



Improving the accuracy of air relative humidity prediction using hybrid machine learning based on empirical mode decomposition: a comparative study

Khaled Merabet¹ · Salim Heddami²

Received: 30 August 2022 / Accepted: 29 March 2023 / Published online: 12 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This paper proposes a hybrid air relative humidity prediction based on preprocessing signal decomposition. New modelling strategy was introduced based on the use of the empirical mode decomposition, variational mode decomposition, and the empirical wavelet transform, combined with standalone machine learning to increase their numerical performances. First, standalone models, i.e., extreme learning machine, multilayer perceptron neural network, and random forest regression, were used for predicting daily air relative humidity using various daily meteorological variables, i.e., maximal and minimal air temperatures, precipitation, solar radiation, and wind speed, measured at two meteorological stations located in Algeria. Second, meteorological variables are decomposed into several intrinsic mode functions and presented as new input variables to the hybrid models. The comparison between the models was achieved based on numerical and graphical indices, and obtained results demonstrate the superiority of the proposed hybrid models compared to the standalone models. Further analysis revealed that using standalone models, the best performances are obtained using the multilayer perceptron neural network with Pearson correlation coefficient, Nash–Sutcliffe efficiency, root-mean-square error, and mean absolute error of approximately ≈ 0.939 , ≈ 0.882 , ≈ 7.44 , and ≈ 5.62 at Constantine station, and ≈ 0.943 , ≈ 0.887 , ≈ 7.72 , and ≈ 5.93 at Sétif station, respectively. The hybrid models based on the empirical wavelet transform decomposition exhibited high performances with Pearson correlation coefficient, Nash–Sutcliffe efficiency, root-mean-square error, and mean absolute error of approximately ≈ 0.950 , ≈ 0.902 , ≈ 6.79 , and ≈ 5.24 , at Constantine station, and ≈ 0.955 , ≈ 0.912 , ≈ 6.82 , and ≈ 5.29 , at Sétif station. Finally, we show that the new hybrid approaches delivered high predictive accuracies of air relative humidity, and it was concluded that the contribution of the signal decomposition was demonstrated and justified.

Keywords Air relative humidity · Modelling · ELM · MLPNN · RFR · Signal decomposition

Responsible Editor: Philippe Garrigues

Highlights

- Predicting air relative humidity using three models, i.e., ELM, MLPNN, and RFR.
- Preprocessing signal decomposition for improving models performances.
- The EMD, VMD, and EWT were used and compared.
- Signal decomposition contributes to the models performances improvement.

✉ Khaled Merabet
kh.merabet@univ-skikda.dz; merabet.khaled@gmail.com

¹ Laboratory of Optimizing Agricultural Production in Subhumid Zones (LOPAZS), Faculty of Science, Agronomy Department, University 20 Août 1955-Skikda, Route El Hadaik, BP 26 Skikda, Algeria

Nomenclature

ANN	Artificial neural networks
DT	Decision trees
DTR	Decision tree regression
ELM	Extreme learning machine
EMD	Empirical mode decomposition
EWT	Empirical wavelet transform
ET_0	Reference evapotranspiration
GBT	Gradient boosting tree
IMFs	Intrinsic mode functions

² Laboratory of Research in Biodiversity Interaction Ecosystem and Biotechnology (LRIBEB), Faculty of Science, Agronomy Department, University 20 Août 1955-Skikda, Route El Hadaik, BP 26 Skikda, Algeria

KNN	K-nearest neighbor
LR	Linear regression
LSTM	Long-short term memory
M5Tree	M5 model tree
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression spline
MLPNN	Multilayer perceptron neural network
MLR	Multiple linear regression
MRA	Multiresolution analysis
NNARX	Nonlinear autoregressive with external input
NSE	Nash-Sutcliffe efficiency
OOB	Out-of-bag
P	Precipitation
Pa	Air pressure
RH	Air relative humidity
R	Correlation coefficient
R^2	Determination coefficient
RBFNN	Radial basis function neural network
RFR	Random forest regression
RMSE	Root-mean-square error
RNN	Recurrent neural networks
SARIMA	Seasonal autoregressive integrated moving average
SLFN	Single-layer feedforward neural network
SR	Solar radiation
SS	Sunshine hours
SVR	Support vector regression
T_{\max}	Maximum air temperature
T_{mean}	Mean temperature
T_{\min}	Minimum air temperature
U_2	Wind speed
VMD	Variational mode decomposition
WD	Wind direction
XGBoost	Extreme gradient boosting

Introduction

Air relative humidity (RH) can be defined as the quantity of water available in the air, and it is one of the most critical weather variables for hydrological and climatic studies, and it was also included in various climatic change investigations (Gunawardhana et al. 2017; Sein et al. 2022). One of the most important uses of RH in water resource management is its inclusion in the standard Penman–Monteith method for reference evapotranspiration (ET_0) calculation (Eccel 2012). Air temperature and RH are considered the most important factors of the thermal environment (Kuang 2020), and their accurate estimation is of great importance. The RH is a highly sensitive weather variable, and it affects several other processes especially the agro-food and the biological items, and it is highly influenced by several other weather variables

especially air temperature, precipitation, and solar radiation (Shrestha et al. 2019). During the last few years, the application of machine learning for water resources planning and management has received great importance and several application can be found in the literature, i.e., modelling pan evaporation (Kisi et al. 2022), predicting solar radiation in semi-arid regions (Jamei et al. 2023), modeling average grain velocity for rectangular channel (Kumari et al. 2022), and also for watershed prioritization (Sarkar et al. 2022). The RH data can be obtained from direct in situ measurement; however, alternatives methods based on modelling approaches can be a good alternative and several applications can be found in the literature.

Tao et al. (2022) used several weather variables collected at two meteorological stations in Irak, i.e., maximum air temperature (T_{\max}), minimum air temperature (T_{\min}), reference evapotranspiration (ET_0), sunshine hours (SS), and wind speed (U_2) for modelling monthly air relative humidity (RH). The authors used the extreme gradient boosting (XGBoost) algorithm for better input variables selection, and they compared the performances of three machines learning models, i.e., support vector regression (SVR), random forest regression (RFR), and multivariate adaptive regression spline (MARS), according to several input variables combinations. According to the obtained results, the RFR was the most accurate at *Kut* station, exhibiting root-mean-square error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE) values of approximately $\approx 4.92\%$, $\approx 3.89\%$, and 0.916, respectively, while at Mosul station, MARS model was the most accurate exhibiting RMSE, MAE, and NSE values of $\approx 3.80\%$, $\approx 2.86\%$, and 0.967, respectively. Yasar et al. (2012) compared between multilayer perceptron neural networks (MLPNN) and multiple linear regression (MLR) in modelling monthly RH in Turkey. For calibrating the models, they used five input variables, namely, monthly precipitation (P), latitude, longitude, altitude, and the month number. It was found that the MLPNN was more accurate exhibiting a correlation coefficient (R) values ranging from ≈ 0.96 to ≈ 0.99 , and from ≈ 0.73 to ≈ 0.94 for the MLPNN and MLR, respectively, while the mean absolute percentage error (MAPE) was ranged from ≈ 1.56 to $\approx 3.32\%$ and from 3.88 to 8.56%, respectively, showing the high contribution of the topographical information, i.e., latitude, longitude, and altitude in improving the forecasting accuracies of the monthly RH. Hanoon et al. (2021) compared between gradient boosting tree (GBT), RFR, linear regression (LR), MLPNN, and radial basis function neural network (RBFNN) in predicting daily and monthly RH in Malaysia. For developing the models, the authors used the RH measured at previous lags times from $(t-1)$ to $(t-6)$ without the inclusion of other climatic variables. From the obtained results, it was found that (i) at daily time scale, the MLPNN was more accurate

exhibiting R , RMSE, and MAE values of approximately ≈ 0.634 , $\approx 2.6\%$, and $\approx 3.9\%$, and (ii) at monthly time scale, the RBFNN was the most accurate with R , RMSE, and MAE values of approximately ≈ 0.713 , $\approx 1.6\%$, and $\approx 2\%$, respectively. Adnan et al. (2021) compared between MARS and M5Tree model for predicting daily RH using P , solar radiation (SR), and mean temperature (T_{mean}) measured at the Hunza River basin, Pakistan. Obtained results revealed that MARS model was more accurate compared to the M5Tree exhibiting RMSE, MAE, and coefficient of determination (R^2) ranging from ≈ 5.86 to $\approx 6.58\%$, ≈ 4.97 to $\approx 5.43\%$, and ≈ 0.806 to ≈ 0.815 , respectively, compared to the value of ≈ 6.08 to $\approx 6.19\%$, ≈ 5.46 to $\approx 5.58\%$, and ≈ 0.762 to ≈ 0.783 obtained using the M5Tree. In addition, they reported that the best forecasting accuracies were obtained using RH and T_{mean} measured at three previous lags as input variables. Shi et al. (2018) applied the MLPNN model for forecasting RH at 6 h, 24 h, and 72 h in advance. High forecasting accuracy was obtained with R^2 values ranging from ≈ 0.654 to ≈ 0.977 .

Thapliyal et al. (2014) used a linear regression model for RH retrieval showing good predictive accuracy with R -value of approximately ≈ 0.91 . Lu and Viljanen (2009) used the nonlinear autoregressive with external input (NNARX) model for predicting RH using T_{mean} and RH measured at previous lags times from $(t-1)$ to $(t-5)$ as input variables showing high predictive accuracy with R value of approximately ≈ 0.99 . Bregaglio et al. (2010) investigated the feasibility of thirteen hourly air relative humidity modelling solutions based on different empirical formula and showing high to moderate correlation between measured and calculated RH with R -values ranging from ≈ 0.660 to ≈ 0.858 . Hussein et al. (2021) compared between five machine learning models, namely, RFR, MLPNN, MLR, XGBoost, and the k -nearest neighbor (KNN) for predicting RH using monthly image data. The image data collected and used were available from NASA GESDISC data archive, and several climatic variables are available, namely, rainfall, evaporation, humidity, temperature, and wind speed. From the obtained results, it was found that the proposed models were able to accurately predict RH with R^2 ranging from ≈ 0.960 to ≈ 0.999 . Suradhaniwar et al. (2021) developed one-step and multi-step ahead forecasting frameworks for RH using suite of machine learning models. They applied the recurrent neural networks (RNN), the long-short term memory (LSTM), the support vector regression (SVR), the seasonal autoregressive integrated moving average (SARIMA), and the MLPNN models. Based on the RMSE values, it was found that the SVR and SARIMA models were more accurate than the MLPNN and the two deep learning models, i.e., the LSTM and RNN. Taking into account the RMSE values, it was found that the SARIMA (RMSE ≈ 1.87) and SVR (RMSE ≈ 1.97) outperformed the MLPNN (RMSE ≈ 2.83), LSTM (RMSE ≈ 2.12), and RNN (RMSE ≈ 2.13) for

one-step and multi-step ahead forecasting: SARIMA (RMSE $\approx 11.31\%$) and SVR (RMSE $\approx 11.30\%$), MLPNN (RMSE $\approx 18.11\%$), LSTM (RMSE $\approx 12.02\%$), and RNN (RMSE $\approx 14.74\%$), respectively. Qadeer et al. (2021) used RFR and SVR for modelling RH using two predictors, namely, dry-bulb temperature and wet-bulb temperature, and they reported slightly and negligible difference between the two models. Arulmozhi et al. (2021) selected a large number of predictors for modelling RH, namely, wind direction (WD), U_2 , T_{mean} , air pressure (Pa), P , SR , and net radiation, and they compared between MLR, MLPNN, RFR, SVR, and decision tree regression (DTR). From the obtained results, the RFR was found to be more accurate and outperformed all others models with R^2 , RMSE, and MAE of 0.954, 2.429%, and 1.470%, respectively.

According to the literature discussed above, it is clear that several attempts have been made for better prediction of RH based on machine learning models. The above listed models, i.e., SVR, MLPNN, DTR, and RNN, have been used in RH modelling studies, whereas many other methods have not. Furthermore, we can argue about the advantages and limitations of each model were governed by the type of data and varied from one region to another. Because air relative humidity and meteorological variables were characterized by linear and nonlinear properties, it become particularly challenging to directly build robust single machine learning models. In order to overcome the limitations of some single machine learning models, the hybridization based on pre-processing signal decomposition has become very popular lately because it can be easily used and their robustness has been demonstrated. Furthermore, to the best of the author's knowledge; no study has reported the application of the pre-processing signal decomposition for predicting RH using climatic variables as predictors. Consequently, there is still-room to investigate new modelling framework and to compare their performances relative to those of already reported in the literature, in an attempt to improve our understanding of the RH prediction. This study is the first to use a number of algorithms, i.e., the empirical mode decomposition (EMD), variational mode decomposition (VMD), and the empirical wavelet transform (EWT) for improving the predictive accuracy of the RH. In this study, we conduct also a comprehensive comparison of the performances of three different machine learning models (MLPNN, extreme learning machine (ELM), and RFR) with and without combination with the EMD, VMD, and EWT algorithms, which make the present study a deeply comparison between single and hybrid models.

Through these intercomparisons, the present investigation could be a sound argument for the judgment of the real benefit and the added utility of the signal decomposition in improving the retrieval of the RH. The paper is organized as follows. "Introduction" is an introduction with in depth literature

review. In "Materials and methods", we present the case study, data used, and the mathematical description of the proposed models. "Study area and data" is reserved to the presentation of the results and discussion. Finally, the conclusion is provided in "Performance assessment of the models".

