



**HAL**  
open science

## **A vector autoregressive methodology for short-term weather forecasting: tests for Lebanon**

Wissam Abdallah, Nassib Abdallah, Jean-Marie Marion, Mohamad Oueidat,  
Pierre Chauvet

► **To cite this version:**

Wissam Abdallah, Nassib Abdallah, Jean-Marie Marion, Mohamad Oueidat, Pierre Chauvet. A vector autoregressive methodology for short-term weather forecasting: tests for Lebanon. SN Applied Sciences, 2020, 2 (9), pp.1555. 10.1007/s42452-020-03292-y . hal-03525823

**HAL Id: hal-03525823**

**<https://uco.hal.science/hal-03525823v1>**

Submitted on 23 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# A vector autoregressive methodology for short-term weather forecasting: tests for Lebanon

Wissam Abdallah<sup>1</sup> · Nassib Abdallah<sup>1</sup> · Jean-Marie Marion<sup>1</sup> · Mohamad Oueidat<sup>1</sup> · Pierre Chauvet<sup>1</sup>

Received: 19 May 2020 / Accepted: 3 August 2020 / Published online: 25 August 2020  
© Springer Nature Switzerland AG 2020

## Abstract

Weather forecasting has been a major challenge due to the uncertain nature of the weather. Numerical models, such as the “Action de Recherche Petite Echelle Grande Echelle” (ARPEGE), the “Global Forecasting System”, the “European Center for Medium-Range Weather Forecasts”, are widely adopted by many meteorological services to forecast weather parameters. Under certain conditions, numerical models may have lower forecast accuracy, which is due to several factors such as the chaotic nature of the partial differential equations that simulate the evolution of the atmosphere and the difficulty of forecasting over countries with a fairly small surface area but with a very varied relief. This paper proposes a time series analysis approach based on the vector autoregression model (VAR) as an alternative and robust solution. The results are very promising (average of about 96.67% of precision between real values and three predicted parameters: temperature minimum, maximum humidity and precipitation) in the field of short-term weather parameter forecasting. In addition, the use of VAR models has solved the major problem posed by the chaotic equations of the ARPEGE model with greater accuracy on the one hand, and the execution time, forecasting accuracy and robustness of the SARIMA univariate models on the other.

**Keywords** Time series analysis TSA · Vector autoregression VAR · Autoregressive moving average ARMA · Weather prediction model

## 1 Introduction

The World Meteorological Organization (WMO) has called for the integration of efforts needed to improve the accuracy of weather forecasting [1].

Indeed, the development of effective weather prediction models has always been particularly difficult and demanding for meteorological services [2]. For example, in Lebanon, which was the study area for this work, the forecast model used by the Civil Aviation Meteorological Service at Beirut Rafik Hariri International Airport (BRHIA) is the ARPEGE model (0.5). Unfortunately, the forecasts provided by ARPEGE have often been erroneous and biased. In the case of Lebanon, the numerical model ARPEGE (0.5) covers certain regions with different climatic

characteristics, such as the Bekaa and Mount Lebanon. In fact, the forecast made by ARPEGE for the Bekaa region is the same as that for the Mount Lebanon region since they are in the same grid; however, the characteristics of the two regions are very different and the results of the ARPEGE forecast are therefore considered to be erroneous.

Temperature forecasting is used in agriculture and hydrology. Precipitation is of interest in Lebanon because the Bekaa plain and the northern part of the country are affected by heavy rainfall. In 2018–2019, floods caused by heavy rainfall led the government to establish a committee of experts to predict and avoid natural disasters. Wind speed and direction are also of interest.

Based on the above, this paper presents an alternative to the numerical model for forecasting all meteorological

✉ Wissam Abdallah, [wissam.abdallah@etud.univ-angers.fr](mailto:wissam.abdallah@etud.univ-angers.fr); Nassib Abdallah, [nassib.abdallah@uco.fr](mailto:nassib.abdallah@uco.fr) | <sup>1</sup>Universite d'Angers, Angers, France.



parameters using vector autoregression. The first section discusses the related work. The second section discusses the methodology for building the forecast model. The third section examines the results and compares them with the existing ARPEGE (0.5) tool. Finally, the last section concludes the work and gives the perspective of this work.

