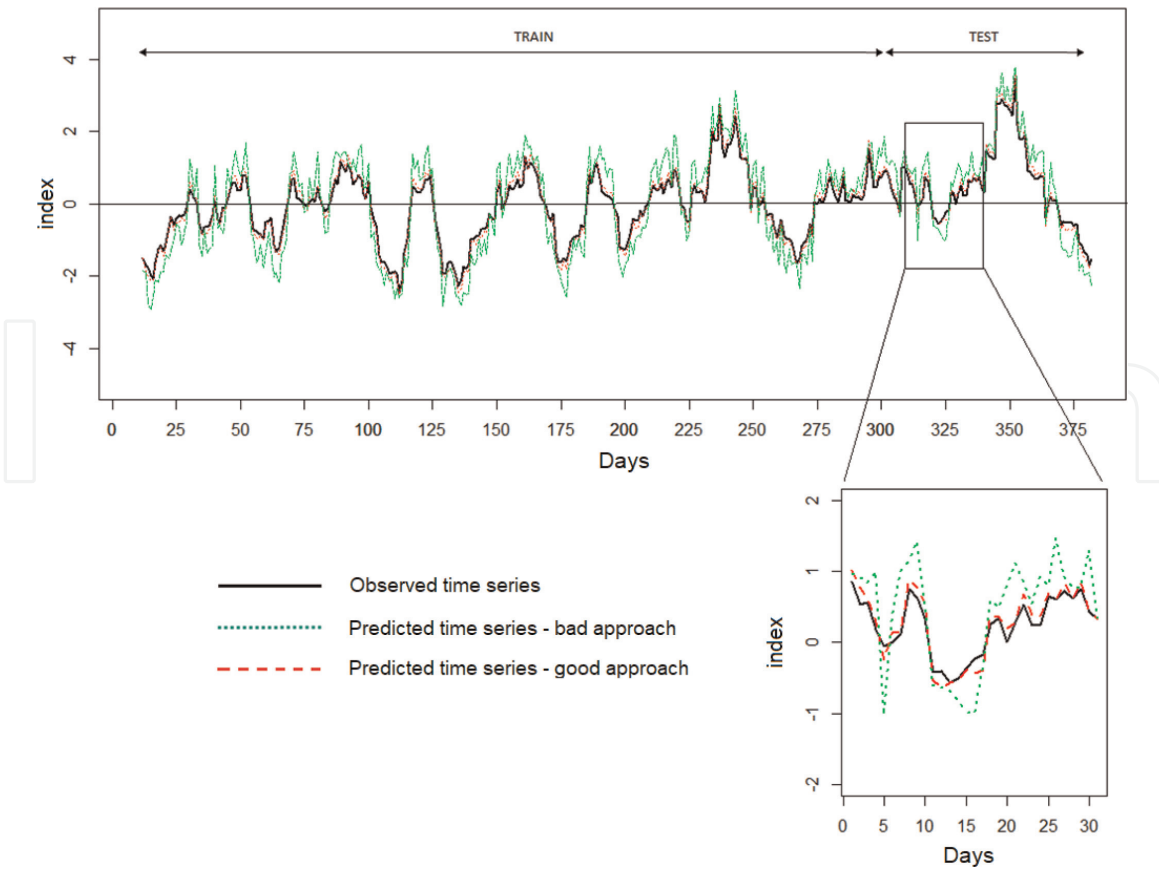
Here we included R and R2, two standard regression criteria, in the group of dimensionless statistics.

Finally, we present just one example of the graphical technique, mainly to show how a training and evaluation process is executed with a ML algorithm (**Figure 8**).

#### 4.1 Case studies

Some inconsistencies in the observations and the duration of satellite records introduce difficulties and uncertainties when applying forecast methods. At least 30 years of data record are required to SPI forecast; therefore, some of the examples we present here are based exclusively on ground gauge data. This situation is very close to reverting since satellite observations are reaching the minimum number of years required and the data are calibrated with ground observations (**Table 1**).

Shirmohammadi et al. [67] evaluated the performance of two ANN architectures (feedforward neural network and Elman or recurrent neural network), different kinds of ANFIS (four different membership functions: Gaussian, bell-shaped, triangular, and Piduetoits shape), WT-ANFIS, and WT-ANN. The wavelets families used here were db4, bior1.1, bior1.5, rboi1.1, rboi1.5, coif2, and coif4.



**Figure 8.**

Generic example of time series forecasting using two different ML methods. The green dotted line indicates a “bad” forecast method. The red dashed line indicates an appropriate method for the data, that is, the curve is closer to the observed time series. Both methods were trained using 80% of the data and tested on the remaining 20%.

Training data came from 1952 to 1992 rain records from East Azerbaijan province (Iran). More than 1000 model structures were tested to predict SPI6 for 1, 2, and 3 months’ lead-time over the test period covering from 1992 to 2011.  $R^2$ , NSE, and RMSE evaluated the performance of the models.

ANFIS models provided more accurate predictions than ANN models, and the inclusion of WT could improve meteorological drought modeling: WT-ANFIS (best RMSE = 0.097), WT-ANN (best RMSE = 0.227), ANFIS (best RMSE = 0.089), and ANN (best RMSE = 1.81).

