Comparative Study on Univariate Forecasting Methods for Meteorological Time Series

Thi-Thu-Hong PHAN *†
†Department of Computer Science
Vietnam National University of Agriculture
Hanoi, Vietnam

Émilie POISSON CAILLAULT*, André BIGAND*
*LISIC
Univ. Littoral Côte d'Opale, EA 4491-LISIC
F-62228 Calais, France

***Abstract*—Time series forecasting has an important role in many real applications in meteorology and environment to understand phenomena as climate change and to adapt monitoring strategy. This paper aims first to build a framework for forecasting meteorological univariate time series and then to carry out a performance comparison of different univariate models for forecasting task. Six algorithms are discussed: Single exponential smoothing (SES), Seasonal-naive (Snaive), Seasonal-ARIMA (SARIMA), Feed-Forward Neural Network (FFNN), Dynamic Time Warping-based Imputation (DTWBI), Bayesian Structural Time Series (BSTS). Four performance measures and various meteorological time series are used to determine a more customized method for forecasting. Through experiments results, FFNN method is well adapted to forecast meteorological univariate time series with seasonality and no trend in consideration of accuracy indices and DTWBI is more suitable as considering the shape and dynamics of forecast values.**

***Keywords*—Univariate time series forecasting; similarity measure; SARIMA; FFNN; BSTS; DTW**

I. Introduction

Time series forecasting is a matter of great importance in numerous domains [1], [2]. In particular, forecasting hydro-meteorological data plays a key step to better understand climate change, environmental change, and then to adapt monitoring strategy, to deploy preventive or corrective actions. This means to define how past events affect future events. But this task is a remaining challenge because hydro-meteorological data are impacted by diverse phenomena and factors from the environment.

Classical methods for forecasting hydro-meteorological time series were investigated to address the issue of linear models [2] such as linear regression, Exponential Smoothing (ES) or model fitting approaches based on moving average. Autoregressive Integrated Moving Average (ARIMA) is one of the most commonly model for this task [3]–[5]. Mahmud et al. [6] investigated seasonal ARIMA model to monthly predict rainfall for 12 future months considering thirty rainfall stations in Bangladesh. Nury et al. [7] employed SARIMA to forecast future values of temperatures in the Sylhet Division of Bangladesh. The authors showed that the SARIMA is a powerful model for short-term forecasting of the two meteorological variables max. and min. temperature. In [8], Li et al. proposed Hadoop-based ARIMA algorithm to forecast weather.

These methods are well adapted to predict generic trends. However, they are not able (i) to determine nonlinear features in data and (ii) to predict quick change inside the process. In the three past decades, numerous approaches have been proposed to improve accuracy and efficiency of time series forecasting, especially using nonlinear models. Cheng et al. [9] pointed out that nonlinear models outperform linear ones for time series forecasting in many applications, such as stock prices [1] and climatology [10].

Artificial Neural Networks (ANN) have become a useful approach to model nonlinear processes such as forecasting rainfall [11], [12], or predicting sea level [13]. In [14], Hung et al. investigated feed-forward neural network model and compared it with a simple persistent method for hourly rainfall forecasting (from 75 rain gauge stations) in Bangkok, Thailand. The results showed that FFNN model illustrated better ability to predict rainfall. Chattopadhyay and Chattopadhya [15] performed a comparison of traditional statistical autoregressive models and autoregressive NN model for univariate prediction of rainfall time series. The results of these studies present the improved performance of NN model when comparing it with the traditional statistical approaches.

Dynamic Time Warping (DTW) [16] is an effective method for measuring similarity between two linear/nonlinear time series. This method is successfully applied in pattern recognition [17], [18], in imputation [19]. For the forecasting task, there are few studies using DTW to predict future values. In [20] Tsinaslanidis and Kugiumtzis used perceptually important points and DTW for stock market forecasting.