## 2 Related works

Weather forecasting can be carried out using many techniques. However, we can separate forecast models into two independent approaches: numerical weather prediction and forecasting based on time series data.

Studies have found that the physics-based models are generally more accurate over longer horizon, while time series methods are more accurate at short horizons. At intermediate horizons, combined models (i.e., physics models adjusted using statistical techniques) are often more accurate [3].

Numerical models like ARPEGE are able to guide meteorologists beyond the first hours of many future weather phenomena such as thunderstorms, snow, cold fronts, warm fronts [4, 5]. As well, numerical models have a particular advantage for forecasters as they can predict a very complex weather phenomenon that propagates on a very large scale, such as tornadoes and turbulence [3]. However, the fundamental problems of numerical models such as ARPEGE appear in their inability to solve the chaotic equations that model some of the atmospheric phenomena such as tornadoes, turbulence and in the low accuracy of forecasts over a geographically defined area [6].

Recently, the cost and complexity of weather forecasting has prompted many researchers to find cheaper and more effective alternatives to short- or long-term forecasts. The prediction models were introduced in the 1930s, with the first univariate models (autoregressive models (AR), moving average models (MA), then autoregressive integrated moving average (ARIMA)). In these models, the history of a variable is used to make predictions [3], but the univariate approach ignores information that could be exploitable from other time series. In order to overcome the limitations of Keynesian Macroeconometric models, Sims [7] proposed a multivariate model: the VAR representation. Doan et al. [8] used forecasting procedure based on a Bayesian method for estimating vector autoregressions.

Today, these models are largely used, alone or in combination with other techniques such as modeling dynamic with artificial neural networks, for example the model vector autoregressive neural network

For that, time series analysis (TSA) has become very useful in the field of forecasting as it is applied in several fields such as weather forecasting, economics, engineering,

environment, medicine [9]. In fact, TSA helps to predict future weather situations by collecting historical data at regular intervals called series [10], these data are collected at specific intervals (hourly, daily, weekly, monthly, yearly, etc.). TSA techniques, such as ARIMA and SARIMA, offers some advantages in terms of reliable results in weather forecasting [9].

Rodrigues et al. [9], in 2017, worked on rainfall prediction using the ARMA model and multiple linear regression. Their models gave an average accuracy of 82% and 87.3%, respectively.

In addition, work published by Kaul et al. [11] in 2017 showed a comparison between data mining techniques in weather forecasting, such as classification and regression trees (CART), artificial neural network (ANN), K-means clustering (KNN). They obtained 85%, 90% and 92% for temperature forecasting, respectively.

Later in 2018, Thi-Thu-Hong et al. [12] conducted a comparative study on univariate time series weather forecasting methods for humidity prediction. They used time-series analysis methods such as seasonal autoregressive integrated moving average models (SARIMA) and artificial neural network (ANN). SARIMA gave an average accuracy of 92% and was more efficient than the results obtained with the other techniques studied: The artificial neural network (ANN) gave an average precision of 91%, dynamic time distortion-based imputation (DTWBI) gave an average precision of 91%, Bayesian structural time series (BSTS) gave an average precision of 88%, single exponential smoothing (SES) gave an average precision of 87%, and finally the Seasonal Naive Technique (SNaive) gave an average precision of 86%.

Furthermore, Kunjumon et al. [13], in 2018, conducted a comparative study of several supervised and unsupervised machine learning techniques for rainfall forecasting. They used the artificial neural network (ANN) and support vector machine (SVM), the K-medoid algorithm and the decision tree (DT) classification algorithm. The SVM gave the best results with an average accuracy of 90%. However, the model execution time was very long.

In 2019, Zabbah et al. [14] used artificial neural network techniques to predict precipitation. Their model combines neural networks with linear and nonlinear methods. As a result, they obtained an average accuracy of 86% of the five-layer (5-10-80-1) neural network.

Recently, Wissam et al. [15] published a paper entitled: A Hybrid Methodology for Short Term Temperature Forecasting which proposes the use of a SARIMA model to predict temperature. Their model gave an average accuracy (91%) higher than that of the numerical model ARPEGE (68%) for the five-day forecast of January 2017.