Belayneh et al. [68] used precipitation records (1970 to 2005) to generate SPI3 and SPI6 time series from 12 stations in the Awash River Basin of Ethiopia (that is, 12 x 2 independent time series). The forecast was performed with ANN (RMSNN trained with the Levenberg-Marquardt back propagation), SVR, and the coupled models: WA-ANN and WA-SVR. About 80% of the data was used for training, 10% for validation, and 10% for testing, and ARIMA forecasting was used as a benchmark [69]. Regarding wavelet decomposition, each time series was decomposed between one and nine levels, and the appropriate level was selected by comparing results among all decomposition levels. The results of all the methods were compared by RMSE, MAE, and  $R^2$ . Overall, the WA-ANN and WA-SVR models were effective in forecasting SPI3 although most WA-ANN models had more accurate estimates (1- or 3-month lead). The WA-ANN model seemed to be more effective in anticipating extreme SPI values (severe drought or heavy precipitation), whereas WA-SVR closely reflected the observed SPI trends but underestimated the extreme events.

Model	SPI3	SPI12	SPI24
SVR	0.54	0.84	0.89
BSVR	0.47	0.86	0.91
BS-SVR	0.62	0.93	0.92
ANN	0.64	0.89	0.93
BANN	0.55	0.87	0.92
BS-ANN	0.67	0.95	0.98
WBANN	0.64	0.87	0.93
WBS-ANN	0.69	0.90	0.95
WBSVR	0.57	0.85	0.90
WBS-SVR	0.67	0.95	0.94

Abbreviations: SVR, support vector regression; BSVR, bootstrap SVR ensemble; BS-SVR, boosting-SVR; ANN, artificial neural networks; BANN, bootstrap ANN ensemble; BS-ANN, boosting-ANN; WBANN, wavelet coupled bootstrap ANN ensemble; WBS-ANN, wavelet boosting-ANN; WBSVR, wavelet coupled bootstrap SVR ensemble; WBS-SVR, wavelet boosting-SVR.

**Table 1.**  
Coefficient of determination ( $R^2$ ) of 10 ML methods to predict 3, 12, and 24 months SPI. Extracted from [71].

For SPI3 (1 month lead time) forecast, the best results in terms of RMSE (0.407) and MAE (0.391) were obtained by the WA-ANN model at the Ziquala station, whereas in terms of  $R^2$  (0.881), the Ginchi station had the best WA-ANN model.

When the lead-time was raised to 3 months, WA-ANN remained the best model. One station (Bantu Liben) had the model with the lowest RMSE and MAE values (0.510 and 0.4941), whereas a second station (Sebeta) had the best results in terms of  $R^2$  (0.7304).

Regarding SPI6 forecasts, the WA-ANN and WA-SVR models provided the best SPI6 forecasts. Neither method was meaningfully better than the other. The predictions for SPI6 were significantly better than SPI3 predictions according to three performance measures. As the forecast lead time increased, the forecast accuracy of all the models declined. This drop was most evident in the ARIMA, ANN, and SVR models.

These results were similar to [70]. Authors used precipitation records (1970–2005) from 20 stations in the same basin of Ethiopia (three different sub-basins) to generate SPI3 and SPI12 series. ANN, SVR, and WA-ANN were evaluated for 1 and 6 months lead time prediction. The comparison was made using RMSE, MAE, and  $R^2$ . Forecasting of SPI 12, for all the models, had better performance results than predicting SPI 3, regardless of the lead time (best  $R^2 = 0.953$ , WA-ANN). The performance of all the models declined when the lead time increased.

Belayneh et al. [71] modeled ANN and SVR as in [68] to forecast SPI 3, SPI12, and SPI24, but they included bootstrap (BANN and BSVR), boosting (BS-ANN, BS-SVR), wavelet coupled bootstrap ensemble (WBANN and WBSVR), and wavelet coupled boosting (WBS-ANN and WBS-SVR) in the analysis.

In general, the performances of SVR and ANN were comparable, although ANN performance was slightly higher. The inclusion of wavelets improved both techniques (wavelet decomposition denoises the time series). All models were more effective at forecasting SPI12 and SPI24 than SPI3 (**Table 1**). All boosting ensemble models were developed in MATLAB (“fitensemble” function).

The WBS-ANN and WBS-SVR models provided better prediction results than all the other types of models evaluated.

Ali, Deo, et al. [43] evaluated the performance of three models (ANFIS, M5Tree, and MPMR) to forecast SPI3, SPI6, and SPI12 calculated from a 35-year rainfall data set (1981–2015) from three (3) stations in Pakistan. SPI data were partitioned into 70% (training) and 30% (testing) periods. M5Tree is a kind of decision tree with linear regression functions on the leaves [72], whereas MPMR stands for minimax probability machine regression [73] and was also applied to benchmark the ensemble-ANFIS model. Regarding SPI3 forecast, ANFIS ( $R = 0.889$  to  $0.946$ ) outperformed MPMR ( $R = 0.843$  to  $0.935$ ) and M5Tree ( $R = 0.831$  to  $0.916$ ). Similarly, SPI6 ANFIS ( $R = 0.968$  to  $0.974$ ) outperformed M5Tree ( $R = 0.950$  to  $0.967$ ) and MPMR ( $R = 0.952$  to  $0.970$ ). For SPI12, ANFIS ( $R = 0.987$  to  $0.993$ ) overcome M5Tree ( $R = 0.950$  to  $0.967$ ) and MPMR ( $R = 0.984$  to  $0.986$ ). The other statistics (e.g., RMSE, WI) corroborated the superior performance of ANFIS. Just as important, the ensemble-ANFIS model achieved the highest accuracy at the three stations when predicting moderate, severe, and extreme droughts.