Materials and methods

Study area and data

The present investigation was carried out in two sites areas in the East of Algeria, where meteorological information are available, as it is detailed in the following paragraphs. The selected two stations were the following (see Fig. 1): (i) Constantine station (Latitude: 36.374°, longitude: 6.562°, and altitude: 698 m) and (ii) Sétif (latitude: 36.374°, longitude: 5.312°, and altitude: 1094 m). For each station, we use data for a total period of 10 years ranging from 2000 to 2009, with total patterns of 3653. For each station, 70% of the data are used for model calibration (training) and 30% for model validation. The meteorological data sets used in the present study are composed from five predictors combined for better prediction of air relative humidity (RH). The selected variables are (i) solar radiation (*SR*: MJ/m²), (ii) the precipitation (*P*: mm), (iii) the wind speed (*U*₂: m/s), and (iv) the maximal and minimal air temperature (*T*_{max} and *T*_{min}: °C). The RH is measured in percent (RH: %). Descriptive statistics parameters for all variables are calculated and provided in Table 1 for the two stations, with the mean, maximal, minimal, standard deviation, coefficient of variation, and the coefficient of correlation calculated between each variable and the RH, which have help in selecting the best input combination, and in total six were selected having various input variables starting from five and ending by two (Table 2). Furthermore, all variables used in the present study were standardized using the Z-score method by subtracting the mean and dividing by the standard deviation.

Performance assessment of the models

In the present study, four performance metrics were selected for model comparison and evaluation: the root-mean-square error (RMSE), mean absolute error (MAE), correlation coefficient (*R*), and Nash–Sutcliffe efficiency (NSE).

$$MAE = \frac{\sum_{i=1}^N |RH_{pre,i} - RH_{obs,i}|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (RH_{obs,i} - RH_{pre,i})^2}{N}} \tag{2}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^N (RH_{obs,i} - RH_{pre,i})^2}{\sum_{i=1}^N (RH_{obs,i} - \overline{RH_{obs}})^2} \right] \tag{3}$$

$$R = \frac{\sum_{i=1}^N (RH_{obs,i} - \overline{RH_{obs}})(RH_{pre,i} - \overline{RH_{pre}})}{\sqrt{\sum_{i=1}^N (RH_{obs,i} - \overline{RH_{obs}})^2 \sum_{i=1}^N (RH_{pre,i} - \overline{RH_{pre}})^2}} \tag{4}$$

$\overline{RH_{obs}}$ and $\overline{RH_{pre}}$ are the mean measured, and mean forecasted air relative humidity, respectively; *RH*_{obs} and *RH*_{pre} specifies the observed and forecasted air relative humidity; and *N* shows the number of data points.

Machine learning models

ANN model

Artificial neural networks (ANN) are mathematical models biologically inspired from the function of the human brain. The ANN model is composed from an ensemble of units called neurons and arranged in a successive parallel layers. In the present paper, we use the multilayer perceptron neural network (MLPNN) model reported as universal approximator (Fig. 2). The basic element of the ANN model is called the neuron, and it can play a key dual role in providing the final response of the model: the summation, and the activation using an activation function. The summation is calculated based on the weighted sum of the inputs variables, while the activation is calculated using an activation function, generally the sigmoidal function (Eq. 5).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

The neurons from the input to the output layers are connected among them through an ensemble of parameters called weights similar to the biological neurons (Haykin 1999; Hornik 1991). The structure of the MLPNN model is determined based on the number of neurons in each layer and the total number of layers. For the input and the output layers, the total number of neurons corresponds exactly to the number of input and output variables, while the number of the neuron in the hidden layer is determined by trial and error. The success of the MLPNN comes from the back-propagation training algorithm, which was developed for improving the capability of the model in handling a non-linear function approximation. During the training process, the cost function calculated between the actual (i.e., measured) and the calculated outputs is minimized. In addition, during the training process, there is a continuous updating of all model parameters (i.e., weights and biases) in both

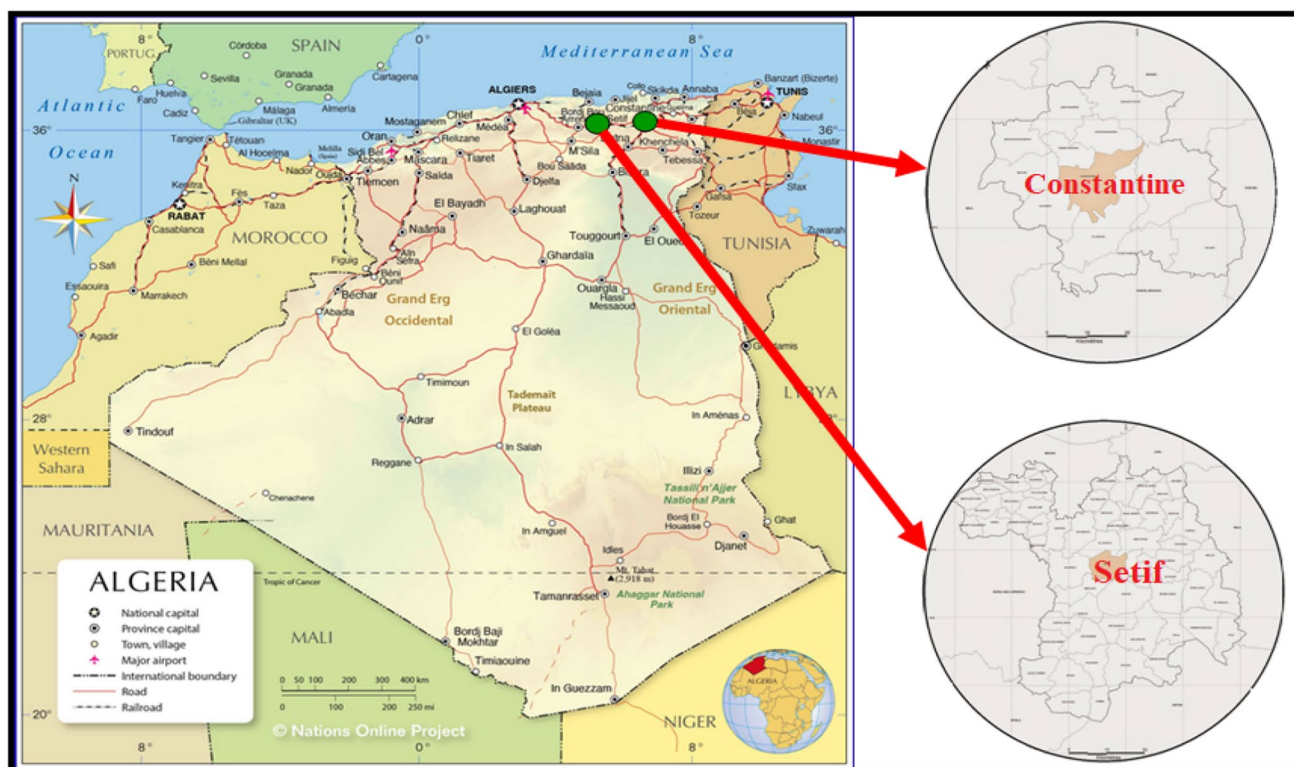


Fig. 1 Location map showing the two stations in the east of Algeria

directions: forward and backward propagation of the gradient. More detail about the ANN can be found in Haykin (1999) and Hornik (1991).

ELM

Suppose we have N observations with input variables x_i and its corresponding output y_i . We do not know any possible relation between x and y variables, and we suppose that the x_i can provide some useful information's for predicting y_i with an acceptable accuracy level. In order to solve the approximation function linking the input (x_i) to the output (y_i), Huang et al. (2006a, b) propose an extreme learning machine algorithm (ELM) for training the single layer feedforward neural network (SLFN) having three layers: input layer (x_i), one hidden layer, and one output layer (y_i) as shown in Fig. 3. Compared to the standalone ANN for which all parameters were updated during the training process, in the ELM model, the input weights and the hidden layer biases were “randomly” selected (i.e., the w_{ij} and b_j), and the output weights linking the hidden layer to the output layer were analytically determined (i.e., the β_j) using the Moore Penrose generalized inverse operation, making the training algorithm a simple linear system.

Suppose that the hidden layer is composed with L hidden neurons, it is assumed that there exist a series of models

parameters designated as: w_{ij} , b_j , β_j , and a nonlinear transformation function G for which the following formulation can be drawn:

$$f(x_n) = \sum_{j=1}^L \beta_j \cdot G(w_{ij} \cdot x_i + b_j) = H\beta \quad (6)$$

where f is the final response of the ELM model, the x_n are the input variables, β_j is the output weights matrix, w_{ij} is the weights linking the input to the hidden layers, b_j is the bias of the hidden layer neuron j , and G is the sigmoid activation function. H is called the hidden layer output matrix (Huang et al. 2006a, b).

RFR

Random forest regression (RFR) is an ensemble method composed of several decision trees models (DT) introduced by Breiman (2001). Each DT is constructed based on a recursive splitting strategy of the input training data (Fig. 4). It is important to note that for each root node, the calibration datasets are arranged into a unique partition, and each DT is induced by an out-of-bag (OOB) approach, which is a sampling with replacement. The OOB leads to two kind of data: a part will be “left out,” while the second

Table 1 Summary statistics of variables

Variables	Subset	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	R
Constantine weather station								
<i>RH</i>	Training	%	63.724	99.667	8.587	21.892	0.344	1.000
	Validation	%	64.048	99.618	10.170	21.656	0.338	1.000
	All data	%	63.821	99.667	8.581	21.819	0.342	1.000
T_{max}	Training	°C	22.858	44.976	1.215	9.735	0.426	-0.895
	Validation	°C	22.568	42.691	0.071	9.592	0.425	-0.884
	All data	°C	22.771	44.976	0.071	9.692	0.426	-0.892
T_{min}	Training	°C	9.700	27.201	-5.335	6.749	0.696	-0.742
	Validation	°C	9.541	28.207	-4.159	6.621	0.694	-0.735
	All data	°C	9.653	28.207	-5.335	6.711	0.695	-0.740
<i>p</i>	Training	mm	1.755	62.366	0.000	4.967	2.831	0.405
	Validation	mm	1.606	32.956	0.000	3.945	2.463	0.447
	All data	mm	1.710	62.366	0.000	4.686	2.741	0.414
U_2	Training	m/s	1.983	6.223	0.590	0.583	0.294	-0.114
	Validation	m/s	1.988	6.757	0.785	0.597	0.300	-0.198
	All data	m/s	1.984	6.757	0.590	0.587	0.296	-0.139
<i>SR</i>	Training	MJ/m ²	17.768	32.476	0.357	8.848	0.498	-0.695
	Validation	MJ/m ²	17.995	32.087	0.430	8.983	0.499	-0.676
	All data	MJ/m ²	17.836	32.476	0.357	8.888	0.498	-0.689
Sétif weather station								
<i>RH</i>	Training	%	64.701	99.665	10.235	23.427	0.362	1.000
	Validation	%	64.340	99.666	9.743	23.053	0.358	1.000
	All data	%	64.593	99.666	9.743	23.313	0.361	1.000
T_{max}	Training	°C	20.972	42.559	-2.737	9.983	0.476	-0.900
	Validation	°C	21.000	41.423	-1.274	9.773	0.465	-0.898
	All data	°C	20.980	42.559	-2.737	9.920	0.473	-0.900
T_{min}	Training	°C	8.436	26.185	-6.272	6.839	0.811	-0.743
	Validation	°C	8.294	25.945	-8.398	6.739	0.813	-0.739
	All data	°C	8.393	26.185	-8.398	6.809	0.811	-0.742
<i>p</i>	Training	mm	2.381	56.023	0.000	5.453	2.290	0.432
	Validation	mm	2.178	46.656	0.000	5.142	2.360	0.414
	All data	mm	2.321	56.023	0.000	5.362	2.311	0.427
U_2	Training	m/s	1.849	6.256	0.554	0.610	0.330	-0.083
	Validation	m/s	1.827	4.948	0.664	0.588	0.322	-0.107
	All data	m/s	1.842	6.256	0.554	0.603	0.327	-0.090
<i>SR</i>	Training	MJ/m ²	17.295	32.709	0.377	8.975	0.519	-0.717
	Validation	MJ/m ²	17.485	32.295	0.435	8.964	0.513	-0.702
	All data	MJ/m ²	17.352	32.709	0.377	8.971	0.517	-0.713

X_{mean} mean, X_{max} maximum, X_{min} minimum, S_x standard deviation, C_v coefficient of variation, R coefficient of correlation with RH

Table 2 The input combinations of different models

ELM	ANN	RFR	Input combination	Output
ELM1	MLPNN1	RFR1	$T_{max}, T_{min}, P, U_2, SR$	RH
ELM2	MLPNN2	RFR2	$T_{max}, T_{min}, U_2, SR$	RH
ELM3	MLPNN3	RFR3	T_{max}, P, U_2, SR	RH
ELM4	MLPNN4	RFR4	T_{max}, P, U_2	RH
ELM5	MLPNN5	RFR5	T_{max}, P, SR	RH
ELM6	MLPNN6	RFR6	T_{max}, P	RH

part will be repeated in the sample. By achieving the training of all DT, RFR model uses an average or majority vote depending on whether a classification or regression task is handled form all the predictions of the single trees.