Many methods have been used in the field of weather forecasting. However, these methods did not consider

multivariate approaches such as the use of vector autoregression (VAR) in weather forecasting. This paper proposes a time series analysis approach based on VAR as an alternative and robust solution to strengthen numerical model decisions. This model is the reference model for multivariate time series analysis. It is powerful, flexible and easy to use. In addition, it is a simple extension of the univariate model to multivariate dynamic data time series. The VAR model has proven useful for describing the dynamic behavior of weather, economic and financial time series and for forecasting [16–18]. However, VAR techniques are short-term prediction models and for long-term prediction it is necessary to add adaptive models. A VAR model relies on different variables, endogenous, external (exogenous) and Lag, to construct a climate model, which can be expressed as follows:

$$X_t = \alpha + A_1 * X_{t-1} + \dots + A_k * X_{t-k} + B * Y_t + \epsilon_t$$

. Where:  $\epsilon_t$ : is a random disturbance that denotes a zero mean, that is generally assumed to follow a normal distribution and is uncorrelated over time [19].  $X_t$ : denotes endogenous variables.  $Y_t$ : denote exogenous variables, including the term trend.  $X_{t-k}$ : denote endogenous variables lagged by  $k$  periods and  $\alpha$  a constant term.  $A_i$ : denote square matrices coefficients [19].

The construction of the proposed VAR model was based on a set of daily weather data from January 1, 2006, to December 31, 2016, and provides a forecast of all the parameters mentioned in a period from January 1, 2017, to January 5, 2017. Therefore, the results were compared with SARIMA and the existing numerical model ARPEGE 0.5.

### 3 Contribution

This study tests three methods. The first is a seasonal ARIMA, the second is the VAR, and the third is the ARPEGE model. The two time series techniques are summarized in Fig. 1. Forecasts are run for five days in January 2017, since values for the ARPEGE model predictions are available at this time. To train the models, the authors used a database of temperature in Beirut for a period of 11 years, from January 1, 2006, through December 31, 2016 [15].

There are three essential parameters to be defined in the SARIMA model.  $p$ : is the autoregressive order,  $d$ : is the order of differentiation (if non stationarity time series) and  $q$ : is the number of coefficients of the moving average model. The Augmented Dickey–Fuller (ADF) statistical test was applied to define the value of “ $d$ .” Indeed, the Augmented Dickey–Fuller (ADF) statistical test checks whether the given series is trend stationary or not. Using BRHIA meteorological data, the ADF shows that there is

stationarity in trend, with a  $p$  value lower than 0.05, which means that it is not necessary to apply the differentiation process and therefore  $d$  is equal to 0.

The parameter “ $p$ ” is defined by the value of the last lag where the curve of the partial autocorrelation function (PACF) crosses and remains in the confidence interval as insignificant. However, since the series is not always in the Quenouille interval, the value of  $p$  is therefore equal to zero and the SARIMA model is a model based on the moving average, so the coefficient  $q$  has yet to be defined. The order “ $q$ ” of our SARIMA model (0,0, $q$ ) is defined by the value of the last lag where the autocorrelation function (ACF) crosses very significantly the Bartlett interval and remains within it. Therefore, the order  $q$  is equal to 35. After defining the SARIMA model coefficients, the ACF and PACF of the model residuals were displayed to ensure that they did not fall outside the confidence interval. Then, the Ljung–Box test was applied to check if the residuals were white noise or not [20]. The  $p$  values obtained are all greater than 0.05 significant, implying that the residuals are white noise and therefore our model is valid. This method of finding  $p, d, q$  coefficients has been adopted to build models to predict the meteorological parameters TMIN, TMAX, UMIN, UMAX, DD, FF, RR, TD and PS [15].

On the other hand, the implementation of the VAR model required the following steps: (i) Data pre-processing; (ii) Testing stationarity; (iii) Selecting Lag Length; (iv) Estimating coefficient matrices; (v) Testing for residuals autocorrelation; (vi) Assessing stability of the VAR model; (vii) Evaluating Granger Causality; (viii) Predicting of the meteorological parameters.

*Data preprocessing* consists of replacing the null data in the series with the mean values of the series itself for a month in which it was found in a specific year. The data, after pre-processing, used are: minimum temperature values (TMIN) in degrees Celsius, maximum temperature values (TMAX) in degrees Celsius, minimum humidity values (UMIN) in percentage, maximum humidity values (UMAX) in percentage, wind direction values (DD) in degrees, wind speed values (FF) meter per second (m/s), precipitation values (RR) in millimeters, dew point values (TD) in degrees Celsius, station pressure values (PS) in hectopascals (hpa) (Table 1).