Khosravi et al. [74] used rainfall data from the Tropical Rainfall Measuring Mission (TRMM) during 2000–2014 in the eastern district of Isfahan to generate 12-month SPI. The first 85% of the data was used to train a single-hidden layer feedforward ANN, an SVR with RBF kernel, an LS-SVR with RBF kernel, and an ANFIS method. Optimum values of SVR and LS-SVR were obtained by a grid search within the range of  $[10^{-3}, 10^{+3}]$  and  $[2^{-3}, 2^{+3}]$  for  $C$  and  $\gamma$  (SVR) or (10, 100 and 1000) for  $g$  and (1, 0.5 and 1) for  $\gamma$  (LS-SVR).

For SPI12, SVR achieved the highest accuracy (RMSE = 0.21), followed by LS-SVR (RMSE = 0.38), ANN (RMSE = 1.24), and ANFIS (RMSE = 1.36). The best ANN model consisted of three layers (input, hidden, and output) with 30, 8, and 1 neuron, respectively.

Chen et al. [75] evaluated RF and ARIMA to forecast SPI3 (short-term drought) with a 1-month lead time and SPI12 (long-term drought). Both models were developed based on data from 1966 to 1995 (four stations in China), and predictions (1 month or 6 months ahead) were made from 1996 to 2004. Overall, RF performed consistently better than ARIMA. Results also suggested that RF is more robust in predicting dry events. Finally, ARIMA lost the capacity to predict SPI12, whereas the accuracy of RF was less affected by the longer lead time.

Agana and Homaifar [76] developed a hybrid model using a denoised empirical mode decomposition [77] and DBN. The proposed method was applied to predict a standardized streamflow index (SSI) across the Colorado River basin (ten stations). The new model was compared with MLP and SVR in predicting SSI12 (1, 6, and 12 months lead time). DBN, SVR, and their hybrid versions displayed rather similar prediction errors. However, DBN and EMD-DBN outperformed all other models for two-step predictions at almost all stations. As in wavelets, the empirical mode decomposition significantly improves the quality of prediction.

Finally, we want to mention two examples where ML was directly applied to rainfall prediction.

El Shafie et al. [77] evaluated a radial basis function neural network (RBFNN) to forecast rainfall in Alexandria City, Egypt. The model was trained using rainfall data from 1960 to 2001 (four stations) and tested with data from 2002 to 2009 to predict yearly and monthly (January and December) precipitation. Regarding yearly model efficiency,  $R^2 = 0.94$  for RBFNN, whereas the control (a multiple linear regression MR model) only reached  $R^2 = 0.21$ . Regarding monthly precipitation, RBFNN was very successful ( $R^2 = 0.899$  for January and  $R^2 = 0.997$  for December) as compared to the control ( $R^2 = 0.997$  and  $0.34$ , respectively).

Sumi et al. [78] compared ANNs, multivariate adaptive regression splines or MARS [79], k-nearest neighbor [80], and SVR with RBF kernel to predict daily and monthly rainfall in Fukuoka, Japan. A preprocessed training set (1975–2004) was



used to train the algorithms with extensive parameter optimization, whereas the test set covered from 2005 to 2009. For monthly rainfall, SVR produced the most accurate forecast (lowest RMSE) and the best rainfall mapping ( $R^2 = 0.93$ ), whereas for the daily rainfall series, the MARS method produced the best  $R^2$  value (0.99). All the metrics were calculated based on single-step ahead forecasting.

## 5. Satellite precipitation products (SPPs)

There is no satellite that can reliably quantify rainfall under all circumstances. However, ground observations, although reliable and with long-term records, do not provide a consistent spatial representation of precipitation, particularly on certain world regions. Therefore, satellite data become necessary, as they provide more homogeneous data quality compared to ground observations [81, 82]. To our knowledge, merged satellite-gauge products are becoming indispensable.

Precipitation data sets may be classified into one of four categories: gauge data sets (e.g., CRU TS [83], APHRODITE [84]), satellite-exclusive (e.g., CHOMPS [85]), merged satellite-gauge products (e.g., GPCP [86], TRMM3B42), and reanalysis (e.g., NCEP1/NCEP2 [87], ERA-Interim [88]). Reanalysis implies integrating irregular observations with models encompassing physical and dynamic processes in order to generate an estimate of the state of the system across a uniform grid and with temporal continuity [89].

Many studies show that satellite precipitation algorithms show different biases, detection probabilities, and missing rainfall ratios in summer and winter. Sources of error include the satellite sensor itself, the retrieval error [90], and spatial and temporal sampling [91, 92].

Algorithms that estimate rainfall from satellite observations are based on either thermal infrared (TIR) bands (inferring cloud-top temperature), passive microwave sensors (PMW), or active microwave sensors (AMW). The TIR-based approach takes into account cold cloud duration or CCD, that is, the time that a cloud has a temperature below the threshold at a given pixel [93]. The PMW-based approach takes advantage of the fact that microwaves can penetrate clouds to explore their internal properties through the interaction of raindrops [94]. AMW is what usually known as precipitation radar [95].

There is a plethora of validation studies of satellite-based rainfall estimates (SREs). Normally, these SREs are compared against ground rainfall estimates [91, 96].