Compared to other methods, only few research has been devoted to predict time series using Bayesian network-based, although Aguilera et al. showed the capability of Bayesian networks in environmental modeling in [21].

Thus this paper does not propose a novel forecasting method. However we emphasize on comparing the performances of different univariate approaches by building a framework for forecasting hydro-meteorological univariate time series. Five time series data are applied to the six models we choose for anticipating future values including SES, Snaive, SARIMA, FFNN, DTWBI, and BSTS. This allows to suggest the most suitable method, among the above-mentioned methods, for predicting hydro-meteorological univariate time series ensuring that results are reliable and high quality.

In addition, for univariate forecasting methods, we must only rely on the available values of this unique variable

ISBN 978-90-827970-1-5 © EURASIP 2018

to estimate future values, without other outside explanatory variables [22]. And, Smith and Agrawal [23] pointed out that "when attempting to forecast univariate time series data, it is generally accepted that parsimonious model techniques are followed".

This paper is organized as follows. Section 2 focuses on univariate forecasting methods. Next, Section 3 introduces our experiments protocol. Results and discussion for forecasting meteorological univariate time series are provided in Section 4. Finally, conclusions are drawn and future work is presented.

II. Time series forecasting methods

In this part, several adapted methods for forecasting meteorological univariate time series are mentioned and then will be deployed.

- SES - Simple Exponential Smoothing: ES methods, including a number of ad hoc techniques, used for extrapolating different types of univariate time series. The new forecast at time $t+1$ is the exponentially weighted average of all t past observations: $y_1, y_2, \ldots, y_t$ [2].
$$y_{t+1|t} = \sum_{n=0}^{t} \alpha(1-\alpha)^n y_{t-n} \quad (1)$$
where $0 \leq \alpha \leq 1$
- Snaive - Seasonal-naive: sets all the forecast values to be the value of the last observation and takes into account the seasonal period as eq.2. Hence, this method considers that the most current observed value is the only important one and all the previous observations do not provide information to estimate future values.
$$y_{t+h} = y_{t+h-km} \quad (2)$$
where m is a seasonal period, $k = 1 + (h-1)/m$, h is a number of periods for forecasting.
- SARIMA - Seasonal-ARIMA: the forecasted values of a stationary time series can be estimated by an additive linear function composed of p past observations (autoregressive) and q random errors (moving average) as eq.3, denoted as ARIMA(p, d, q) [2], and d is the differencing number used to make a series y to be stationary.
$$y_t = \sum_{n=1}^{p} \alpha_n \times y_{t-n} + \epsilon_t + \sum_{n=1}^{q} \beta_n \times \epsilon_{t-n} \quad (3)$$
Seasonal ARIMA model is developed from ARIMA by taking into account seasonal factors. SARIMA is labeled as SARIMA$(p, d, q)(P, D, Q)s$, where upper-cases are counterpart of ARIMA model for the seasonal model and s is number of periods per season.
- BSTS - Bayesian Structural Time Series: This model applies Markov Chain Monte Carlo (MCMC) to sample the posterior distribution of a Bayesian structural time series model. The model involves 3 major steps:
Kalman filter: This step consists in decomposing a time series. Various state variables such as trend, seasonality, regression can be added in this step.
Spike-and-slab: This step selects the most important regression predictors.
Bayesian model averaging: This step combines the results and calculates prediction values.
- DTWBI: In a previous study [19], we proposed DTWBI approach for completing missing values. Here, we consider forecasting values as missing data, and then we apply DTWBI method to compute these future values. Forecasting process is based on past values. This is fully compatible with DTWBI approach that fills missing values according to the recorded data.
The approach consists in finding the most similar sub-sequence Qs to a query Q (the sub-sequence before the predicted position) by sliding windows based on the combination of shape-feature extraction algorithm [24] and DTW. This allows some distortions both in the temporal and value axis. Once the most similar window is identified, the following sub-sequence Qfs of the Qs is considered as the forecast values. The dynamics and the shape of data before the forecast values are key-point of this technique (see [19]).
- FFNN - Feed-forward neural network: Artificial Neural Network is proposed from inspiring the interconnection neurons of the human. FFNN maps the set of inputs to the set of outputs (both data inputs and outputs are digital). FFNN allows to automatically extract global features before the last decision step (output layer) considering only one hidden layer. A FFNN with no hidden layers is also called linear perceptron: its inputs are directly mapped to the outputs unit via the weighted connections.