Signal decomposition methods

In the present paper, three-signal decomposition were used, namely, the empirical mode decomposition (EMD),

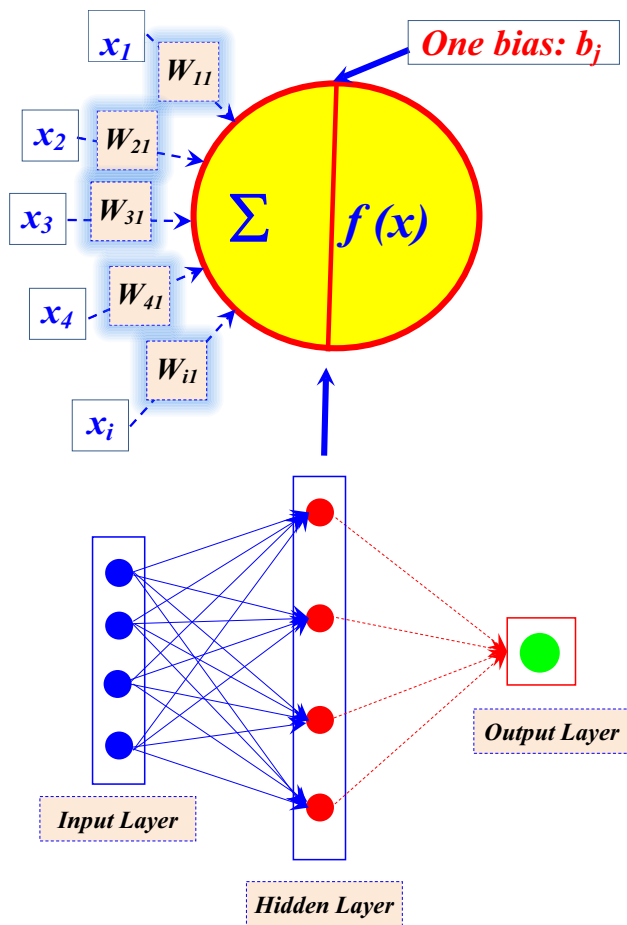


Fig. 2 The multilayer perceptron neural network (MLPNN) architecture

the variational mode decomposition (VMD), and the empirical wavelet transform (EWT). An example of the application of the three algorithms for maximal air temperature decomposition is provided in Fig. 5.

Empirical mode decomposition

Empirical mode decomposition (EMD) was proposed by Huang et al. (1998). The EMD algorithm is a preprocessing signal decomposition mainly used for filtering any nonlinear signal and making it as a series of sub signal called intrinsic mode functions (IMFs). In order to be an IMF oscillatory component, some conditions should be respected as the phase and amplitude of each one changes slowly: (i) only one zero value is attributed to one IMF between two consecutive extremes and (ii) the local average of the IMF is equal to zero (Alaodolehei et al. 2020; Abdulhay et al. 2020). Decomposition of nonlinear signal using the EMD, i.e., the “sifting” process, involves several stages, which can be summarized as follows (Abdollahpoor and Lotfivand 2020):

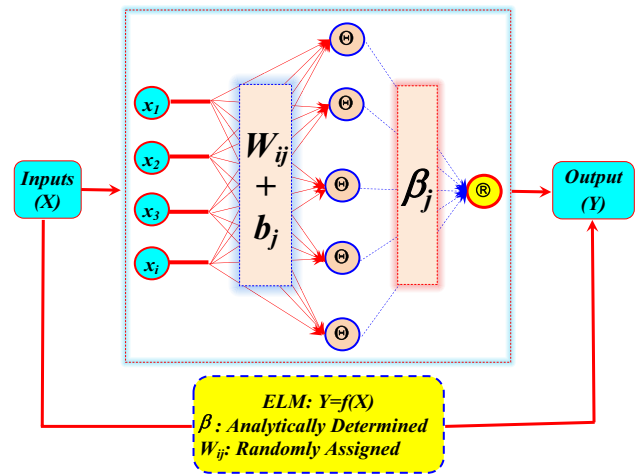


Fig. 3 The extreme learning machine (ELM) architecture

1. The calculated IMFs should be ordered from lower scale to higher scale.
2. The process of extracting the IMF will start by calculating the local maxima and minima if the signal $x(t)$.
3. The obtained local maxima and minima are used by the cubic spline curve for setting the upper and lower envelopes, and their average value, i.e., the $m_1(t)$, is then calculated.
4. Calculates the first component, i.e., $C_1(t)$ as follow:

$$C_1(t) = x(t) - C_1(t) \tag{7}$$

5. This process (i.e., the “sifting” process) is iterative and will continue until the first IMF is extracted.
6. While the process is iterative, there is a shutoff parameter for stopping the “sifting” process, i.e., the standard deviation (SD) calculated as follows (Abdollahpoor and Lotfivand 2020):

$$SD = \sum_{t=0}^T \left[\frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)} \right] \tag{8}$$

7. Finally, the original signal $x(t)$ can be reformulated as follow:

$$x(t) = \sum_{i=1}^N IMF_i(t) + R_N(t) \tag{9}$$

where N is the number of IMF and the R_N is called the residue (El Bouny et al. 2019).

VMD

Dragomiretskiy and Zosso (2014) propose the variational mode decomposition (VMD) for signal decomposition. The

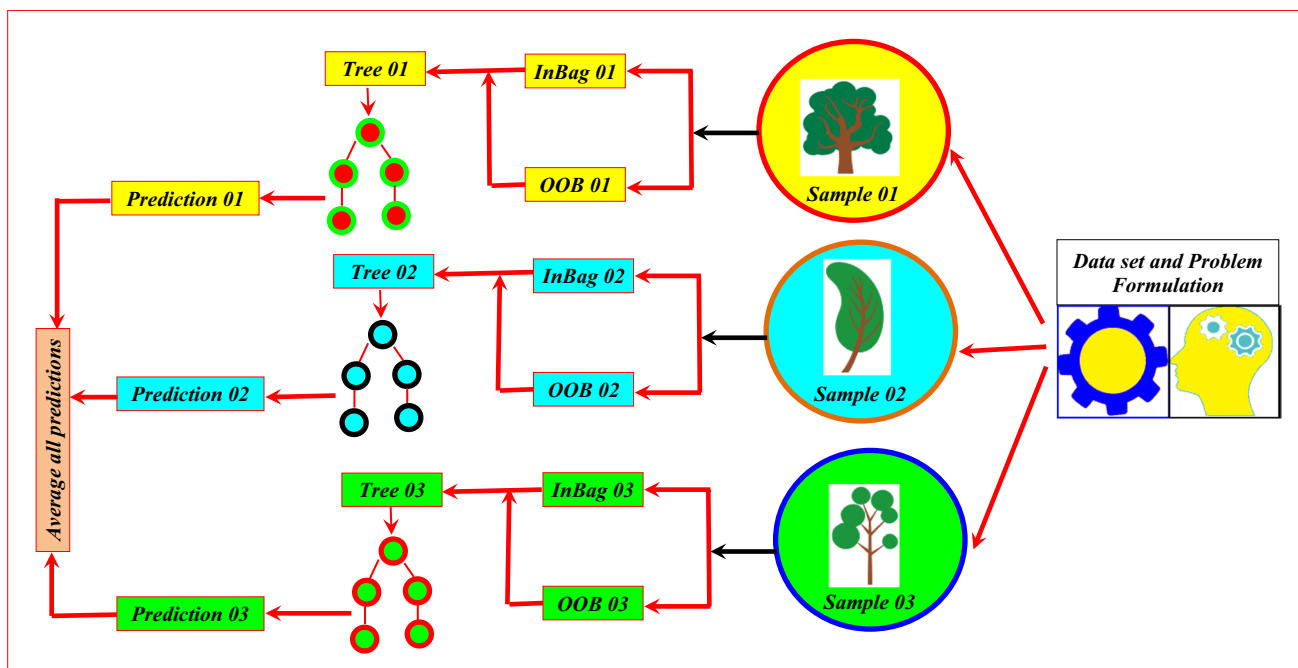


Fig. 4 The random forest regression (RFR) architecture

VMD uses an adaptive decomposition process for extracting a series of intrinsic mode functions (IMFs) characterized by specific sparsity properties (Li et al. 2022). The VMD estimates the modes, i.e., the IMFs and their respective center frequencies using an adaptively and concurrently algorithm

(Peng et al. 2020). Each mode calculated using the VMD could be formulated as follow:

$$u_k(t) = A_k(t)\cos[\varphi_k(t)] \tag{10}$$

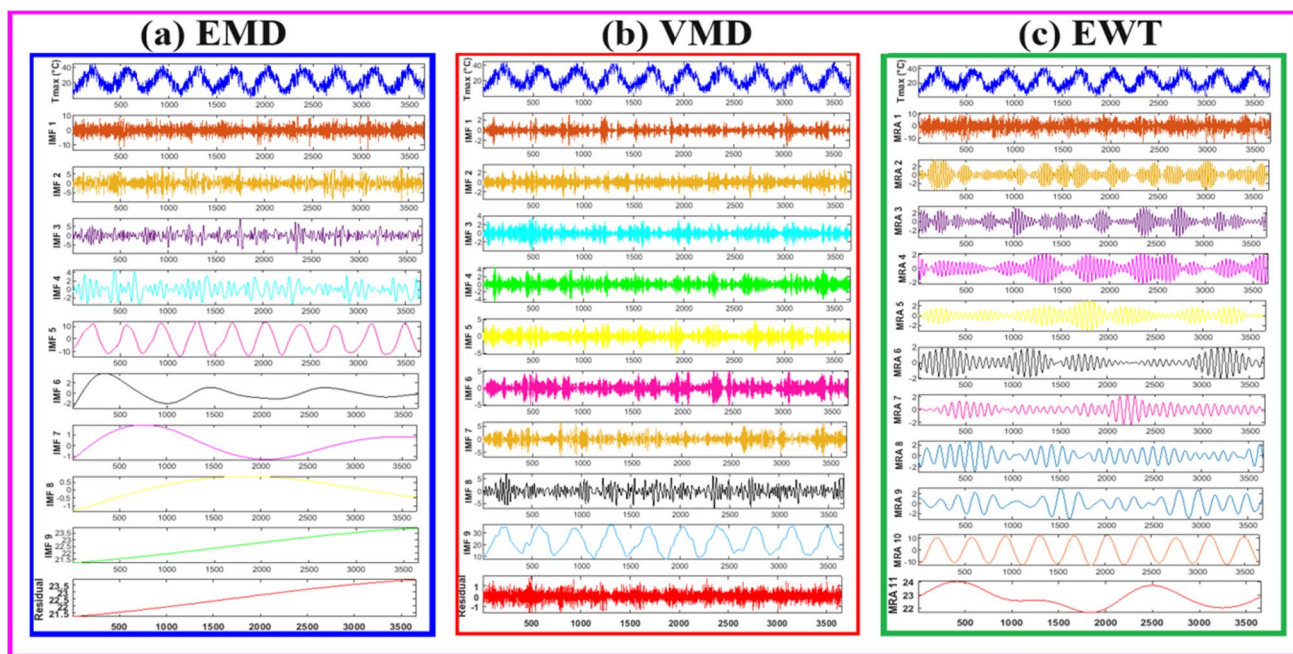


Fig. 5 An example of maximal air temperature (T_{max}) signal decomposition using a the EMD, b the VMD, and c the EWT

where $u_k(t)$ is the k th mode component, $A_k(t)$ is a non-negative envelope, and $\phi_k(t)$ is instantaneous phase, respectively (Li et al. 2022). Using the VMD, we suppose that each calculated mode corresponds to an IMF having a finite bandwidth and a central frequency, which were determined using an iterative searching process for an optimal solution (Zhang et al. 2020). Given any nonlinear and nonstationarity signal $f(t)$, the VMD can be seen as a minimization of a constrained optimization problem and the VMD can be constructed as follows:

$$\begin{cases} \min_{u_k, w_k} \left\{ \sum_{k=1}^K \|\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \|^2 \right\} \\ \text{subject to } \sum_{k=1}^K u_k = f(t) \end{cases} \quad (11)$$

where u_k ($k = 1, 2, 3, \dots, K$) represents the K IMFs using VMD, w_k is the corresponding center frequencies of each IMF, t is the time script, K represents the total number of sub-signals, $\delta(t)$ is the Dirac distribution, j is the imaginary unit, and \otimes is the convolution operator, and finally the expression $\left(\delta(t) + \frac{j}{\pi t} \right)$ corresponds to the Hilbert transform of $u_k(t)$ (Liu et al. 2021; Li et al. 2022; Peng et al. 2020).