Sun et al. [97] reviewed 30 currently available global precipitation (gauge-based, satellite-related, or reanalysis) data sets. The degree of variability of the precipitation estimates varies by region. Large differences in annual and seasonal estimates were found in tropical oceans, complex mountain areas, Northern Africa, and some high-latitude regions. Systematic errors are the main sources of errors over large parts of Africa, northern South America, and Greenland. Random errors are the dominant kinds of error in large regions of global land, especially at high latitudes. Regarding satellite assessments, PERSIANN-CCS has larger systematic errors than CMORPH, TRMM 3B42, and PERSIANN-CDR. The spatial distribution of systematic errors is similar for all reanalysis products [97].

**Table 2** presents a comparison of several representative satellite rainfall products. More information regarding these and other products can be found in [97, 98].

Abbreviations: (IR) infrared satellite imagery, (MW) microwave estimates, (GG) ground gauges, (AMSU) Advanced Microwave Sounding Unit, (AMSU-B) Advanced Microwave Sounding Unit-B, (SSM/I) Special Sensor Microwave/Imager; (AMSR-E) Advanced Microwave Scanning Radiometer for the Earth Observing



Product	Spatial coverage	Temporal coverage	Inputs	Access
CHIRPSv2.0 Funk et al. [99]	0.05° × 0.05° 50°S–50°N	Daily, pentadal and monthly 1981 to near present	IR + GG	<a href="http://chg.geog.ucsb.edu/data/">http://chg.geog.ucsb.edu/data/</a>
PERSIANN Nguyen et al. [100]	0.25° × 0.25° 60°S–60°N	1,3,6 hours March 2000 to present	IR (GOES-8, GOES-10, GMS-5, Metsat-6, and Metsat-7) corrected with MW(DMSP 7, 8, and 9 and NOAA-15, 16, and 17)	<a href="https://chrsdata.eng.uci.edu/">https://chrsdata.eng.uci.edu/</a>
PERSIANN- CDR Ashouri et al. [101]	0.25° × 0.25° 60°S–60°N	Daily, monthly 1983 to 2017	IR (GRIDSAT-B1) +GPCP correction	<a href="http://chrsdata.eng.uci.edu/">http://chrsdata.eng.uci.edu/</a>
CMORPH Joyce et al. [102]	0.25° × 0.25° 60°S–60°N	30 min 2002 to present	SSM/I (DMSP 13, 14, and 15) AMSU-B (NOAA-15, 16, 17, and 18) AMSU-E (Aqua), TMI (TRMM) Geostationary satellite IR	<a href="http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html">http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html</a>
RFE2.0 Xie and Arkin [103]	0.1° × 0.1° 40°N–40°S 20°W–55°E	daily 2001 to present	MW (SSM/I, AMSU-B) IR (GTS)	<a href="https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/">https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/</a>
TRMM3B42 Huffman et al. [104]	0.25° × 0.25° 50°N–50°S	3 hourly/ daily 1998 to 2015.	MW (TRMM, SSM/I, AMSR, AMSU), IR	<a href="https://pmm.nasa.gov/data-access/downloads/trmm">https://pmm.nasa.gov/data-access/downloads/trmm</a>

**Table 2.**  
*Representative satellite rainfall products.*

System, (MHS) Microwave Humidity Sounder, (GPCP) Global Precipitation Climatology Centre, (GOES) Geostationary Operational Environmental Satellite, (Metsat) meteorological satellite, (NOAA) NASA-provided TIROS series of weather forecasting satellite run by the National Oceanic and Atmospheric Administration, (DMSP) Defense Meteorological Satellite Program, (GRIDSAT-B1) geostationary IR channel brightness temperature.

6. Accessing and processing the data

The capacity to acquire information from remote sensing data has been improved to an unprecedented level, accumulating overwhelming amounts of information. For example, the Google Earth Engine (GEE) [105] is updated at a rate of nearly 6000 scenes per day from active missions (a typical image 10 km by 10 km requires 50–200 million bytes of memory). Such a large amount of data requires not only vast amounts of memory data but also higher-level services with high-performance computing systems [106]. Successful experiences have already been recorded [97], but the GEE is worth mentioning [105]. GGE stores a multi-petabyte catalog of satellite imagery and geospatial data sets collected from different resources and provides high-performance computing systems that can be

accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE). It also possess a library with more than 800 functions, ranging from simple mathematical functions to powerful geostatistical, machine learning, and image processing operations [105].

In many situations, intense computing resources are required for image processing operations [105], but more friendly solutions can be suggested to those who do have not the necessary skills.

As mentioned before, many satellite precipitation products are freely available (**Table 1**). Most of them are in network Common Data Form (netCDF) format [95]. R users can access this format using the “ncdf4” [96] or “raster” [97] packages. These data were already processed and can be used to forecast and perform complementary analyses [98]. We have already mentioned the SPEI [59] or the SPI [60] packages used to generate, for example, the SPI index.

Regarding the ML methods discussed here, almost all of them are available in packages deposited at the CRAN or CRAN-like repositories, for example, “Random Forest” package [43], “rminer” [99] that implements ANN, SVR and boosting [99], etc. A full list of packages implementing ML algorithms is available at <https://cran.r-project.org/web/views/MachineLearning.html>.

Finally, also available at the repositories are plenty of packages that are really helpful for visualizing and interpreting the results [107, 108].

## 7. Conclusion

Climate change is shifting global rainfall patterns and will increase the intensity and duration of drought around the world; this produces the need to take contingency actions to prevent the impact of famine. ML models, an evolving research area, are a valuable complement to methods previously proposed for forecasting drought. Results obtained so far for predicting meteorological indices are very satisfactory, especially with hybrid models such as WT-ANN or WT-SVR.