III. Experiments protocol

We have conducted a set of experiments on five meteorological time series using six different univariate models to evaluate their forecasting performance. R language is used to compute all experiments. We utilize the latest R-packages of forecast [25] (for FFNN, Ses and Snaive), astsa [26] for SARIMA, bsts [27] (for BSTS). For DTWBI, we develop ourselves (upon request). For SARIMA, auto.arima() [25] is employed to optimize the parameters p, d, q, P, D, Q. For FFNN, we use the default parameters: input nodes are the number of seasonal lags applied to seasonally adjusted data, and number of nodes in the hidden layer is half the number of input nodes. And for BSTS, we choose $niter = 50$ with specified seasonality component for each forecasting rate.

A. Data presentation

In this section, we describe the data used for the study. Five hydro-meteorological time series are used for evaluating the performance of forecasting methods (table I). These five datasets were collected at three meteorological stations in Vietnam. They have different sampling frequency and time measurement duration (short or long period). In order to obtain useful information from the datasets and to make the datasets easily exploitable, we analyzed these series. Table I summarizes their characteristics. All the five datasets have

ISBN 978-90-827970-1-5 © EURASIP 2018

a seasonality component (i.e. an annual cycle), without any linear trend.

TABLE I
CHARACTERISTICS OF TIME SERIES

N0	Dataset name	Period	# Samples	Frequency
1	Ba Tri humidity	2003-2007	7,304	6 hours
2	Ba Tri air temperature	2003-2007	7,304	6 hours
3	Cua Ong air temperature	1973-1999	9859	daily
4	Phu Lien humidity	1961-2015	692	monthly
5	Phu Lien air temperature	1961-2014	684	monthly

B. Experiments process

To assess the capacity of forecasting algorithms, we used a technique including three steps. In the first step, data segments are deleted from each time series with different size of consecutive data. In the second step, all forecasting algorithm are applied as mentioned above to estimate the forecast values. Finally, after forecasting data, four performance indicators are computed between the predicted segment and the deleted true values.

In this study, 5 forecasting data levels are considered on 5 datasets. For Phu Lien datasets with monthly sampling frequency, we predict 6, 12, 18, 24 and 30 future months. For the infra daily series, the forecasting size is ranged from 0.5% 0.75%, 1%, 1.25% and 1.5% of the dataset size. For each forecasting level, all the algorithms are conducted 5 times by back-warding the predicted position of each repetition with a size of forecasting. We then run 25 iterations for each dataset.

C. Performance indicator

After the prediction of future values, we compared the performance of six different forecasting methods based on four metrics described as follows:

1) Similarity define the percentage of similar values between the predicted values y and the actual values x. It is calculated by:

$$Sim(y,x) = \frac{1}{T}\sum_{i=1}^{T}\frac{1}{1+\frac{|y_i-x_i|}{\max(x)-\min(\mathrm{x})}} \tag{4}$$

Where T is the number of forecast values. A higher similarity (Sim value $\in [0,1]$) highlights a better ability method for the forecasting task.

2) NMAE: The Normalized Mean Absolute Error between the predicted values y and the actual ones x is computed as:

$$NMAE(y,x) = \frac{1}{T}\sum_{i=1}^{T}\frac{|y_i-x_i|}{V_{max}-V_{min}} \tag{5}$$

Where V_{max}, V_{min} are the maximum and the minimum values of input time series (time series used for forecasting). A lower NMAE value means better performance method for the prediction task.