EWT

The empirical wavelet transform (EWT) was introduced by Gilles (2013). The EWT was formulated based on the assumption that a signal $x(t)$ results from the sum of the individual subcomponents characterized by a compact support in Fourier spectrum (Liu et al. 2020). The EWT is used for providing a series of sub signal called multiresolution analysis (MRA) (Wang and Hu 2015). The EWT uses two distinguished functions, namely: (i) empirical wavelet functions ($\hat{\varphi}_n(\omega)$) (i.e., the band-pass filters) and (ii) empirical scale function $\hat{\psi}_n(\omega)$ (i.e., the low-pass filters). The two functions are defined by Eqs. (12) and (13), respectively (Hu et al. 2015; Si et al. 2019):

$$\hat{\varphi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq \omega_n - \tau_n \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} |\omega| - \omega_n + \tau_n \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

and

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_{n+1}} |\omega| - \omega_{n+1} + \tau_{n+1} \right) \right] & \text{if } \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_n + \tau_n \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} |\omega| - \omega_n + \tau_n \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

For the two above equations, (ω) is the n th maxima of the Fourier spectrum. The selection of the best (τ) value

should be proportional to (ω_n) for which $(\tau_n = \gamma \cdot \omega_n)$ where $(0 < \gamma < 1)$. The function $\beta(x) \in C^k([0, 1])$ is an arbitrary function and expressed as follows (Liu et al. 2020; Hu et al. 2015; Si et al. 2019):

$$\beta(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases} \text{ and } \beta(x) + \beta(1 - x) = 1 \forall x \in [0, 1] \quad (14)$$

Results and discussion

Model development

This study compares between single models, i.e., MLPNN, ELM, and RFR, and hybrid models based on signal decomposition, i.e., the EMD, VMD, and EWT. Hence, the hybrid models were designated hereafter as MLPNN_EMD, MLPNN_VMD, and MLPNN_EWT, and this is identical with the ELM and RFR models. For each single and hybrid models, six input combinations were tested and compared for showing the effect of varying the models structure on the predictive accuracy (Table 3). Furthermore, in this section, we try to investigate if we can obtain accurate RH prediction values and range, and more precisely, if the signal decomposition algorithms can help in rendering the predictive results more accurate. A series of evaluation metrics was used for comparison and to determine whether the hybrid models were better than the single models, i.e., the R, NSE, RMSE, and MAE. In addition, a graphical comparison using the scatterplot, boxplot, violin plot, and Taylor diagram between measured and predicted data was provided, and the results were presented for each station separately. The flowchart of the proposed modeling framework is depicted in Fig. 6.

Results at Constantine station

Table 3 lists the numerical performance of the three single models used in this study. Hereafter, we focus on the results during the validation stage.

According to Table 3, the MLPNN model yields better accuracy for all input combinations, but its numerical performances were slightly superior to those of the ELM and RFR, implying its superiority. First, using all five input variables, the MLPNN1 exhibited the high R (≈ 0.939) and NSE (≈ 0.882) values, and the lowest RMSE (≈ 7.44) and MAE (≈ 5.62) values, respectively. The ELM1 and the RFR1 yielded equally numerical performances slightly lower than the MLPNN1. The results in Table 3 show that in terms of the mean values, the MLPNN models yielded the biggest mean R (≈ 0.924) and NSE (≈ 0.851) values, and the lower mean RMSE (≈ 8.33) and MAE (≈ 6.43) values, followed

by the ELM models, and the RFR models were found to be the less accurate.

More precisely, in terms of RMSE and MAE, the biggest improvement among all models are gained using the MLPNN models with $\approx 1.662\%$ and $\approx 3.064\%$ compared to the ELM models and $\approx 4.82\%$ and $\approx 6.29\%$ compared to the RFR models. Beyond the first input combination, it is clear from Table 3 that the models based on four input variables (i.e., combination two and three) exhibited slightly lower performances and it is clear that the inclusion of the precipitation plays minor role in models performances improvement. The MLPNN2 and ELM2 for which the *P* was excluded worked equally having the same numerical performances superior to the RFR2. Furthermore, using only three input variables, i.e., the fourth and fifth input combinations, it is clear that the performances of the models were significantly decreased from the first to the fifth combination and the inferiority of the RFR5 became more obvious. The MLPNN1 improve the MLPNN5 by decreasing the RMSE and MAE by approximately $\approx 17.119\%$ and $\approx 19.611\%$, respectively, and the ELM1 improve the ELM5 by decreasing the RMSE and MAE by approximately $\approx 12.449\%$ and $\approx 12.802\%$, respectively, while the RMSE and MAE of the RFR5 was improved by $\approx 15.949\%$ and $\approx 19.124\%$, respectively.

Finally, our analysis revealed that the poorest performances among all proposed models were exhibited using the models having only two input variables for which the RMSE and MAE were significantly increased and the *R* and NSE values were dramatically decreased, and more precisely, the

RFR6 is the only one for which the *R* and the NSE were decreased below the values of ≈ 0.900 and ≈ 0.800 , respectively (Table 3).

In the second part of the present study, we tried to improve the RH predictive accuracy by applying the signal decomposition, and in total, three algorithms were compared, i.e., the EMD, VMD, and EWT. Obtained results are depicted in Table 4. A further discussion of the differences between the models with and without signal decomposition is warranted hereafter. An analysis of the ensemble results obtained using the EMD signal decomposition revealed that the contribution of the EMD in improving the performances of the ELM and RFR is too small; on the contrary, the numerical performances were slightly deteriorated.

The analysis show that the mean RMSE and MAE values of the single ELM models were relatively equal to those obtained using the ELM_EMD, while the mean *R* and NSE values remained unchangeable, from (≈ 0.921 and ≈ 0.845) to (≈ 0.920 and ≈ 0.846), respectively, showing the limitation of the EMD algorithm in improving the performances of the single ELM models. Regarding the RFR models, we can clearly see that mean RMSE and MAE were increased from (≈ 8.755 and ≈ 6.866) to (≈ 9.068 and ≈ 7.225), respectively. However, it is important to note that the performances of the single RFR5 and RFR6 were slightly improved using the EMD algorithm for which the values of RMSE and MAE were dropped from (≈ 9.424 and ≈ 7.53) to (≈ 9.209 and ≈ 7.302) between RFR5 and RFR_EMD5, and from (≈ 9.753 and ≈ 7.869) to (≈ 9.047 and ≈ 7.246) between RFR6 and RFR_EMD6. Concerning the MLPNN models, it is clear

Table 3 Performances of different standalone models at the Constantine station

Models	Training				Validation			
	<i>R</i>	NSE	RMSE	MAE	<i>R</i>	NSE	RMSE	MAE
ELM1	0.938	0.880	7.568	5.822	0.931	0.865	7.940	6.096
ELM2	0.935	0.875	7.745	5.849	0.932	0.868	7.861	6.009
ELM3	0.935	0.874	7.779	5.983	0.925	0.855	8.235	6.303
ELM4	0.932	0.868	7.960	6.170	0.925	0.852	8.333	6.551
ELM5	0.923	0.852	8.429	6.550	0.911	0.824	9.069	6.991
ELM6	0.918	0.843	8.670	6.751	0.903	0.811	9.400	7.870
MLPNN1	0.946	0.895	7.093	5.407	0.939	0.882	7.446	5.628
MLPNN2	0.937	0.877	7.674	5.827	0.932	0.868	7.871	6.062
MLPNN3	0.938	0.878	7.635	5.850	0.930	0.861	8.066	6.229
MLPNN4	0.932	0.869	7.919	6.115	0.925	0.853	8.313	6.494
MLPNN5	0.923	0.852	8.420	6.546	0.912	0.828	8.984	7.001
MLPNN6	0.915	0.837	8.825	6.858	0.904	0.815	9.313	7.186
RFR1	0.975	0.950	4.890	3.704	0.931	0.866	7.921	6.090
RFR2	0.971	0.940	5.339	3.967	0.922	0.850	8.376	6.303
RFR3	0.974	0.947	5.056	3.837	0.926	0.856	8.207	6.310
RFR4	0.955	0.908	6.641	5.166	0.915	0.833	8.847	7.092
RFR5	0.950	0.899	6.959	5.417	0.902	0.810	9.424	7.530
RFR6	0.940	0.880	7.586	5.910	0.894	0.797	9.753	7.869

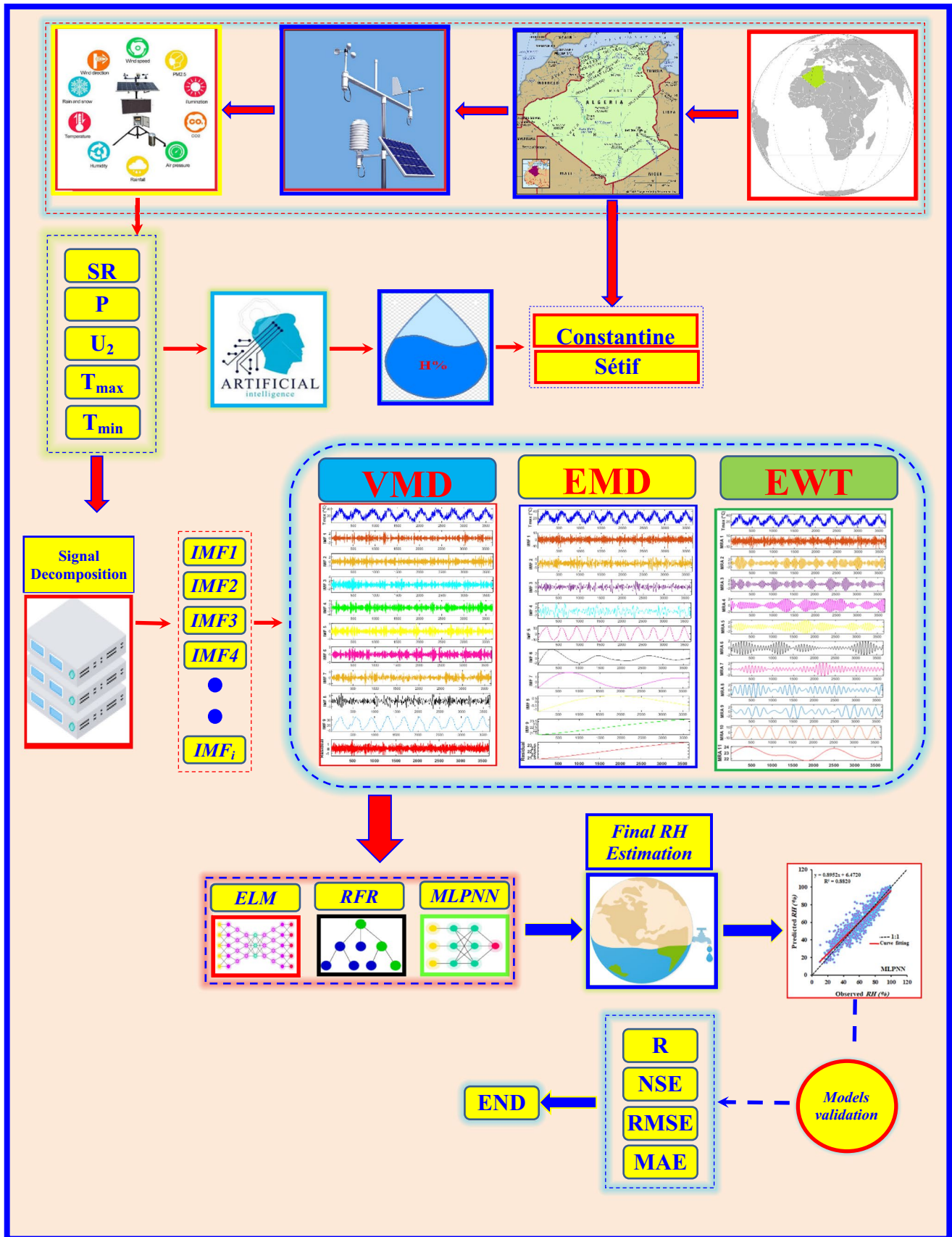


Fig. 6 Flowchart of the modelling strategy for air relative humidity

from the results reported in Table 4 that the MLPNN models are the only models for which significant improvement was gained using the EMD algorithm. All six models (i.e., from MLPNN1 to MLPNN6) showed their numerical performances improved by an increase in the R and NSE values and a decrease in the RMSE and MAE values. Overall comparison between models revealed that the mean RMSE and MAE values calculated using the MLPNN models were decreased from (≈ 8.332 and ≈ 6.433) to (≈ 7.371 and ≈ 5.795) exhibiting an improvement rates of approximately $\approx 11.532\%$ and $\approx 9.917\%$, respectively. Among the six input combination, it is clear that the MLPNN_EMD1 is the unique model that benefit most from the EMD showing its performances significantly increased with an improvement rates of approximately $\approx 1.064\%$, $\approx 2.040\%$, $\approx 7.89\%$, and $\approx 4.47\%$ in terms of R , NSE, RMSE, and MAE, respectively. As we can conclude that the EMD is an interesting algorithm for improving the MLPNN performances, but this cannot be generalized to the other machines learning models, i.e., the ELM and the RFR for which no improvement was gained.

It can be seen from Table 4 that when the VMD algorithm was used for signal decomposition, no improvement was gained and all machine-learning models have shown their performances decreased significantly, highlighting the limitation and the poor contribution of the VMD in improving the predictive accuracy of the RH. In order to evaluate the prediction performance of the proposed hybrid VMD models compared to the single models, the comparison between the mean four metrics, i.e., the R , NSE, RMSE, and MAE values, is discussed hereafter (Table 4). We can see that the mean R (≈ 0.908), NSE (≈ 0.823), RMSE (≈ 9.100), and MAE (≈ 7.211) of the ELM_VMD, the mean R (≈ 0.921), NSE (≈ 0.841), RMSE (≈ 8.555), and MAE (≈ 6.656) of the MLPNN_VMD, and the mean R (≈ 0.890), NSE (≈ 0.776), RMSE (≈ 10.259), and MAE (≈ 8.321) of the RFR_VMD were all less than the values obtained using the single ELM, MLPNN, and RFR models, which leads to the conclusion that further efforts are required to understand the very limitation and the poor contribution of the VMD in improving the accuracy of the RH estimation.