Most of the work that we reported here is based on the standardized precipitation index or SPI, which is a reliable measure of drought used in more than 60 countries. The leading month or the number of months over which SPI is calculated significantly influences the prediction values.

Unfortunately, many of the examples were based on ground gauge data. The brevity (and noise) of the records obstructs the use of many satellite products. However, as time progresses and data retrieval improves, satellite products will be long and accurate enough to generate reliable results.

The exponential growth of public and free satellite imagery sources and of open-source software, as well as cheaper access to cloud-based technology, will provide powerful forecasting tools to a greater number of researchers, allowing them to forecast drought before it occurs.

## Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

IntechOpen

IntechOpen

### **Author details**

Leonardo Ornella, Gideon Kruseman and Jose Crossa\*  
International Maize and Wheat Improvement Center (CIMMYT), Apdo,  
Mexico City, Mexico

\*Address all correspondence to: [j.crossa@cgiar.org](mailto:j.crossa@cgiar.org)

### **IntechOpen**

---

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] van Dijk AI, Beck HE, Crosbie RS, de Jeu RAM, Liu YY, Podger GM, et al. The millennium drought in Southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*. 2013;**49**:1040-1057. DOI: 10.1002/wrcr.20123
- [2] Keshavarz M, Karami E, Vanclay F. The social experience of drought in rural Iran. *Land Use Policy*. 2013;**30**:120-129. DOI: 10.1016/j.landusepol.2012.03.003
- [3] Harris C. Heat, Hardship and Horrible Harvests: Europe's Drought Explained. Euronews [Internet]. 2018. Available from: <https://www.euronews.com/2018/08/10/explained-europe-s-devastating-drought-and-the-countries-worst-hit>
- [4] Hao Z, AghaKouchak A, Nakhjiri N, Farahmand A. Global integrated drought monitoring and prediction system. *Scientific Data*. 2014;**1**:140001. DOI: 10.1038/sdata.2014.1
- [5] Wilhite DA, Svoboda MD, Hayes MJ. Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resources Management*. 2007;**21**: 763-774. DOI: 10.1007/s11269-006-9076-5
- [6] Leng G, Huang M. Crop yield response to climate change varies with crop spatial distribution pattern. *Scientific Reports*. 2017;**7**:1463-1463. DOI: 10.1038/s41598-017-01599-2
- [7] Gobin A. Modelling climate impacts on crop yields in Belgium. *Climate research (Open Access for articles 4 years old and older)*. 2010;**44**:55-68. DOI: 10.3354/cr00925
- [8] Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, et al. Stationarity is dead: Whither water management? *Science*. 2008;**319**:573-574. DOI: 10.1126/science.1151915 %JScience
- [9] Mishra AK, Singh VP. Drought modeling – A review. *Journal of Hydrology*. 2011;**403**:157-175. DOI: 10.1016/j.jhydrol.2011.03.049
- [10] Vereecken H, Huisman JA, Pachepsky Y, Montzka C, van der Kruk J, Bogaen H, et al. On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology*. 2014;**516**:76-96. DOI: 10.1016/j.jhydrol.2013.11.061
- [11] Maity R, Suman M, Verma N. Drought prediction using a wavelet based approach to model the temporal consequences of different types of droughts. *Journal of Hydrology*. 2016; **539**:417-428
- [12] Moyé LA, Kapadia AS, Cech IM, Hardy RJ. The theory of runs with applications to drought prediction. *Journal of Hydrology*. 1988;**103**:127-137. DOI: 10.1016/0022-1694(88)90010-8
- [13] Paulo AA, Pereira LS. Prediction of SPI drought class transitions using Markov chains. *Water Resources Management*. 2007;**21**:1813-1827. DOI: 10.1007/s11269-006-9129-9
- [14] Moreira EE, Coelho CA, Paulo AA, Pereira LS, Mexia JT. SPI-based drought category prediction using loglinear models. *Journal of Hydrology*. 2008;**354**: 116-130. DOI: 10.1016/j.jhydrol.2008.03.002
- [15] Loaiciga H. On the probability of droughts: The compound renewal model. *Water Resources Research*. 2005;**41**. DOI: 10.1029/2004WR003075
- [16] Dzupire NC, Ngare P, Odongo L. A Poisson-gamma model for zero inflated