3) RMSE: The Root Mean Square Error is defined as the average squared difference between the forecast values y and the respective true values x. This indicator is very useful for measuring overall precision or accuracy. In general, the most effective method would have the lowest RMSE.

$$RMSE(y,x) = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_i-x_i)^2} \tag{6}$$

4) FB (Fractional Bias): This parameter determines whether the predicted values are overestimated or underestimated relatively to those observed. A model is considered as perfect when its FB tends to zero, and as acceptable when $-0.3 \leq FB \leq 0.3$

$$FB(y,x) = 2 * \frac{\mathrm{mean}(y)-\mathrm{mean}(x)}{\mathrm{mean}(y)+\mathrm{mean}(x)} \tag{7}$$

IV. RESULTS AND DISCUSSION

Table II presents average results of different forecasting algorithms on 5 univariate time series for the 4 indicators. The best results for each forecasting rate are bold highlighted.

These results show that FFNN method demonstrates better performance for forecasting future data on Phu Lien temperature, Ba Tri humidity and Ba Tri temperature series: the highest similarity, the lowest NMAE and RMSE at every forecasting levels. The highest similarity (close to 1 with Sim $\in [0,1]$), lowest NMAE and RMSE highlight an improved capability for the forecasting task. The results illustrate that the forecast values generated from the FFNN method are close to the real values. However, when considering the FB index, the indicator presents the bias of estimated values with real values, the FFNN only yields the best results at some levels.

On Phu Lien temperature data, following the FFNN approach is DTWBI as predicting values from 6 to 30 months on the first 3 indices (Similarity, RMSE and NMAE). For FB index, DTWBI outperformed other methods for larger sizes of forecasting values, from 18 to 30 months. The third one is BSTS on this dataset for all indicators.

As reported in table II, in contrast to the three above datasets, BSTS method shows the best predictability on Sim, RMSE and NMAE measurements for all ratios on Phu Lien humidity. The second rank is SARIMA when considering the three indicators (excluding 2^{nd} level for Sim index).

In addition, all five series have a seasonality component, so we choose SARIMA to make a prediction. Although ARIMA is a benchmark method for the forecasting task and for each time series we use R function auto.arima() [25] to optimize parameters but with these time series this model does not illustrate its ability.

Looking at Cua Ong temperature dataset (table II), FFNN continues to demonstrate its predictability for meteorological univariate time series at the first two forecast levels (0.5% and 0.75%). But at higher ratios from level 3 to level 5, DTWBI proves its predictability: the largest value for Similarity, and the smallest value considering error and bias indices.

Ses and Snaive methods were proposed for forecasting data with seasonality or no trend. When considering accuracy indices, they yield quite good results (table II).

ISBN 978-90-827970-1-5

In this study, we also compare the visualization performance of forecasting values generated from different methods.

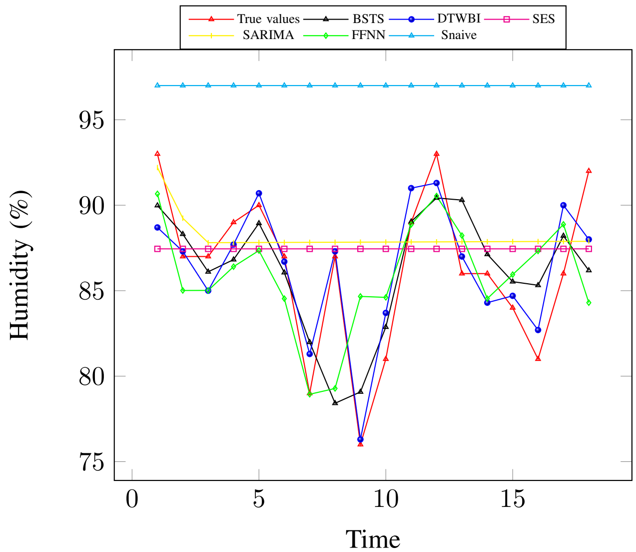

Fig. 1. True values and forecast values generated from different univariate methods on Phu Lien humidity series (forecast size of 18 months)