*Testing stationarity* Augmented Dickey–Fuller (ADF) Statistical Test was very useful for testing the stationarity of each endogenous variable.

This table presents the Augmented Dickey–Fuller (ADF) statistical test that was applied to test the stationarity of each weather parameter. The result shows that the value of the significant  $P$  value (Prob. is between 0.0000 and 0.0009) of the ADF statistical test for each parameter is below the significance level of 0.05. This means that all weather parameters have a stationary trend.

**Table 1** Augmented Dickey–Fuller for testing meteorological parameters, whether they are stationary or not

Test type	Parameter	Null hypothesis	t Statistics	Prob*
Augmented Dickey–Fuller	TMIN	TMIN has a unit root	−3.3229	0.0009
Augmented Dickey–Fuller	TMAX	TMAX has a unit root	−4.1106	0.0009
Augmented Dickey–Fuller	UMIN	UMIN has a unit root	−10.9737	0.0000
Augmented Dickey–Fuller	UMAX	UMAX has a unit root	−11.9537	0.0000
Augmented Dickey–Fuller	RR	RR has a unit root	−43.3741	0.0000
Augmented Dickey–Fuller	TD	TD has a unit root	−4.1050	0.0010
Augmented Dickey–Fuller	DD	DD has a unit root	−32.2076	0.0000
Augmented Dickey–Fuller	PS	PS has a unit root	−6.0889	0.0000
Augmented Dickey–Fuller	FF	FF has a unit root	−38.2602	0.0000

The P value is significant (relevant) if the value is under 0.05. In statistics, it means that we reject the hypothesis H0, which assert that our series are not stationary, and accept the hypothesis H1, concluding that our series are stationary

**Table 2** This table presents the coefficients of the model of TMIN with other independent endogenous meteorological parameters with a lag=3

Coef TMIN	TMIN	TMAX	UMIN	UMAX	DD	FF	RR	TD	PS	PM
Lag=1	0.311304	0.332589	0.011209	−0.04143	−0.0004	−0.04949	−0.00999	0.152508	0.085409	−0.10927
Lag=2	0.017496	−0.13078	−0.01989	−0.02911	0.000335	−0.01896	0.017984	0.158909	−0.35336	0.331046
Lag=3	0.072752	−0.08549	−0.01329	−0.02001	5.93E-05	−0.00603	0.005738	0.127376	−0.14462	0.148169

**Table 3** This table presents the coefficients of the model of TMAX with other independent endogenous meteorological parameters with a lag=3

Coef TMAX	TMIN	TMAX	UMIN	UMAX	DD	FF	RR	TD	PS	PM
Lag=1	0.03992	0.902183	0.075732	0.02428	−0.00279	−0.19002	−0.01896	−0.31405	0.473568	−0.41235
Lag=2	0.055578	0.066165	0.009347	−0.01039	0.001728	0.032206	0.008168	0.002354	−0.22291	0.200376
Lag=3	−0.0425	−0.0254	−0.01232	−0.02593	0.001121	0.004958	−0.00747	0.208183	−0.90588	0.860628

**Table 4** This table presents the coefficients of the model of RR with other independent endogenous meteorological parameters with a lag=3

Coef RR	TMIN	TMAX	UMIN	UMAX	DD	FF	RR	TD	PS	PM
Lag=1	−0.30532	−0.45881	−0.06176	−0.02136	−0.00137	0.305904	0.229748	0.3706	0.26369	−0.67403
Lag=2	0.500652	0.250315	0.068982	0.047457	0.000283	−0.0507	−0.0194	−0.58884	0.05796	0.230682
Lag=3	0.087609	−0.02277	9.55E-05	0.001079	0.000744	−0.06492	0.066466	−0.06567	0.566093	−0.54405

*Selecting lag length* The selection of the Lag of the model was made by taking one of the lowest values between the orders of criteria calculated using the statsmodel library in python 3.7. The different orders calculated correspond to the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the final prediction error (FPE) and the Hannan–Quinn information criterion (HQIC).

*Estimating coefficient matrices* This step provides equations for certain meteorological measurements relative to other endogenous variables.