- rainfall data. *Journal of Probability and Statistics*. 2018;**2018**:1-12. DOI: 10.1155/2018/1012647
- [17] Macesic N, Polubriaginof F, Tatonetti NP. Machine learning: Novel bioinformatics approaches for combating antimicrobial resistance. *Current Opinion in Infectious Diseases*. 2017;**30**:511-517. DOI: 10.1097/qco.0000000000000406
- [18] González-Camacho JM, Ornella L, Pérez-Rodríguez P, Gianola D, Dreisigacker S, Crossa J. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *The Plant Genome*. 2018;**11**. DOI: 10.3835/plantgenome2017.11.0104
- [19] Athey S. The Impact of Machine Learning on Economics. In: Agrawal A, Gans J, Goldfarb A, editors. *The Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press; 2018
- [20] Lesiv M, Schepaschenko D, Moltchanova E, Bun R, Dürauer M, Prishchepov AV, et al. Spatial distribution of arable and abandoned land across former Soviet Union countries. *Scientific Data*. 2018;**5**:180056. DOI: 10.1038/sdata.2018.56
- [21] Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*. 2018;**39**:2784-2817. DOI: 10.1080/01431161.2018.1433343
- [22] Li Y, Zhang H, Xue X, Jiang Y, Shen Q. Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2018;**8**:e1264. DOI: 10.1002/widm.1264
- [23] Berry T, Harlim J. Correcting biased observation model error in data assimilation. *Monthly Weather Review*. 2017;**145**:2833-2853. DOI: 10.1175/mwr-d-16-0428.1
- [24] Abbot J, Marohasy J. The application of machine learning for evaluating anthropogenic versus natural climate change. *GeoResJ*. 2017;**14**:36-46. DOI: 10.1016/j.grj.2017.08.001
- [25] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;**20**:273-297. DOI: 10.1023/a:1022627411411
- [26] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*. 2004;**14**:199-222. DOI: 10.1023/b:Stco.0000035301.49549.88
- [27] Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. *Least Squares Support Vector Machines*. Singapore: World Scientific; 2002
- [28] Balabin RM, Lomakina-Rumyantseva E. Support vector machine regression (SVR/LS-SVM) - an alternative to neural networks (ANN) for analytical chemistry- comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *The Analyst*. 2011;**136**:1703-1712. DOI: 10.1039/c0an00387e
- [29] Karasuyama M, Nakano R, editors. *Optimizing SVR Hyperparameters via fast cross-validation using AOSVR*. In: 2007 International Joint Conference on Neural Networks. 2007
- [30] Wang L. *Support Vector Machines: Theory and Applications*. Berlin Heidelberg: Springer-Verlag; 2005
- [31] Sheela KG, Deepa SN. Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering*. 2013;**2013**:11. DOI: 10.1155/2013/425740
- [32] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Computation*. 2006;**18**:



1527-1554. DOI: 10.1162/neco.2006.18.7.1527

[33] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006; **313**:504-507. DOI: 10.1126/science.1127647

[34] Hinton GE. A practical guide to training restricted Boltzmann machines. In: Montavon G, Orr GB, Müller K-R, editors. *Neural Networks: Tricks of the Trade*. 2nd ed. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp. 599-619

[35] Breiman L. Bagging Predictors. *Machine Learning*. Vol. 241996. pp. 123-140. DOI: 10.1023/a:1018054314350

[36] Efron B. Better bootstrap confidence intervals. *Journal of the American Statistical Association*. 1987; **82**:171-185. DOI: 10.1080/01621459.1987.10478410

[37] Awajan AM, Ismail MT, AL Wadi S. Improving forecasting accuracy for stock market data using EMD-HW bagging. *PLoS One*. 2018; **13**. DOI: 10.1371/journal.pone.0199582

[38] Breiman L. Random Forests. *Machine Learning*. Vol. 452001. pp. 5-32. DOI: 10.1023/a:1010933404324

[39] Mayumi Oshiro T, Santoro Perez P, Baranauskas J. How Many Trees in a Random Forest? Lecture notes in computer science 7376 MLDM'12 Proceedings of the 8th international conference on Machine Learning and Data Mining in Pattern Recognition; 2012. pp. 154-168

[40] Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002; **2**:18-22

[41] Jang J. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE*

*Transactions on Systems, Man, and Cybernetics*. 1993; **23**:665-685. DOI: 10.1109/21.256541

[42] Cheng C-T, Lin J-Y, Sun Y-G, Chau K. Long-Term Prediction of Discharges in Manwan Hydropower Using Adaptive-Network-Based Fuzzy Inference Systems Models. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005. pp. 1152-1161

[43] Ali M, Deo R, Downs N, Maraseni T. An ensemble-ANFIS based uncertainty assessment model for forecasting multi-scalar standardized precipitation index. *Atmospheric Research*. 2018; **207**:155-180. DOI: 10.1016/j.atmosres.2018.02.024

[44] Şahin M, Rızvan E. Prediction of attendance demand in European football games: Comparison of ANFIS, fuzzy logic, and ANN. *Computational Intelligence and Neuroscience*. 2018; **2018**:14. DOI: 10.1155/2018/5714872

[45] Schapire RE. The boosting approach to machine learning: An overview. In: Denison DD, Hansen MH, Holmes CC, Mallick B, Yu B, editors. *Nonlinear Estimation and Classification*. New York, NY: Springer New York; 2003. pp. 149-171

[46] Friedman JH. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*. 2001; **29**:1189-1232

[47] Khashei M, Bijari M. A new class of hybrid models for time series forecasting. *Expert Systems with Applications*. 2012; **39**:4344-4357. DOI: 10.1016/j.eswa.2011.09.157

[48] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003; **50**:159-175. DOI: 10.1016/S0925-2312(01)00702-0