Figure 1 presents the shape of forecast values yielded by different methods on the Phu Lien humidity series. From this figure, it is clear that SES and Snaive methods do not produce a similar shape as the shape of true values. When comparing the quantitative indicators, DTWBI is only second or third rank, but when considering the shape of forecasting values, DTWBI is better than other methods. The dynamics and the form of predicted values produced by the DTWBI method are very similar to the form of true values.

In this paper, Cross-Correlation (CC) coefficients between the query and each sliding window (as defined in DTWBI method) are also calculated, and the maximum coefficient is computed. CC indicates the similarity of two series. For forecasting task, this coefficient demonstrates how past values affect future ones. High CC means that predicted values are close to past values. In table III, we see that CC coefficients are very high only for Phu Lien temperature series, (approximate 1). These CC values make it possible to explain why the predicted values (generated from DTWBI, FFNN, SARIMA and BSTS) and the actual values are nearly identical: similarity values are very high, error and bias indexes are very low.

From the above results and analysis, we suggest to use DTWBI approach for forecasting meteorological univariate time series when considering the shape of predicted values and to apply FFNN when regarding the quantitative accuracy.

V. Conclusions and Future Work

This paper proposes a framework for meteorological univariate time series forecasting. Quantitative performance of different methods are compared on 5 various datasets using 4 quantitative indicators (similarity, NMAE, RMSE and FB). The visual performance of these methods is also evaluated. The obtained results clearly demonstrate that FFNN yielded improved performance when considering accuracy of forecast values and DTWBI is more appropriate when regarding the shape and dynamics of predicted values for forecasting meteorological univariate time series. These results are original for hydro-meteorological univariate time series. The present work will allow to compare different type of univariate time series and to forecast multivariate time series in the future.

References

[1] C. Chatfield, *The Analysis of Time Series: An Introduction*. Chapman and Hall/CRC, Mar. 2016, google-Books-ID: qKzyAbdaDFAC.

[2] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443–473, Jan. 2006.

[3] I. Kaushik and S. M. Singh, "Seasonal ARIMA model for forecasting of monthly rainfall and temperature," *Journal of Environmental Research and Development*, vol. 3, no. 2, 2008.

[4] P. E. N. M. Momani and P. E. N. M. Momani, "Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis," *American Journal of Environmental Sciences*, vol. 5, no. 5, pp. 599–604, Oct. 2009.

[5] B. S. Kim, S. Z. Hossein, and G. Choi, "Evaluation of temporal-spatial precipitation variability and prediction using seasonal ARIMA model in Mongolia," *KSCE Journal of Civil Engineering*, vol. 15, no. 5, pp. 917–925, May 2011.

[6] I. Mahmud, S. H. Bari, and M. T. U. Rahman, "Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method," *Environmental Engineering Research*, vol. 22, no. 2, pp. 162–168, Nov. 2016.

[7] A. H. Nury, M. Koch, and M. J. B. Alam, "Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh," in *4th International Conference on Environmental Aspects of Bangladesh (ICEAB), August*, 2013, pp. 24–26.

[8] L. Li, Z. Ma, L. Liu, and Y. Fan, "Hadoop-based ARIMA Algorithm and its Application in Weather Forecast," *International Journal of Database Theory and Application*, vol. 6, no. 5, pp. 119–132, Oct. 2013.

[9] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. J. Kong, and S. T. S. Bukkapatnam, "Time series forecasting for nonlinear and non-stationary processes: A review and comparative study," *IIE Transactions*, vol. 47, no. 10, pp. 1053–1071, Oct. 2015.

[10] M. Mudelsee, *Climate Time Series Analysis*. Springer, 2013.