As shown in Table 4, compared with the VMD, the results of the EWT algorithm are more practical and can provide more support to the prediction of the RH. As shown in Table 4, it is obvious that the performances of the MLPNN and RFR models were improved and the used of the EWT leads to a significant increase in the mean R , NSE, RMSE, and MAE values, while the performances of the ELM models were decreased. Using the EWT, the mean R , NSE, RMSE, and MAE values of the MLPNN models were improved by $\approx 1.516\%$, $\approx 3.250\%$, $\approx 9.853\%$ and $\approx 8.528\%$, respectively. Similarly, the mean R , NSE, RMSE, and MAE values of the RFR models were improved by $\approx 1.494\%$, $\approx 2.035\%$, $\approx 4.959\%$, and $\approx 3.442\%$, respectively. Among

all proposed models, it is clear that the MLPNN_EWT1 was the best model showing its performances significantly improved compared to the single MLPNN1 model, exhibiting improvement rates of approximately $\approx 1.171\%$, $\approx 2.267\%$, $\approx 8.796\%$, and $\approx 6.876\%$, in terms of R , NSE, RMSE, and MAE values, respectively.

Furthermore, if all models were compared one by one, the mean RMSE and MAE values of the MLPNN6 having only the T_{\max} and P as input variables were decreased by approximately $\approx 15.215\%$ and $\approx 13.665\%$, respectively, which constitute the high improvement rates among all proposed models. From Tables 3 and 4, the numerical values of the performances metrics lead to conclude that the performances of the hybrid models based on signal decomposition are generally higher than the single models except the VMD algorithm who failed to give any improvement in terms of predictive accuracy. In addition, the experimental results show that the RMSE and MAE values of the EWT based models are the lowest, while the R and NSE values were the highest compared to the values obtained using the EMD algorithm. In conclusion, the hybrid EWT-based models have the best predictive performance and the relatively better air relative humidity estimation.

The models were further compared based on graphical comparison as shown in Figs. 7 and 8. In Fig. 7, we drawn the scatterplot of measured and predicted air relative humidity for the best single and hybrid models for which we can conclude that plotted data were less scattered using the MLPNN models with and without decomposition and the models based on the EWT were the best accurate models. According to the boxplot (Fig. 8a) and the violin plot (Fig. 8b), the models based on EWT were the most accurate and the MLPNN with and without decomposition improve all other models showing the high similarity with the measured one. However, according to the Taylor diagram (Fig. 8c), the RFR_VMD was the poorest model among all proposed models.

Results at Sétif station

Table 5 shows the predictive results for the training and validation data using the different singles methods and based on the six input combinations for Sétif station. When comparing the three single models (ELM, MLPNN, and RFR) taking into account the mean values of the performances metrics, it is obvious that the RFR models were worse than the ELM and the MLPNN models.

The ELM models yielded an R and NSE values ranging from ≈ 0.905 to ≈ 0.938 (mean ≈ 0.924), and from ≈ 0.818 to ≈ 0.879 (mean ≈ 0.854), the MLPNN models yielded an R and NSE values ranging from ≈ 0.908 to ≈ 0.943 (mean ≈ 0.926), and from ≈ 0.824 to ≈ 0.887 (mean ≈ 0.857), while the values obtained using the RFR were

Table 4 Performances of hybrid models based on EMD, VMD and EWT at the Constantine station

Models	Training				Validation			
	<i>R</i>	NSE	RMSE	MAE	<i>R</i>	NSE	RMSE	MAE
ELM_EMD1	0.945	0.894	7.129	5.659	0.926	0.857	8.173	6.453
ELM_EMD2	0.952	0.907	6.674	5.247	0.915	0.836	8.760	6.906
ELM_EMD3	0.956	0.913	6.448	5.073	0.927	0.858	8.168	6.456
ELM_EMD4	0.958	0.918	6.274	4.907	0.933	0.869	7.843	6.104
ELM_EMD5	0.946	0.895	7.094	5.615	0.910	0.827	9.015	7.159
ELM_EMD6	0.949	0.901	6.878	5.361	0.911	0.829	8.950	6.976
MLPNN_EMD1	0.974	0.947	5.052	3.903	0.949	0.900	6.858	5.376
MLPNN_EMD2	0.969	0.938	5.431	4.237	0.946	0.894	7.044	5.445
MLPNN_EMD3	0.969	0.938	5.433	4.219	0.946	0.896	6.994	5.466
MLPNN_EMD4	0.961	0.918	6.270	4.913	0.938	0.878	7.558	6.061
MLPNN_EMD5	0.955	0.911	6.546	5.075	0.935	0.874	7.669	6.055
MLPNN_EMD6	0.945	0.893	7.176	5.638	0.927	0.860	8.105	6.365
RFR_EMD1	0.982	0.954	4.687	3.657	0.917	0.828	8.989	7.133
RFR_EMD2	0.981	0.953	4.763	3.708	0.916	0.825	9.045	7.184
RFR_EMD3	0.980	0.950	4.887	3.838	0.914	0.822	9.136	7.288
RFR_EMD4	0.979	0.948	4.980	3.912	0.919	0.828	8.979	7.195
RFR_EMD5	0.978	0.947	5.048	3.958	0.912	0.819	9.209	7.302
RFR_EMD6	0.976	0.943	5.231	4.108	0.916	0.825	9.047	7.246
ELM_VMD1	0.938	0.879	7.615	5.995	0.916	0.838	8.722	6.921
ELM_VMD2	0.928	0.861	8.158	6.422	0.897	0.804	9.579	7.613
ELM_VMD3	0.943	0.888	7.313	5.766	0.918	0.841	8.620	6.823
ELM_VMD4	0.941	0.885	7.408	5.817	0.920	0.846	8.485	6.742
ELM_VMD5	0.930	0.865	8.046	6.334	0.898	0.805	9.547	7.514
ELM_VMD6	0.930	0.866	8.022	6.330	0.896	0.801	9.646	7.655
MLPNN_VMD1	0.971	0.942	5.267	4.051	0.937	0.877	7.600	5.906
MLPNN_VMD2	0.957	0.915	6.371	4.966	0.929	0.859	8.122	6.311
MLPNN_VMD3	0.908	0.805	9.661	7.556	0.879	0.736	11.125	8.682
MLPNN_VMD4	0.957	0.915	6.394	4.937	0.936	0.876	7.607	5.933
MLPNN_VMD5	0.949	0.901	6.877	5.314	0.923	0.852	8.322	6.436
MLPNN_VMD6	0.939	0.881	7.559	5.913	0.919	0.844	8.553	6.666
RFR_VMD1	0.981	0.949	4.924	3.925	0.892	0.779	10.177	8.276
RFR_VMD2	0.980	0.946	5.080	4.030	0.892	0.780	10.153	8.227
RFR_VMD3	0.980	0.947	5.043	4.027	0.890	0.775	10.270	8.343
RFR_VMD4	0.979	0.944	5.196	4.145	0.891	0.777	10.232	8.328
RFR_VMD5	0.978	0.943	5.206	4.135	0.886	0.770	10.376	8.372
RFR_VMD6	0.976	0.940	5.355	4.255	0.886	0.772	10.347	8.382
ELM_EWT1	0.959	0.919	6.232	4.904	0.927	0.859	8.139	6.411
ELM_EWT2	0.957	0.915	6.366	5.041	0.919	0.843	8.579	6.792
ELM_EWT3	0.959	0.919	6.230	4.941	0.920	0.845	8.534	6.830
ELM_EWT4	0.959	0.919	6.237	4.905	0.923	0.851	8.344	6.686
ELM_EWT5	0.953	0.908	6.626	5.225	0.904	0.815	9.316	7.323
ELM_EWT6	0.954	0.910	6.566	5.134	0.910	0.828	8.977	6.949
MLPNN_EWT1	0.979	0.957	4.528	3.509	0.950	0.902	6.791	5.241
MLPNN_EWT2	0.974	0.947	5.040	3.929	0.946	0.895	7.014	5.472
MLPNN_EWT3	0.944	0.891	7.239	5.735	0.918	0.842	8.609	6.834
MLPNN_EWT4	0.969	0.938	5.446	4.226	0.947	0.896	6.994	5.479
MLPNN_EWT5	0.962	0.924	6.033	4.703	0.934	0.871	7.763	6.078
MLPNN_EWT6	0.955	0.911	6.528	5.107	0.931	0.867	7.896	6.204
RFR_EWT1	0.985	0.963	4.206	3.266	0.931	0.857	8.193	6.500

Table 4 (continued)

Models	Training				Validation			
	<i>R</i>	NSE	RMSE	MAE	<i>R</i>	NSE	RMSE	MAE
RFR_EWT2	0.983	0.961	4.347	3.364	0.928	0.852	8.327	6.603
RFR_EWT3	0.984	0.961	4.322	3.360	0.931	0.856	8.214	6.535
RFR_EWT4	0.983	0.959	4.432	3.470	0.930	0.853	8.309	6.669
RFR_EWT5	0.981	0.956	4.571	3.541	0.927	0.849	8.414	6.700
RFR_EWT6	0.980	0.954	4.719	3.677	0.925	0.847	8.466	6.769

ranged from ≈ 0.902 to ≈ 0.933 (mean ≈ 0.919), and from ≈ 0.800 to ≈ 0.871 (mean ≈ 0.834). Using the MLPNN instead of the ELM and RFR allows achieving better performances, however, it shows generally slightly better performances compared to the ELM in terms of the mean *R*, NSE, RMSE, and MAE for which the difference was completely negligible. Overall, among all proposed models, it is clear that the MLPNN1 was the most accurate, followed by the ELM1 the RFR2, respectively. Regarding our analysis based on the number of input variables included, we can conclude that the use of more input leads to more reliable prediction. In fact, the fifth and six input combination appear the poorest and more precisely, the lowest accuracy was obtained with the MLPNN6 having an *R* ≈ 0.908 , NSE ≈ 0.824 , RMSE ≈ 9.660 , and MAE ≈ 7.588 , followed by the ELM6 *R* ≈ 0.905 , NSE ≈ 0.818 , RMSE ≈ 9.839 , and MAE ≈ 7.521 , while the RFR6 was the poorest with *R* ≈ 0.902 , NSE ≈ 0.800 , RMSE ≈ 10.316 , and MAE ≈ 7.719 , respectively. We conduct a systematic analysis of the model performance based on signal decompositions algorithms. According to Table 6, using the VMD algorithm, only the MLPNN models have shown their performances improved, while the performances of the ELM and RFR have significantly deteriorated. First, the mean *R*, NSE, RMSE, and MAE values of the MLPNN were slightly improved using the VMD algorithm showing improvement rates of $\approx 0.66\%$, $\approx 1.342\%$, $\approx 3.858\%$, and $\approx 1.806\%$, respectively. The improvement begin to be considered as significant beyond the third (MLPNN_VMD3) until the last input combinations (MLPNN_VMD6), for which the most significant improvement was gained exhibiting an improvement rates of approximately $\approx 1.762\%$, $\approx 3.641\%$, $\approx 8.820\%$, and $\approx 8.448\%$, respectively (MLPNN_VMD6 compared to the MLPNN6).

The performances and effectiveness of the hybrid models based on EMD algorithm are presented and discussed. Based on the results in Table 6, the improved percentages of each single model by the proposed hybrid models have been calculated as follow: (i) the mean *R*, NSE, RMSE, and MAE of the ELM models were decreased by $\approx 0.25\%$, $\approx 0.53\%$, $\approx 1.685\%$, and $\approx 3.957\%$, respectively, showing the limitation of the EMD algorithm improving the single ELM models; (ii) the mean *R*, NSE, RMSE, and MAE of the MLPNN

models were increased by $\approx 2.753\%$, $\approx 5.583\%$, $\approx 13.975\%$, and $\approx 14.51\%$, respectively, showing the significant contribution of the EMD in improving the performances of the MLPNN models; and (iii) the mean *R*, NSE, and RMSE of the RFR models were enhanced by $\approx 0.363\%$, $\approx 0.280\%$, and ≈ 0.455 , respectively, exhibiting a negligible improvement of the single RFR models.

From Table 6, it can be found that (a) compared with the single models, the MLPNN_EMD1 to MLPNN_EMD4 have obtained the highest prediction accuracy, and the MLPNN_EMD1 shows better performances that all other models exhibiting the highest *R* (≈ 0.949) and NSE (≈ 0.901) values, and the lowest RMSE (≈ 7.268) and MAE (≈ 5.609) values; (b) the prediction accuracies of the hybrid MLPNN_EMD1 to MLPNN_EMD4 are little difference and the ELM-based hybrid models have a little better prediction performances than the RFR-based hybrid models. This indicate that the ELM is more appropriate to building robust predictive models; (c) using only two input variables for all the predictive models, the MLPNN_EMD6 shows the best prediction performances with *R* (≈ 0.933) and NSE (≈ 0.870), RMSE (≈ 8.310), and MAE (≈ 6.487) exhibiting improvement rates of approximately $\approx 2.753\%$, $\approx 5.583\%$, $\approx 13.975\%$, and $\approx 14.510\%$, compared to the MLPNN6.