[49] Sanghani A, Bhatt N, Chauhan NC, editors. *A Novel Hybrid Method for*

- Time Series Forecasting Using Soft Computing Approach. Cham: Springer International Publishing; 2019
- [50] Addison PS. The Illustrated Wavelet Transform Handbook. Introductory Theory and Applications in Science, Engineering, Medicine and Finance. Boca Raton, FL: CRC Press; 2017
- [51] Najmi A-H, Sadowsky J, Morlet O, Transform W. The continuous wavelet transform and variable resolution time-frequency analysis. Johns Hopkins APL Technical Digest (Applied Physics Laboratory). 1997;**18**:134-139
- [52] Sundararajan D. Discrete Wavelet Transform: A Signal Processing Approach. First ed. John Wiley & Sons: Singapore Pte Ltd.; 2015
- [53] Mishra A, Singh V. A review of drought concepts. Journal of Hydrology. 2010;**391**:202-216. DOI: 10.1016/j.jhydrol.2010.07.012
- [54] Heim Jr RR. A review of twentieth-century drought indices used in the United States. Bulletin of the American Meteorological Society. 2002;**83**: 1149-1166. DOI: 10.1175/1520-0477-83.8.1149
- [55] World Meteorological Organization (WMO) and Global Water Partnership (GWP) Integrated Drought Management Programme (IDMP). Handbook of Drought Indicators and Indices. Svoboda M, Fuchs BA, editors 2016
- [56] McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration of time scales. In: 8th Conference of Applied Climatology. CA, USA: Anaheim; 1993. pp. 179-184
- [57] Hayes M, Svoboda M, Wall N, Widhalm M. The Lincoln declaration on drought indices: Universal meteorological drought index recommended. Bulletin of the American Meteorological Society. 2011;**92**:485-488. DOI: 10.1175/2010BAMS3103.1
- [58] R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013
- [59] Beguería S, Vicente-Serrano SM, Reig F, Latorre B. Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. International Journal of Climatology. 2014;**34**: 3001-3023. DOI: 10.1002/joc.3887
- [60] Neves J. Package ‘spi’. Compute the SPI index using R.; 2011
- [61] Byun H-R, Wilhite DA. Objective quantification of drought severity and duration. Journal of Climate. 1999;**12**: 2747-2756. DOI: 10.1175/1520-0442(1999)012<2747:Oqodsa>2.0.Co;2
- [62] Bazrafshan J, Khalili A. Spatial analysis of meteorological drought in Iran from 1965 to 2003. Desert. 2013;**18**: 63-71
- [63] Tian Y, Nearing GS, Peters-Lidard CD, Harrison KW, Tang L. Performance metrics, error modeling, and uncertainty quantification. Monthly Weather Review. 2016;**144**:607-613. DOI: 10.1175/mwr-d-15-0087.1
- [64] Waseem M, Mani N, Andiego G, Usman M. A review of criteria of fit for hydrological models. International Research Journal of Engineering and Technology (IRJET). 2017;**4**:1765-1772
- [65] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology. 1970;**10**:282-290. DOI: 10.1016/0022-1694(70)90255-6

- [66] Willmott CJ. On the validation of models. *Physical Geography*. 1981;2: 184-194. DOI: 10.1080/02723646.1981.10642213
- [67] Shirmohammadi B, Moradi H, Moosavi V, Semiromi M, Zeinali A. Forecasting of meteorological drought using wavelet-ANFIS hybrid model for different time steps (case study: Southeastern part of East Azerbaijan province, Iran). *Natural Hazards*. 2013; **69**:389-402. DOI: 10.1007/s11069-013-0716-9
- [68] Belayneh A, Adamowski J, Khalil B. Short-term SPI drought forecasting in the Awash River basin in Ethiopia using wavelet transforms and machine learning methods. *Sustainable Water Resources Management*. 2016;2: 87-101. DOI: 10.1007/s40899-015-0040-5
- [69] Mishra AK, Desai VR, Singh VP. Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*. 2007;12:626-638. DOI: 10.1061/(ASCE)1084-0699(2007)12:6(626)
- [70] Belayneh A, Adamowski J. Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Applied Computational Intelligence and Soft Computing*. 2012; **2012**:1-13. DOI: 10.1155/2012/794061
- [71] Belayneh A, Adamowski J, Khalil B, Quilty J. Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmospheric Research*. 2016;172:37-47. DOI: 10.1016/j.atmosres.2015.12.017
- [72] Frank E, Wang Y, Inglis S, Holmes G, Witten IH. Using model trees for classification. *Machine Learning*. 1998; **32**:63-76. DOI: 10.1023/a:1007421302149
- [73] Strohmann T, Grudic G. A formulation for minimax probability machine regression. In: *NIPS'02 Proceedings of the 15th International Conference on Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press; 2003. pp. 785-792
- [74] Khosravi I, Jouybari-Moghaddam Y, Sarajian MR. The comparison of NN, SVR, LSSVR and ANFIS at modeling meteorological and remotely sensed drought indices over the eastern district of Isfahan, Iran. *Natural Hazards*. 2017; **87**:1507-1522. DOI: 10.1007/s11069-017-2827-1
- [75] Chen J, Li M, Wang W. Statistical uncertainty estimation using random forests and its application to drought forecast. *Mathematical Problems in Engineering*. 2012;2012:1-12. DOI: 10.1155/2012/915053
- [76] Agana AN, Homaifar A. EMD-based predictive deep belief network for time series prediction: An application to drought forecasting. *Hydrology*. 2018;5. DOI: 10.3390/hydrology5010018
- [77] El Shafie AH, El-Shafie A, Almukhtar A, Taha M, El Mazoghi HG, Abou Kheira A. Radial basis function neural networks for reliably forecasting rainfall. *Journal of Water and Climate Change*. 2012;3:125
- [78] Sumi SM, Zaman MF, Hirose H. A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. *International Journal of Applied Mathematics and Computer Science*. 2012;22:841-854. DOI: 10.2478/v10006-012-0062-1
- [79] Friedman JH. Multivariate adaptive regression splines. *Annals of Statistics*. 1991;19:1-67
- [80] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical*



Learning: Data Mining, Inference, and Prediction. New York: Springer; 2009

[81] Javanmard S, Yatagai A, Nodzu M, Bodaghjamali J, Kawamoto H. Comparing high-resolution gridded precipitation data with satellite rainfall estimates of TRMM\_3B42 over Iran. *Advances in Geosciences*. 2010;**25**: 119-125. DOI: 10.5194/adgeo-25-119-2010