[11] T. Mandal and V. Jothiprakash, "Short-term rainfall prediction using ANN and MT techniques," *ISH Journal of Hydraulic Engineering*, vol. 18, no. 1, pp. 20–26, Mar. 2012.

[12] A. El-Shafie, A. Noureldin, M. Taha, A. Hussain, and M. Mukhlisin, "Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia," *Hydrology and Earth System Sciences*, vol. 16, no. 4, pp. 1151–1169, Apr. 2012.

[13] M. Imani, R.-J. You, and C.-Y. Kuo, "Caspian Sea level prediction using satellite altimetry by artificial neural networks," *International Journal of Environmental Science and Technology*, vol. 11, no. 4, pp. 1035–1042, May 2014.

[14] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand," *Hydrology and Earth System Sciences*, vol. 13, no. 8, pp. 1413–1425, 2009.

[15] S. Chattopadhyay and G. Chattopadhyay, "Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN," *Comptes Rendus Geoscience*, vol. 342, no. 2, pp. 100–107, Feb. 2010.

[16] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 16, pp. 43–49, 1978.

[17] D. J. Bemdt and Clifford, "Using Dynamic Time Warping to find patterns in time series," 1994.

[18] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognition*, vol. 44, no. 9, pp. 2231–2240, Sep. 2011.

ISBN 978-90-827970-1-5

TABLE II
PERFORMANCE INDEXES OF VARIOUS FORECASTING ALGORITHMS ON 5 DATASETS (**BEST RESULTS IN BOLD**)