Tables 2, 3, 4, 5, and 6, respectively, represent the equations of the parameters TMIN, TMAX, UMIN, UMAX and RR taking into account current values of all meteorological variables (TMIN, TMAX, UMIN, UMAX, DD, FF, RR, TD, PS and PM) to generate the forecast values of these parameters for the first five days of January 2017.

*Testing for residuals autocorrelation* The Breusch–Godfrey serial correlation LM test is necessary to test if correlation remains between the residuals of the model. According to the result of the Breusch statistical test, p

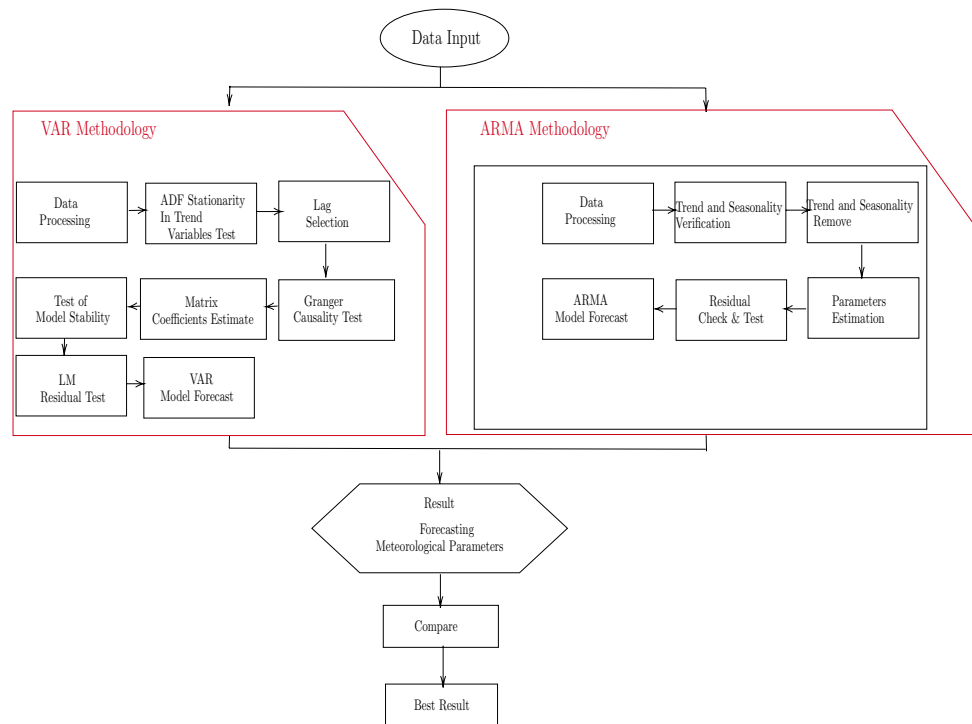
**Table 5** This table presents the coefficients of the model of UMIN with other independent endogenous meteorological parameters with a lag=3

Coef UMIN	TMIN	TMAX	UMIN	UMAX	DD	FF	RR	TD	PS	PM
Lag=1	-0.27186	-1.88216	-0.09657	-0.06899	0.011132	0.301477	0.051618	2.930018	-1.36184	0.779974
Lag=2	-0.7766	-0.02281	-0.03854	-0.08238	-0.00617	-0.17407	-0.03833	0.222342	-0.97323	1.085997
Lag=3	-0.05853	-0.20459	-0.03693	-0.0916	-0.00057	0.116626	8.00E-05	0.500303	6.443474	-6.43083

**Table 6** This table presents the coefficients of the model of UMAX with other independent endogenous meteorological parameters with a lag=3

Coef UMAX	TMIN	TMAX	UMIN	UMAX	DD	FF	RR	TD	PS	PM
Lag=1	-0.07308	0.429983	0.075682	0.444886	-0.00274	-0.08562	0.118835	0.151024	-1.09041	0.395694
Lag=2	-0.25001	0.102837	-0.00044	0.030948	9.98E-05	0.088648	-0.05821	-0.21248	0.163234	0.283925
Lag=3	-0.52316	0.103156	-0.012	0.024858	-0.00075	0.011575	-0.027	0.156103	-1.73323	1.612043

**Fig. 1** Methodologies of the two presented approaches (VAR & SARIMA)



value = 0.07 greater than 0.05 for Lag=3, which means that there is no correlation in the residual series when we select lag =3 as the order for our proposed model.