[82] Dembélé M, Zwart SJ. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *International Journal of Remote Sensing*. 2016;**37**: 3995-4014. DOI: 10.1080/01431161.2016.1207258

[83] Harris I, Jones PD, Osborn TJ, Lister DH. Updated high-resolution grids of monthly climatic observations – The CRU TS3.10 dataset. *International Journal of Climatology*. 2014;**34**: 623-642. DOI: 10.1002/joc.3711

[84] Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A. APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*. 2012;**93**:1401-1415. DOI: 10.1175/bams-d-11-00122.1

[85] Joseph R, Smith TM, Sapiano MRP, Ferraro RR. A new high-resolution satellite-derived precipitation dataset for climate studies. *Journal of Hydrometeorology*. 2009;**10**:935-952. DOI: 10.1175/2009jhm1096.1

[86] Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P-P, Janowiak J, et al. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*. 2003;**4**: 1147-1167. DOI: 10.1175/1525-7541(2003)004<1147:Tvgpcp>2.0.Co;2

[87] Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, Fiorino M, et al.

NCEP–DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*. 2002;**83**:1631-1644. DOI: 10.1175/bams-83-11-1631

[88] Berrisford P, Dee DP, Poli P, Brugge R, Mark F, Manuel F, et al. The ERA-Interim Archive Version 2.0. Shinfield Park, Reading: ECMWF; 2011

[89] Jian-Jian F, Shuanglin L. Intercomparison of the south Asian high in NCEP1, NCEP2, and ERA-40 reanalyses and in station observations. *Atmospheric and Oceanic Science Letters*. 2012;**5**:189-194. DOI: 10.1080/16742834.2012.11446989

[90] Maggioni V, Nikolopoulos EI, Anagnostou EN, Borga M. Modeling satellite precipitation errors over mountainous terrain: The influence of gauge density, seasonality, and temporal resolution. *IEEE Transactions on Geoscience and Remote Sensing*. 2017; **55**:4130-4140. DOI: 10.1109/TGRS.2017.2688998

[91] AghaKouchak A, Mehran A, Norouzi H, Behrangi A. Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*. 2012;**39**. DOI: 10.1029/2012GL051592

[92] Tian Y, Peters-Lidard CD, Eylander JB, Joyce RJ, Huffman GJ, Adler RF, et al. Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research-Atmospheres*. 2009;**114**. DOI: 10.1029/2009JD011949

[93] Domenikiotis C, Spiliotopoulos M, Galakou E, Dalezios N, editors. Assessment of the Cold Cloud Duration (Ccd) Methodology for Rainfall Estimation in Central Greece; 2003

[94] Laviola S, Levizzani V. Passive microwave remote sensing of rain from satellite sensors. In: Mukherjee M, editor. *Advanced Microwave and*

Millimeter Wave Technologies  
 Semiconductor Devices Circuits and  
 Systems. Rijeka: InTech; 2010.  
 pp. 549-572

[95] Kawanishi T, Kuroiwa H, Kojima M, Oikawa K, Kozu T, Kumagai H, et al. TRMM precipitation radar. *Advances in Space Research*. 2000;25:969-972. DOI: 10.1016/S0273-1177(99)00932-1

[96] Bell TL, Kundu PK. Comparing satellite rainfall estimates with rain gauge data: Optimal strategies suggested by a spectral model. *Journal of Geophysical Research-Atmospheres*. 2003;108. DOI: 10.1029/2002JD002641

[97] Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu K-L. A review of global precipitation data sets: Data sources, Estimation, and Intercomparison. 2018;56:79-107. DOI: 10.1002/2017RG000574

[98] Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk AIJM, Weedon GP, et al. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*. 2017;21:6201-6217. DOI: 10.5194/hess-21-6201-2017

[99] Funk C, Peterson P, Landsfeld MF, Pedreros D, Verdin JP, Rowland J, et al. A Quasi-Global Precipitation Time Series for Drought Monitoring. USGS professional paper 832; 2014

[100] Nguyen P, Shearer EJ, Tran H, Ombadi M, Hayatbini N, Palacios T, et al. The CHRS data portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Scientific Data*. 2019;6:180296. DOI: 10.1038/sdata.2018.296

[101] Ashouri H, Hsu K-L, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD, et al. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for

hydrological and climate studies. *Bulletin of the American Meteorological Society*. 2015;96:69-83. DOI: 10.1175/bams-d-13-00068.1

[102] Joyce RJ, Janowiak JE, Arkin PA, Xie P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*. 2004;5:487-503. DOI: 10.1175/1525-7541(2004)005<0487:Camtpg>2.0.Co;2

[103] Xie P, Arkin PA. Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *AMS*. 1996; 9:840-858. DOI: 10.1175/1520-0442(1996)009<0840:Aogmpu>2.0.Co;2

[104] Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, et al. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*. 2007;8: 38-55. DOI: 10.1175/jhm560.1

[105] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 2017;202: 18-27. DOI: 10.1016/j.rse.2017.06.031

[106] Lokers R, Knapen R, Janssen S, van Randen Y, Jansen J. Analysis of big data technologies for use in agro-environmental science. *Environmental Modelling and Software*. 2016;84: 494-504. DOI: 10.1016/j.envsoft.2016.07.017

[107] Rahlf T. *Data Visualisation with R*. Springer International Publishing; 2017. 385 p

[108] Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016. 213 p