Metrhod		Phu Lien temperature				Phu Lien humidity				Ba Tri humidity				Ba Tri temperature			
	Size	Sim	NMAE	RMSE	FB	Sim	NMAE	RMSE	FB	Sim	NMAE	RMSE	FB	Sim	NMAE	RMSE	FB
DTWBI	6	0.922	0.06	1.29	-0.03	0.76	0.12	5.5	**0**	0.85	0.13	11.75	0.02	0.83	0.08	23.75	-0.01
FFNN		**0.93**	**0.05**	**1.18**	-0.02	0.75	0.11	4.97	0.02	**0.89**	**0.08**	**6.46**	**0**	**0.9**	**0.04**	**12.19**	0.01
SARIMA		0.88	0.09	1.78	-0.03	0.77	0.11	4.71	0.02	0.87	0.09	8.03	0.01	0.79	0.09	25.76	0.04
BSTS		0.921	0.06	1.18	**0**	**0.8**	**0.08**	**3.51**	0.01	0.81	0.16	13.14	-0.01	0.77	0.11	35.23	**0**
ses		0.76	0.26	6.2	-0.01	0.78	0.11	4.84	0.02	0.83	0.14	11.75	0.02	0.81	0.08	25.21	0.01
snaive		0.76	0.26	6.2	-0.01	0.69	0.17	6.99	0.06	0.78	0.18	14.55	-0.03	0.77	0.11	32.35	0.02
DTWBI	12	0.925	0.07	1.63	-0.02	0.8	0.12	5.86	-0.02	0.84	0.14	12.71	0.02	0.82	0.08	27.09	0.03
FFNN		**0.94**	**0.05**	**1.24**	**0**	0.83	0.1	4.5	0.01	**0.9**	**0.08**	**6.62**	**0**	**0.86**	**0.06**	**17.59**	0.02
SARIMA		0.92	0.07	1.63	-0.01	0.82	0.1	4.71	0.01	0.87	0.11	8.81	0.01	0.81	0.08	24.37	0.03
BSTS		0.92	0.07	1.48	**0**	**0.86**	**0.08**	**3.64**	**0**	0.78	0.21	17.1	0.05	0.8	0.09	28.5	**0.01**
ses		0.71	0.35	7.23	-0.26	0.83	0.1	4.73	0.01	0.83	0.14	11.9	0.01	0.82	0.08	25.29	**0.01**
snaive		0.71	0.35	7.23	-0.26	0.8	0.11	5.48	0.02	0.8	0.19	15.56	0	0.78	0.11	34.51	**0.01**
DTWBI	18	0.93	0.06	1.49	**0**	0.83	0.11	5.48	-0.02	0.84	0.14	13.1	0.03	0.84	0.08	25.46	0.03
FFNN		**0.94**	**0.06**	**1.3**	**0**	0.83	0.1	4.67	-0.03	**0.88**	**0.1**	**8.59**	-0.01	**0.89**	**0.05**	**15.95**	**0**
SARIMA		0.92	0.08	1.82	-0.01	0.84	0.1	4.67	**0**	0.87	0.11	9.55	**0**	0.83	0.08	24.27	0.02
BSTS		0.93	0.07	1.54	0.02	**0.88**	**0.07**	**3.41**	-0.01	0.75	0.25	18.85	-0.13	0.72	0.17	52.62	-0.03
ses		0.76	0.31	7.18	0.01	0.84	0.1	4.81	-0.01	0.83	0.15	12.81	**0**	0.84	0.07	23.46	**0**
snaive		0.76	0.31	7.18	0.01	0.8	0.14	6.15	0.03	0.75	0.25	19.56	-0.06	0.78	0.11	33.82	0.01
DTWBI	24	**0.94**	0.06	1.45	**0**	0.83	0.12	5.8	-0.03	0.85	0.15	13.87	-0.01	0.84	0.08	24.75	0.03
FFNN		**0.94**	**0.05**	**1.24**	-0.01	0.85	0.11	5.08	0.01	**0.89**	**0.09**	**8.61**	**0**	**0.87**	**0.06**	**18.49**	**0**
SARIMA		0.91	0.08	1.8	-0.01	0.85	0.1	4.95	0.01	0.86	0.12	10.36	0	0.84	0.08	23.64	0.02
BSTS		0.92	0.08	1.67	0.01	**0.87**	**0.09**	**3.85**	**-0.01**	0.76	0.25	20.42	-0.06	0.81	0.1	30.91	-0.02
ses		0.74	0.31	6.55	-0.22	0.84	0.11	5.03	0.01	0.83	0.16	13.26	0	0.84	0.07	23.71	0.01
snaive		0.74	0.31	6.55	-0.22	0.84	0.11	5.15	0.02	0.79	0.22	19.05	-0.04	0.81	0.1	31.73	-0.01
DTWBI	30	0.93	0.07	1.6	**0**	0.84	0.11	5.34	**-0.01**	0.89	0.11	10.46	0.01	0.86	0.07	21.9	0.03
FFNN		**0.94**	**0.05**	**1.27**	-0.01	0.84	0.12	5.69	**0.01**	**0.91**	**0.08**	**7.91**	-0.01	**0.89**	**0.05**	**16.52**	0.01
SARIMA		0.91	0.08	1.8	-0.01	0.85	0.11	5.08	**0.01**	0.86	0.13	10.84	**0**	0.84	0.08	24.06	0.02
BSTS		0.92	0.07	1.66	0.01	**0.86**	**0.1**	**4.62**	0.02	0.77	0.24	19.56	-0.06	0.77	0.13	40.17	0.01
ses		0.77	0.29	6.49	0.04	0.85	0.11	5.28	**0.01**	0.83	0.17	13.77	**0**	0.84	0.08	24.8	**0**
snaive		0.77	0.29	6.49	0.04	0.85	0.11	5.56	0.03	0.82	0.19	16.81	0.05	0.82	0.09	30.07	-0.04