It can be seen from Table 6 that (a) the improved percentages of the three models using the EWT algorithm are almost relatively equal just except that the mean values of the *R* (≈ 0.939), NSE (≈ 0.875), RMSE (≈ 8.159), and MAE (≈ 6.468) obtained using the RFR_EWT were slightly superior to those obtained using the ELM_EWT (*R* ≈ 0.931 , NSE ≈ 0.867 , RMSE ≈ 8.421 , MAE ≈ 6.670), and those of the MLPNN_EWT (*R* ≈ 0.932 , NSE ≈ 0.865 , RMSE ≈ 8.400 , MAE ≈ 6.662), respectively, which indicates that the proposed EWT algorithm has obviously enhanced the prediction accuracy; (b) among all proposed models, the MLPNN_EWT1 yielded the best prediction performances improvement for which the *R* and NSE values were remarkably increased from (≈ 0.943 and ≈ 0.887) to (≈ 0.955 and ≈ 0.912) exhibiting an improvement rates of approximately $\approx 1.273\%$ and $\approx 2.818\%$, and the RMSE and MAE values were dramatically decreased from (≈ 7.729 and ≈ 5.933) to (≈ 6.820 and ≈ 5.293) exhibiting an enhancement rates of approximately $\approx 11.761\%$ and $\approx 10.787\%$, respectively;

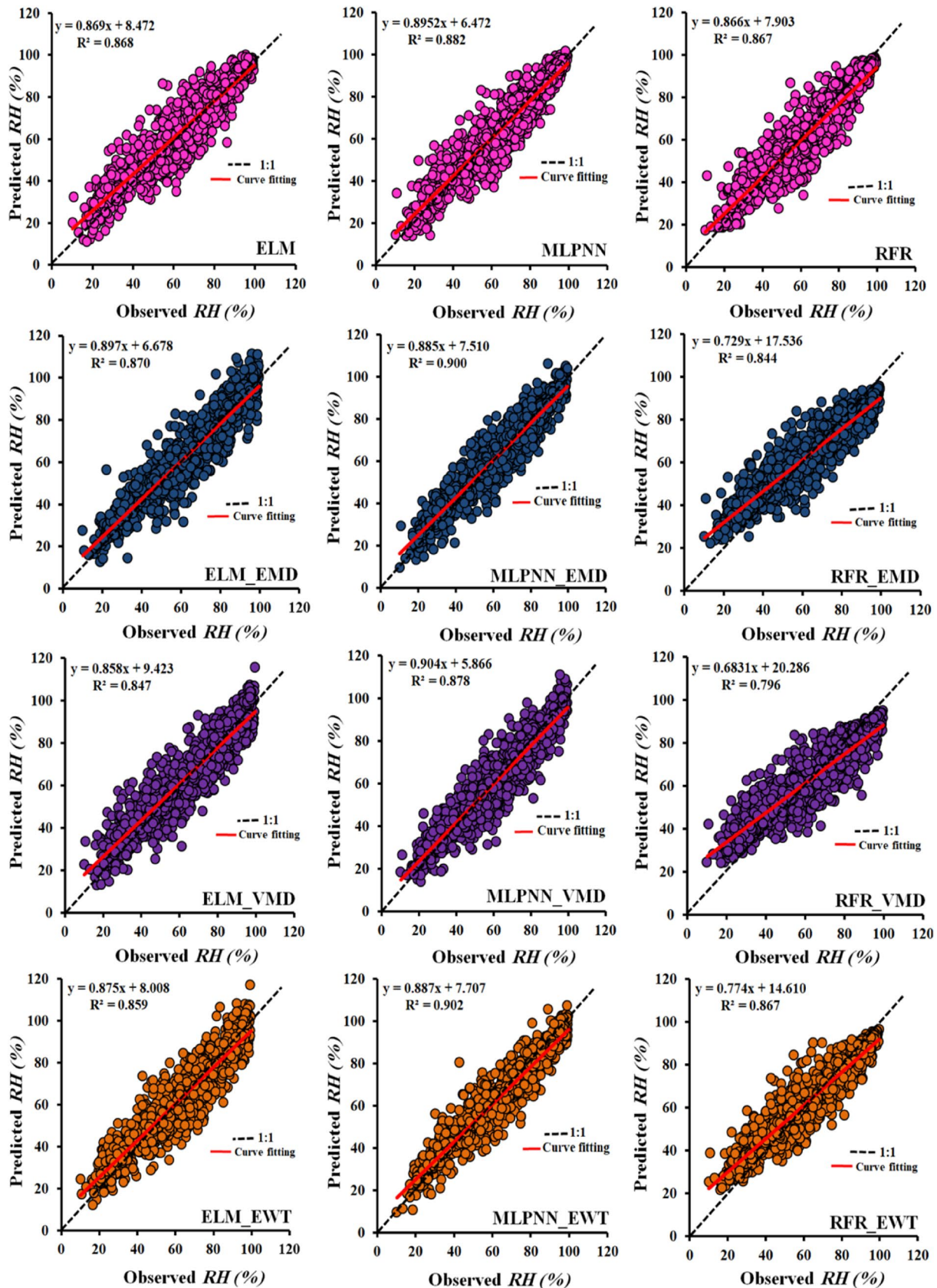


Fig. 7 Scatterplot of measured against predicted RH using the best single and hybrid models for the Constantine station: validation stage

Fig. 8 Graphs showing the comparison between measured and predicted air relative humidity during the validation stage: **a** boxplot, **b** violin plot, and **(c)** Taylor diagram

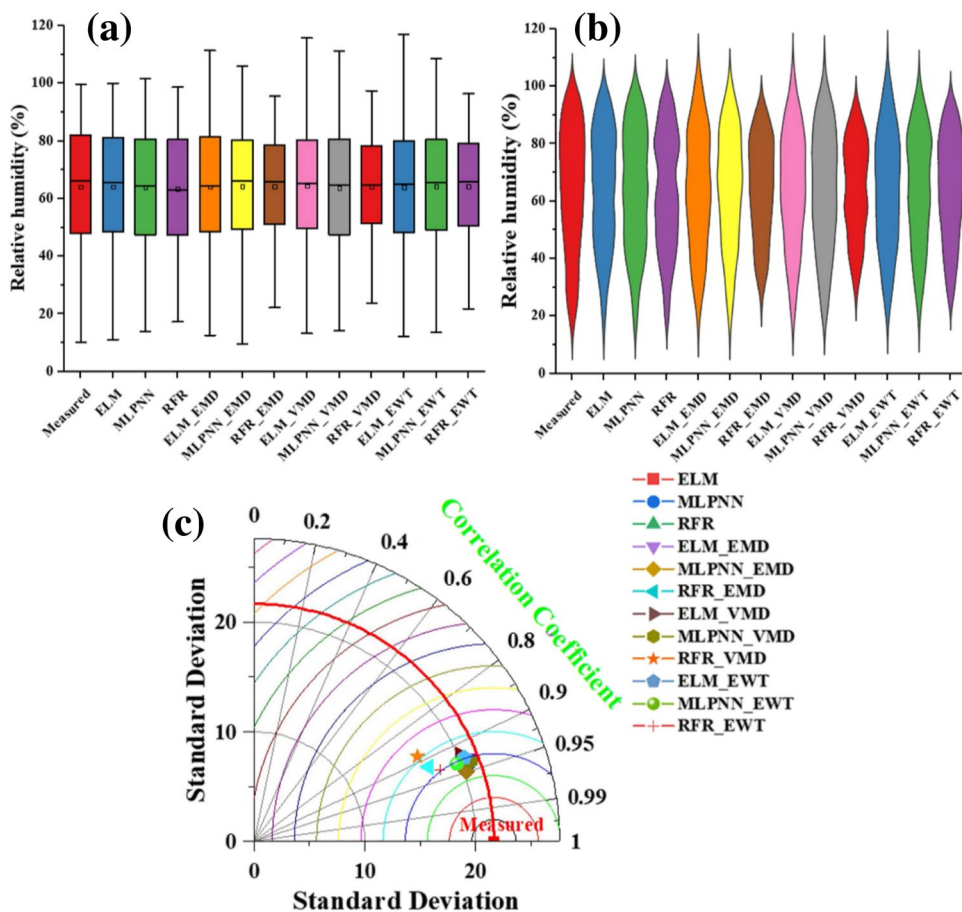


Table 5 Performances of different standalone models at the Sétif station

Models	Training				Validation			
	R	NSE	RMSE	MAE	R	NSE	RMSE	MAE
ELM1	0.942	0.888	7.837	5.930	0.938	0.879	8.016	6.179
ELM2	0.939	0.882	8.045	5.965	0.936	0.877	8.095	6.153
ELM3	0.936	0.876	8.252	6.225	0.927	0.860	8.634	6.695
ELM4	0.935	0.874	8.327	6.298	0.924	0.853	8.830	6.864
ELM5	0.925	0.855	8.919	6.672	0.913	0.834	9.386	7.219
ELM6	0.921	0.849	9.099	6.924	0.905	0.818	9.839	7.521
MLPNN1	0.951	0.903	7.282	5.513	0.943	0.887	7.729	5.933
MLPNN2	0.945	0.892	7.705	5.732	0.939	0.882	7.929	6.038
MLPNN3	0.941	0.884	7.968	5.957	0.930	0.864	8.501	6.467
MLPNN4	0.937	0.876	8.232	6.252	0.925	0.854	8.790	6.810
MLPNN5	0.927	0.858	8.817	6.669	0.912	0.832	9.435	7.316
MLPNN6	0.922	0.850	9.061	6.920	0.908	0.824	9.660	7.588
RFR1	0.975	0.950	5.244	3.835	0.930	0.855	8.779	6.494
RFR2	0.972	0.944	5.556	4.050	0.933	0.871	8.284	6.236
RFR3	0.973	0.946	5.430	3.971	0.923	0.842	9.146	6.788
RFR4	0.960	0.920	6.622	4.964	0.919	0.829	9.540	7.311
RFR5	0.957	0.914	6.866	5.111	0.907	0.806	10.144	7.614
RFR6	0.948	0.897	7.508	5.568	0.902	0.800	10.316	7.719

Table 6 Performances of hybrid models based on EMD, VMD, and EWT at the Sétif station

Models	Training				Validation			
	<i>R</i>	NSE	RMSE	MAE	<i>R</i>	NSE	RMSE	MAE
ELM_EMD1	0.954	0.911	7.001	5.479	0.920	0.847	9.005	7.109
ELM_EMD2	0.952	0.907	7.150	5.582	0.915	0.837	9.296	7.284
ELM_EMD3	0.956	0.915	6.837	5.296	0.930	0.864	8.507	6.699
ELM_EMD4	0.957	0.917	6.767	5.313	0.925	0.856	8.747	6.898
ELM_EMD5	0.953	0.908	7.111	5.544	0.915	0.836	9.343	7.336
ELM_EMD6	0.956	0.914	6.860	5.363	0.924	0.854	8.807	6.979
MLPNN_EMD1	0.975	0.950	5.223	4.051	0.949	0.901	7.268	5.609
MLPNN_EMD2	0.970	0.940	5.738	4.445	0.949	0.899	7.338	5.793
MLPNN_EMD3	0.973	0.947	5.417	4.171	0.949	0.899	7.317	5.717
MLPNN_EMD4	0.967	0.935	5.986	4.619	0.947	0.895	7.450	5.812
MLPNN_EMD5	0.932	0.833	9.574	7.640	0.908	0.795	10.438	8.347
MLPNN_EMD6	0.954	0.910	7.020	5.399	0.933	0.870	8.310	6.487
RFR_EMD1	0.983	0.958	4.794	3.693	0.927	0.845	9.081	7.274
RFR_EMD2	0.982	0.956	4.927	3.784	0.927	0.845	9.061	7.283
RFR_EMD3	0.981	0.954	5.004	3.839	0.923	0.836	9.318	7.455
RFR_EMD4	0.980	0.951	5.203	4.020	0.920	0.830	9.506	7.629
RFR_EMD5	0.979	0.951	5.199	3.984	0.921	0.835	9.374	7.542
RFR_EMD6	0.978	0.948	5.332	4.107	0.916	0.826	9.613	7.715
ELM_VMD1	0.937	0.878	8.179	6.402	0.914	0.835	9.356	7.452
ELM_VMD2	0.928	0.862	8.702	6.917	0.908	0.824	9.661	7.824
ELM_VMD3	0.941	0.886	7.916	6.197	0.918	0.842	9.165	7.256
ELM_VMD4	0.942	0.887	7.863	6.147	0.919	0.844	9.087	7.182
ELM_VMD5	0.936	0.876	8.253	6.475	0.905	0.818	9.831	7.849
ELM_VMD6	0.936	0.876	8.251	6.484	0.908	0.824	9.676	7.639
MLPNN_VMD1	0.971	0.942	5.633	4.370	0.940	0.882	7.906	6.294
MLPNN_VMD2	0.965	0.931	6.170	4.816	0.937	0.878	8.047	6.373
MLPNN_VMD3	0.965	0.931	6.143	4.768	0.934	0.871	8.263	6.435
MLPNN_VMD4	0.961	0.923	6.490	5.006	0.934	0.872	8.230	6.485
MLPNN_VMD5	0.955	0.911	6.972	5.380	0.925	0.855	8.782	6.893
MLPNN_VMD6	0.949	0.901	7.381	5.616	0.924	0.854	8.808	6.947
RFR_VMD1	0.982	0.954	5.001	3.920	0.908	0.812	9.988	8.042
RFR_VMD2	0.981	0.951	5.166	4.045	0.904	0.804	10.193	8.182
RFR_VMD3	0.981	0.952	5.120	4.027	0.909	0.813	9.977	8.090
RFR_VMD4	0.981	0.950	5.231	4.124	0.910	0.813	9.972	8.126
RFR_VMD5	0.979	0.949	5.304	4.126	0.907	0.811	10.025	8.101
RFR_VMD6	0.978	0.947	5.390	4.207	0.909	0.814	9.945	8.057
ELM_EWT1	0.958	0.918	6.704	5.329	0.933	0.871	8.286	6.514
ELM_EWT2	0.957	0.916	6.769	5.384	0.923	0.852	8.878	6.966
ELM_EWT3	0.961	0.924	6.458	5.107	0.937	0.877	8.088	6.478
ELM_EWT4	0.960	0.921	6.564	5.186	0.938	0.880	7.990	6.385
ELM_EWT5	0.957	0.915	6.817	5.379	0.929	0.862	8.556	6.756
ELM_EWT6	0.955	0.912	6.937	5.463	0.926	0.857	8.727	6.921
MLPNN_EWT1	0.979	0.958	4.773	3.722	0.955	0.912	6.820	5.293
MLPNN_EWT2	0.931	0.863	8.680	6.876	0.904	0.816	9.874	7.866
MLPNN_EWT3	0.965	0.931	6.171	4.855	0.943	0.888	7.716	6.022
MLPNN_EWT4	0.951	0.903	7.303	5.759	0.935	0.874	8.185	6.483
MLPNN_EWT5	0.965	0.930	6.184	4.814	0.942	0.887	7.761	6.081
MLPNN_EWT6	0.925	0.831	9.623	7.582	0.914	0.810	10.043	7.984
RFR_EWT1	0.986	0.968	4.201	3.225	0.943	0.880	7.968	6.327