Cua Ong temperature

	Size	0.05				0.075				0.1				0.125			
DTWBI		0.79	0.09	30.29	-0.02	0.82	0.11	34.59	**0**	**0.83**	**0.11**	**35.21**	**-0.04**	**0.85**	**0.11**	**34.34**	**-0.06**
FFNN		**0.83**	**0.07**	**24.5**	0.05	**0.84**	**0.09**	**28.85**	0.08	0.82	0.13	40.66	0.1	0.83	0.14	45.77	0.07
SARIMA		0.7	0.14	42.12	**-0.03**	0.76	0.14	43.61	**0**	0.78	0.14	43.21	0.01	0.78	0.15	44.71	-0.04
BSTS		0.52	0.65	202.04	0.65	0.43	0.6	188.4	-0.8	0.61	0.5	143.9	-0.89	0.48	0.5	155.7	0.042
ses		0.79	0.1	31.8	0.04	**0.84**	0.1	30.85	**0**	0.81	0.14	42.95	0.11	0.8	0.16	48.97	0.04
snaive		0.79	0.1	31.8	0.04	**0.84**	0.1	30.85	**0**	0.81	0.14	42.95	0.11	0.8	0.16	48.97	0.04
DTWBI	0.15	**0.86**	**0.11**	**35.1**	**-0.05**												
FFNN		0.85	0.13	41.57	0.09												
SARIMA		0.8	0.16	49.46	-0.02												
BSTS		0.68	0.53	199.1	0.2												
ses		0.8	0.18	55.3	0.06												
snaive		0.8	0.18	55.3	0.06												

This work was kindly supported by the Ministry of Education and Training Vietnam International Education Development FEDER - the region Hauts-de-France (CPER 2014-2020 MARCO) and carried out using the CALCULCO computing platform, supported by SCoSI/ULCO.

TABLE III
THE MAXIMUM OF CROSS-CORRELATION BETWEEN THE QUERY AND SLIDING WINDOWS.

Size	#1	#2	Size (%)	#3	#4	#5
6	0.997	0.872	0.5	0.93	0.91	0.78
12	0.988	0.89	0.75	0.91	0.9	0.76
18	0.979	0.84	1	0.87	0.88	0.75
24	0.977	0.8	1.25	0.86	0.87	0.76
30	0.974	0.75	1.5	0.86	0.85	0.78

#1-Phu Lien temperature, #2-Phu Lien humidity, #3-Ba Tri temperature, #4-Ba Tri humidity, #5-Cua Ong temperature

[19] T.-T.-H. Phan, É. Poisson Caillault, A. Lefebvre, and A. Bigand, "Dynamic time warping-based imputation for univariate time series data," *Pattern Recognition Letters*, Aug. 2017.

[20] P. E. Tsinaslanidis and D. Kugiumtzis, "A prediction scheme using perceptually important points and dynamic time warping," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6848–6860, Nov. 2014.

[21] P. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón, "Bayesian networks in environmental modelling," *Environmental Modelling & Software*, vol. 26, no. 12, pp. 1376–1388, Dec. 2011.

[22] C. Frank, A. Garg, L. Sztandera, and A. Raheja, "Forecasting women's apparel sales using mathematical modeling," *International Journal of Clothing Science and Technology*, vol. 15, no. 2, pp. 107–125, 2003.

[23] M. Smith and R. Agrawal, "A Comparison of Time Series Model Forecasting Methods on Patent Groups." in *MAICS*, 2015, pp. 167–173.

[24] T. T. H. Phan, E. P. Caillault, and A. Bigand, "Comparative study on supervised learning methods for identifying phytoplankton species," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*. IEEE, Jul. 2016, pp. 283–288.

[25] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 26, no. 3, pp. 1–22, 2008.

[26] D. Stoffer, *astsa: Applied Statistical Time Series Analysis*, 2017, r package version 1.8. [Online]. Available: https://CRAN.R-project.org/package=astsa

[27] S. L. Scott, *bsts: Bayesian Structural Time Series*, 2017, r package version 0.7.1. [Online]. Available: https://CRAN.R-project.org/package=bsts