Table 6 (continued)

Models	Training				Validation			
	<i>R</i>	NSE	RMSE	MAE	<i>R</i>	NSE	RMSE	MAE
RFR_EWT2	0.985	0.965	4.407	3.376	0.940	0.876	8.130	6.425
RFR_EWT3	0.986	0.966	4.290	3.288	0.941	0.877	8.074	6.391
RFR_EWT4	0.985	0.964	4.437	3.441	0.939	0.874	8.177	6.501
RFR_EWT5	0.984	0.963	4.531	3.477	0.937	0.872	8.255	6.546
RFR_EWT6	0.982	0.960	4.682	3.623	0.935	0.869	8.352	6.619

(c) using only two input variables, we can see that the RFR_EWT6 was the best model showing and improvement rates of approximately $\approx 3.659\%$, $\approx 8.625\%$, $\approx 19.038\%$, and $\approx 14.251\%$ compared to the single RFR6, which is the high improvement rate gained using all three signal decomposition algorithms. The scatterplots of the measured and predicted air relative humidity for the best single and hybrid models are depicted in Fig. 9. The boxplot, violin plot, and the Taylor diagram are depicted in Fig. 10.

Discussion

This section further discusses the results of the above experiments with already published works. To prove that the presented combined models based on signal decomposition indeed improves the air relative humidity prediction accuracy, the values of the performances indices obtained using our approach are compared with those reported in the literature in a similar studies. In comparison to our study, Tao et al. (2022) produce higher correlations between the measured and predicted RH ($R \approx 0.984$) using the MARS model at Mosul station, Irak, compared to the best value obtained in our study ($R \approx 0.955$) at Sétif Station, but produce lower correlation between the measured and predicted RH at Kut station with values ranging from $R \approx 0.890$ to $R \approx 0.946$, compared to the values of $R \approx 0.955$ obtained using the MLPNN_EWT proposed in our study. Relative to our study, Yasar et al. (2012) have higher correlation between measured and predicted RH, exhibiting excellent correlation coefficient ranging from $R \approx 0.960$ to $R \approx 0.999$. Again, this leads them to conclude a significant and critical role of the latitude, longitude, and altitude in increasing the predictive accuracy. In the investigation conducted by Hanoon et al. (2021), they certainly downplayed the potential advantage of machine learning models, i.e., the MLPNN and the RBFNN used for modelling RH (the R values were ranged from ≈ 0.634 to ≈ 0.713), and therefore reported a lesser role of the modelling scenario based only on the inclusion of the RH measured at several time lags.

The motivation of Adnan et al. (2021) study was to investigate whether the inclusion of the RH measured at previous lag combined with mean air temperature can yielded high

predictive accuracy. They obtained a correlation coefficients ranging from ≈ 0.898 to ≈ 0.903 using the MARS model, and ranging from ≈ 0.873 to ≈ 0.885 using the M5Tree model; hence, our correlation coefficients are likely more accurate ($R \approx 0.955$). For instance, Shi et al. (2018) simulate the RH at three different time scale, i.e., 6, 24, and 72 h. Depending on the time scale, they show that there is a significant decrease in the MLPNN model performances, and more precisely, the high R value (≈ 0.988) obtained for 6-h interval of time was dropped to be (≈ 0.852) for the model based on 24 h, and further deteriorated to be very low (≈ 0.806) for the model at 72-h interval of time, which are less than the values achieved in our present study. Finally, the modelling strategy proposed by Hussein et al. (2021) for which the RH was predicted using monthly image data, it was found that excellent predictive accuracy can be achieved with R values ranging from ≈ 0.980 to ≈ 0.999 ; however, their approach was criticized for employing high precision satellite images, which are not always available for the majors part of the world.

The significant improvements in air relative humidity prediction have made the proposed signal decomposition to be more attractive tool. As the needs for continuous measurement of air relative humidity continue to increase, the issues caused by some single and standalone machine learning models have become more relevant, which have motivated the introduction of new robust modelling frameworks. In this context, the improvement of air relative humidity estimation has boosted the development of new modelling strategy based on the combination of two paradigms: machine learning and signal decomposition. However, despite the increasing number of published paper from different countries, making generalization and conclusions more difficult because the obtained results from one study to another varied according to the type of data, the extent of data, how the models were evaluated, and how the modelling strategy was formulated. Based on the idea that air relative humidity variability is governed by the fluctuation of various weather variables, our approach was based on testing various input combination. In summary, to introduce an accurate air relative humidity prediction, a novel combined model based on EMD, VMD, EWT, and machine learning models is proposed. Among the proposed decomposition

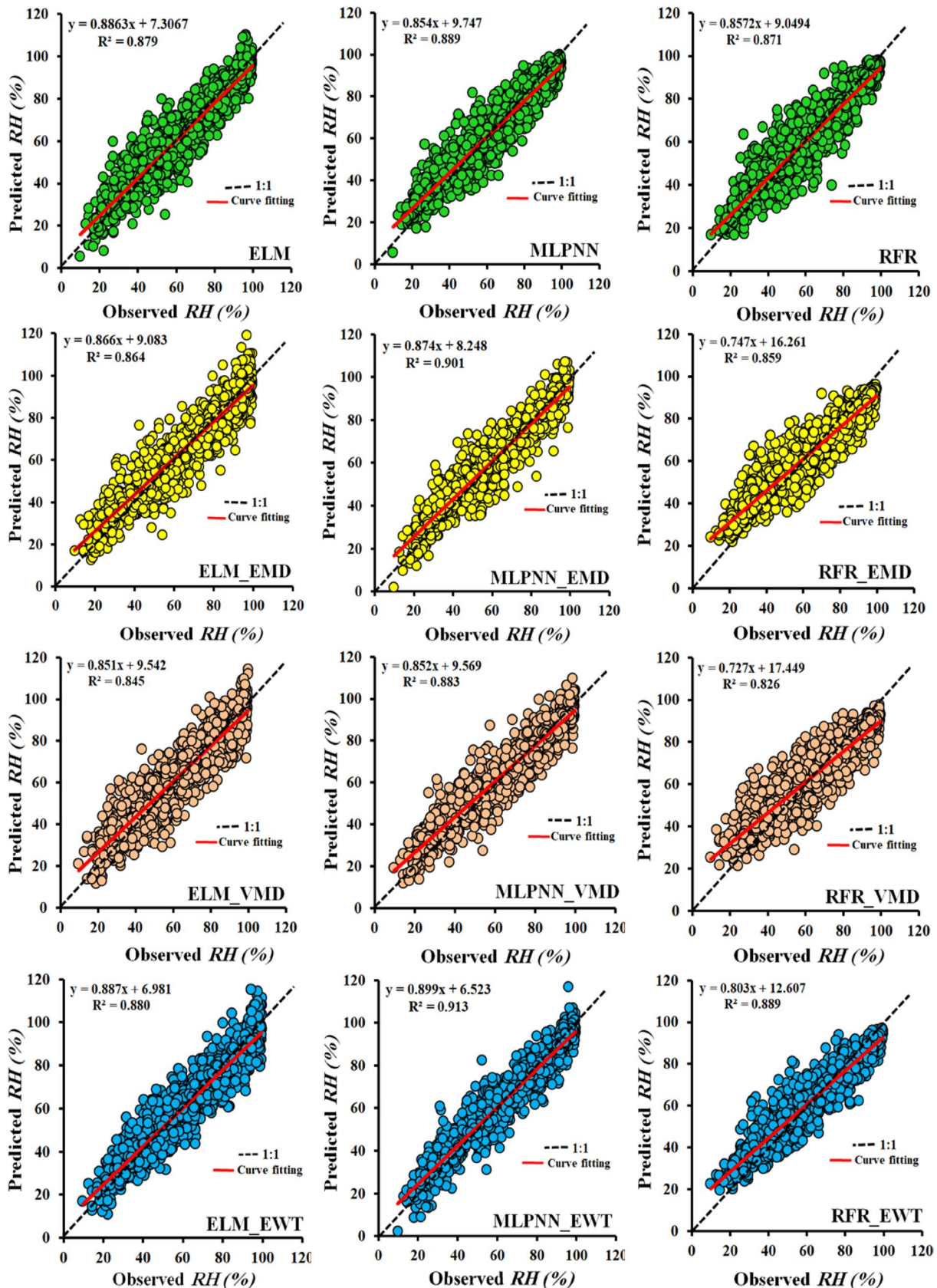
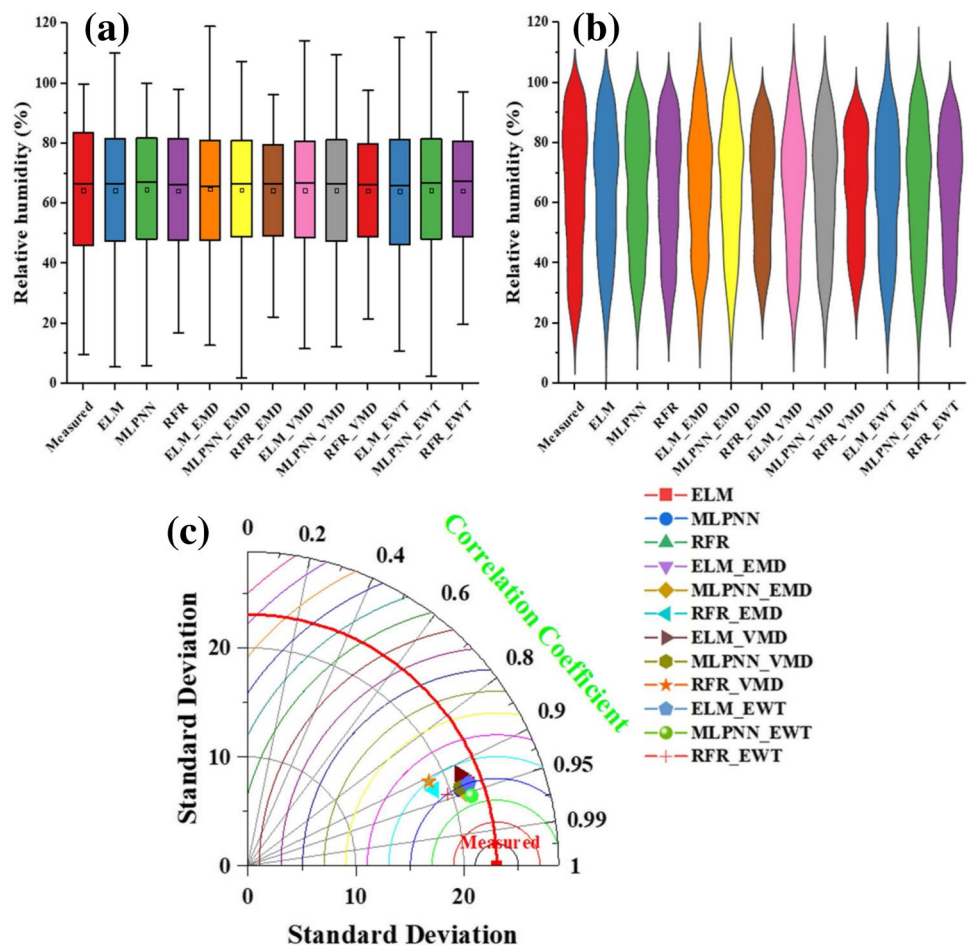


Fig. 9 Scatterplot of measured against predicted RH using the best single and hybrid models for Sétif station: validation stage

Fig. 10 Graphs showing the comparison between measured and predicted air relative humidity for Sétif station during the validation stage: **a** boxplot, **b** violin plot, and **c** Taylor diagram



algorithms, the EWT was found to be the most significant algorithm for improving the estimation of the air relative humidity, and this was confirmed for all machine-learning models. Meanwhile, the VMD was found to be the poorest algorithm compared to the EMD and EWT. Ultimately, the results obtained using the MLPNN was clearly superior to those obtained using the RFR and the ELM models based on all decomposition algorithms. At Constantine station, the MLPNN_EWT1 showed significantly better performances than the MLPNN_EMD1 and the MLPNN_VDM1 with R , NSE, RMSE, and MAE of ≈ 0.950 , ≈ 0.902 , ≈ 6.791 , and ≈ 5.241 , respectively. At Sétif station, the same hybrid model, i.e., the MLPNN_EWT1, exhibited the high performances with R , NSE, RMSE, and MAE of ≈ 0.955 , ≈ 0.912 , ≈ 6.820 , and ≈ 5.293 , respectively. In summary, the above experiments show that hybrid model based on EWT algorithm predicts the air relative humidity accurately with more precision and it leverages the strong correlations between measure and calculated data for different sites and they significantly enhance the single models and they have demonstrated to be more suitable for ensuring better generalization ability than single machine learning models.

Conclusion

In this paper, we propose a new data driven approaches for better prediction of air relative humidity (RH) based on signal decomposition algorithms and standalone machine learning hybrid models. While the use of signal decomposition algorithms is broadly reported in the literature for hydrological modelling studies and agrometeorological variables prediction, few investigations were related to their application for RH prediction. Our hybrid models have demonstrated their suitability in improving the performances of the standalone models especially using fewer input variables, and it was found that the VMD was the less accurate algorithm exhibiting poor performances compared to the EMD and EWT algorithms. Further analysis revealed that while the two studied stations were located in the same climatic zone, i.e., the semi-arid climate, the proposed standalone and hybrid models worked differently depending on the input variables combinations. In overall, it was demonstrated that the MLPNN model was more accurate compared to the ELM and RFR with and without signal decomposition, and it takes full advantages from the EMD and EWT rather than the VMD. More precisely, the improvements rates gained

from the application of the signal decomposition were more obvious using the MLPNN compare to the ELM and RFR models.

According to the obtained results, we can report the following finding for futures studies:

1. In the future, we plan to extend the present modeling framework to a large dataset located in different climate regions.
2. This investigation provides a clear discussion about the factors affecting air relative humidity and the significant input combinations were selected. It is identified that, solar radiation, the minimal and maximal temperatures are the major's factors affecting relative humidity. Therefore, the predictive accuracy of the proposed models can be certainly improved by using optimizations algorithms and better selection of the studied input variables.
3. We also plan to investigate the possible other factors that can help in improving the predictive accuracies, especially the possible inclusion of other weathers variables.
4. It is highly recommended to explore new models and new signal decompositions algorithms to achieve the higher prediction accuracy.

However, at the end of the present study, some limitations of our methodology should be highlighted. It is worth mentioning that the performances of the hybrid models was significantly affected by the change in the meteorological input variables, and the high predictive accuracy was guaranteed with the inclusion of the variables having high coefficient of correlation with the air relative humidity. Secondly, it was found that the robustness and success of the signal decomposition algorithms could not be generalized and the results of each model will be different. For example, it was found that the VMD algorithm does not work accurately with all machine learning models. In summary, we argue that air relative humidity is affected by many external factors such as solar radiation, air temperature, and wind speed, making it highly nonlinear. It was demonstrated that single models often fail to correctly predict air relative humidity; consequently, it is worthwhile to explore the capability of the complex hybrid models in order to achieve high predictive accuracies.

Author contribution Conceptualization: Khaled Merabet and Salim Heddami; data curation: Khaled Merabet and Salim Heddami; formal analysis: Khaled Merabet and Salim Heddami; validation: Khaled Merabet and Salim Heddami; supervision: Khaled Merabet and Salim Heddami; writing original draft: Khaled Merabet and Salim Heddami; visualization: Khaled Merabet and Salim Heddami; investigation: Khaled Merabet and Salim Heddami. All authors have read and agreed to the published version of the manuscript.

Data availability The data presented in this study will be available on interested request from the corresponding author.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication All the authors have declared their consent to publish the manuscript.

Institutional Review Board statement Not applicable.

Informed consent Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Abdollahpoor R, Lotfivand N (2020) Fully adaptive denoising of ECG signals using empirical mode decomposition with the modified indirect subtraction and the adaptive window techniques. *Circ Syst Signal Process* 39(8):4021–4046. <https://doi.org/10.1007/s00034-020-01350-9>
- Abdulhay E, Alafeef M, Alzghoul L, Al Momani M, Al Abdi R, Arunkumar N, ... & de Albuquerque VHC (2020). Computer-aided autism diagnosis via second-order difference plot area applied to EEG empirical mode decomposition. *Neural Comput Appl* 32(15):10947–10956. <https://doi.org/10.1007/s00521-018-3738-0>.
- Adnan M, Adnan RM, Liu S, Saifullah M, Latif Y & Iqbal M (2021). Prediction of relative humidity in a high elevated basin of western Karakoram by using different machine learning models. *Weather Forecast* 59:. <https://doi.org/10.5772/intechopen.98226>.
- Alaodolehei B, Jafarian K, Sheikhan A, Beni HM (2020) Performance enhancement of an achalasia automatic detection system using ensemble empirical mode decomposition denoising method. *J Med Biol Eng* 40(2):179–188. <https://doi.org/10.1007/s40846-019-00497-4>
- Arulmozhi E, Basak JK, Sihalath T, Park J, Kim HT, Moon BE (2021) Machine learning-based microclimate model for indoor air temperature and relative humidity prediction in a swine building. *Animals* 11(1):222. <https://doi.org/10.3390/ani11010222>
- Bregaglio S, Donatelli M, Confalonieri R, Acutis M, Orlandini S (2010) An integrated evaluation of thirteen modelling solutions for the generation of hourly values of air relative humidity. *Theoret Appl Climatol* 102(3):429–438. <https://doi.org/10.1007/s00704-010-0274-y>
- Breiman L (2001) Random Forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. *IEEE Trans Signal Process* 62(3):531–544. <https://doi.org/10.1109/TSP.2013.2288675>
- Eccel E (2012) Estimating air humidity from temperature and precipitation measures for modelling applications. *Meteorol Appl* 19(1):118–128. <https://doi.org/10.1002/met.258>
- El Bouny L, Khalil M, Adib A (2019) ECG signal filtering based on CEEMDAN with hybrid interval thresholding and higher order statistics to select relevant modes. *Multimedia Tools Appl* 78(10):13067–13089. <https://doi.org/10.1007/s11042-018-6143-x>
- Gilles J (2013) Empirical wavelet transform. *IEEE Trans Signal Process* 61(16):3999–4010. <https://doi.org/10.1109/TSP.2013.2265222>
- Gunawardhana LN, Al-Rawas GA, Kazama S (2017) An alternative method for predicting relative humidity for climate change studies. *Meteorol Appl* 24(4):551–559. <https://doi.org/10.1002/met.1641>

- Hanoon MS, Ahmed AN, Zaini NA, Razzaq A, Kumar P, Sherif M, ... & El-Shafie A (2021). Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia. *Sci Rep* 11(1):1–19. <https://doi.org/10.1038/s41598-021-96872-w>
- Haykin S (1999) *Neural networks a comprehensive foundation*. Prentice Hall, Upper Saddle River
- Hornik K (1991) Approximation capabilities of multilayer feedforward networks. *Neural Netw* 4(2):251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T). 485
- Hu J, Wang J, Ma K (2015) A hybrid technique for short-term wind speed prediction. *Energy* 81:563–574. <https://doi.org/10.1016/j.energy.2014.12.074>
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceed R Soc London: Series A: Math, Physical Eng Sci* 454(1971):903–995
- Huang GB, Chen L, Siew CK (2006a) Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Networks* 17(4):879–892. <https://doi.org/10.1109/TNN.2006.875977>
- Huang GB, Zhu QY, Siew CK (2006b) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Hussein EA, Ghaziagar M, Thron C, Vaccari M, Bagula A (2021) Basic statistical estimation outperforms machine learning in monthly prediction of seasonal climatic parameters. *Atmosphere* 12(5):539. <https://doi.org/10.3390/atmos12050539>
- Jamei M, Bailek N, Bouchouicha K, Hassan MA, Elbeltagi A, Kuriqi A, ... & El-Kenawy ESM (2023). Data-driven models for predicting solar radiation in semi-arid regions. *Comput Mater Continua* 74(1):1625–1640. <https://doi.org/10.32604/cmc.2023.031406>
- Kisi O, Mirboluki A, Naganna SR, Malik A, Kuriqi A, Mehraein M (2022) Comparative evaluation of deep learning and machine learning in modelling pan evaporation using limited inputs. *Hydrol Sci J* 67(9):1309–1327. <https://doi.org/10.1080/02626667.2022.2063724>
- Kuang W (2020) Seasonal variation in air temperature and relative humidity on building areas and in green spaces in Beijing, China. *Chin Geogr Sci* 30(1):75–88. <https://doi.org/10.1007/s11769-020-1097-0>
- Kumari A, Kumar A, Kumar M, Kuriqi A (2022) Modeling average grain velocity for rectangular channel using soft computing techniques. *Water* 14(9):1325. <https://doi.org/10.3390/w14091325>
- Li G, Chen K, Yang H (2022) A new hybrid prediction model of cumulative COVID-19 confirmed data. *Process Saf Environ Prot* 157:1–19. <https://doi.org/10.1016/j.psep.2021.10.047>
- Liu H, Yu C, Wu H, Duan Z, Yan G (2020) A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting. *Energy* 202:117794. <https://doi.org/10.1016/j.energy.2020.117794>
- Liu Y, Feng G, Tsui KL, Sun S (2021) Forecasting influenza epidemics in Hong Kong using Google search queries data: a new integrated approach. *Expert Syst Appl* 185:115604. <https://doi.org/10.1016/j.eswa.2021.115604>
- Lu T, Viljanen M (2009) Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural Comput Appl* 18(4):345–357. <https://doi.org/10.1007/s00521-008-0185-3>
- Peng T, Zhang C, Zhou J, Nazir MS (2020) Negative correlation learning-based RELM ensemble model integrated with OVMD for multi-step ahead wind speed forecasting. *Renew Energy* 156:804–819. <https://doi.org/10.1016/j.renene.2020.03.168>
- Qadeer K, Ahmad A, Qyyum MA, Nizami AS, Lee M (2021) Developing machine learning models for relative humidity prediction in air-based energy systems and environmental management applications. *J Environ Manag* 292:112736. <https://doi.org/10.1016/j.jenvman.2021.112736>
- Sarkar P, Kumar P, Vishwakarma DK, Ashok A, Elbeltagi A, Gupta S, Kuriqi A (2022) Watershed prioritization using morphometric analysis by MCDM approaches. *Ecol Inform* 70:101763. <https://doi.org/10.1016/j.ecoinf.2022.101763>
- Sein ZMM, Ullah I, Iyakaremye V, Azam K, Ma X, Syed S, Zhi X (2022) Observed spatiotemporal changes in air temperature, dew point temperature and relative humidity over Myanmar during 2001–2019. *Meteorol Atmos Phys* 134(1):1–17. <https://doi.org/10.1007/s00703-021-00837-7>
- Shi X, Lu W, Zhao Y, Qin P (2018) Prediction of indoor temperature and relative humidity based on cloud database by using an improved BP neural network in Chongqing. *IEEE Access* 6:30559–30566. <https://doi.org/10.1109/ACCESS.2018.2844299>
- Shrestha AK, Thapa A & Gautam H (2019). Solar radiation, air temperature, relative humidity, and dew point study: Damak, jhapa, Nepal. *Int J Photoenergy* 2019: <https://doi.org/10.1155/2019/8369231>.
- Si Y, Zhang Z, Kong L, Li S, Wang Q, Kong C, Li Y (2019) Aging condition identification of viscoelastic sandwich structure based on empirical wavelet transform and Hilbert envelope demodulation. *Compos Struct* 215:13–22. <https://doi.org/10.1016/j.compstruct.2019.02.039>
- Suradhaniwar S, Kar S, Durbha SS, Jagarlapudi A (2021) Time series forecasting of univariate agrometeorological data: a comparative performance evaluation via one-step and multi-step ahead forecasting strategies. *Sensors* 21(7):2430. <https://doi.org/10.3390/s21072430>
- Tao H, Awadh SM, Salih SQ, Shafik SS, Yaseen ZM (2022) Integration of extreme gradient boosting feature selection approach with machine learning models: application of weather relative humidity prediction. *Neural Comput Appl* 34(1):515–533. <https://doi.org/10.1007/s00521-021-06362-3>
- Thapliyal PK, Shukla MV, Bisht JH, Pal PK, Navalgund RR (2014) Improvement in the retrieval of humidity profiles using hybrid regression technique from infrared sounder data: a simulation study. *Meteorol Appl* 21(2):301–308. <https://doi.org/10.1002/met.1330>
- Wang J, Hu J (2015) A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (Gaussian process regression) model. *Energy* 93:41–56. <https://doi.org/10.1016/j.energy.2015.08.045>
- Yasar A, Simsek E, Bilgili M, Yucel A, Ilhan I (2012) Estimation of relative humidity based on artificial neural network approach in the Aegean Region of Turkey. *Meteorol Atmos Phys* 115(1):81–87. <https://doi.org/10.1007/s00703-011-0168-2>
- Zhang Y, Pan G, Chen B, Han J, Zhao Y, Zhang C (2020) Short-term wind speed prediction model based on GA-ANN improved by VMD. *Renew Energy* 156:1373–1388. <https://doi.org/10.1016/j.renene.2019.12.047>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.