In fact, selecting the value of lag on the basis of the AIC and FPE criteria generally tends to overestimate the real order of the model with respect to BIC; on the other hand, HQIC is a compromise between AIC and BIC, it often gives a value of particular order between that provided by AIC and BIC. On the basis of this scenario, the selection of lag=3 corresponds well to the order of

the model we propose. Later, the result of the model stability shows that the model is stable on lag =3.

*Assessing stability of the VAR model* The CUSUM test was applied to test the stability of the generated model in time. As shown in Fig. 2, the CUSUM test shows that the model is stable.

*Evaluating Granger causality* Granger’s test was very useful to see what are the relevant features that jointly cause each endogenous variable. Unidirectional

relationships are indicated by (→), but bidirectional relationships between parameters can also be found.

As shown in Table 7, concerning the wind parameters, the DD parameter is strongly influenced by the TMAX and UMAX parameters (*p* values are less than 0.05). On the other hand, the FF parameter is conditioned by the parameters TD, TMIN, UMAX and UMIN. For temperature parameters, TMAX is influenced by TD, TMIN, UMAX and UMIN. On the other hand, for the TMIN parameter, the parameters with strong causality are : FF, TD and UMAX. TD is influenced by FF, TMAX, UMAX and UMIN. For humidity, UMAX is influenced by FF, RR and UMIN. For UMIN, strong causality is found by TD, TMIN, TMAX and UMAX. For station pressure, PS is caused by parameters DD, FF, RR, TD, TMIN, UMAX and UMIN. Finally, RR precipitation is strongly influenced by the following parameters : FF, TD, TMAX, TMIN, UMAX and UMIN.

*Predicting of the meteorological parameters* After validation of the model estimated on the basis of the training data used from 01/01/2006 to 31/12/2016. The forecasting process was applied for all meteorological parameters for the first five days of January 2017 (from 01/01/2017 to 05/01/2017).

## 4 Results and discussion

This section presents the empirical findings. We compare the ARIMA, VAR and ARPEGE model forecasts to the actual data. The results are summarized in Tables 8, 9 and 10.

Concerning the results of ARPEGE, the values were extracted from the weather bulletins prepared by the forecasters of the BRHIA meteorological service. Indeed, ARPEGE forecasts and generates maps and graphs related to weather parameters for 4 days after the current date. Thereafter, the forecasters of the meteorological service, during the preparation of the daily weather bulletin, analyze the graphs and generated maps of meteorological parameters in order to estimate the forecast meteorological values.

As shown in Table 8 and compared to SARIMA and ARPEGE, the VAR model gives the best accuracy for minimum temperature prediction (TMIN) with a percentage of 97%, compared to SARIMA with an accuracy of 91% and ARPEGE with an accuracy of 68%. Furthermore, as shown in Table 9, the VAR model provided the best percentage accuracy (95%) compared to SARIMA which gave an accuracy of 90% and ARPEGE(0.5) which provided an accuracy of 88% with respect to the actual values of maximum humidity (UMAX). Moreover, as presented in Table 10, for precipitation forecast, the VAR model offers the best accuracy with a percentage of 98% against a percentage

**Table 7** Causality table between the different meteorological parameters

Dependant variable	<i>P</i> _value Granger causality Test // Independant variable
DD	0.0018 → PS
	0.0017 → PM
	0.0095 → TMAX
	0.0000 ←
	0.0000 ← UMAX
FF	0.0000 → RR
	0.0000 → TD
	0.0001 ←
	0.0000 → TMIN
	0.0001 ←
TMAX	0.0000 → UMAX
	0.0000 ←
	0.0002 ← UMIN
	0.0000 → TD
	0.0000 ←
TMIN	0.0006 → TMIN
	0.0000 ←
	0.0000 ← UMAX
	0.0000 → UMIN
	0.0000 ←
PS	0.0000 → FF
	0.0000 ←
	0.0000 → RR
	0.0000 ← TD
	0.01 → UMAX
RR	0.0000 ←
	0.0018 ← DD
	0.0000 ← FF
	0.0006 ← RR
	0.0007 ← TD
TD	0.0000 ← TMIN
	0.0000 ← UMAX
	0.0000 ← UMIN
	0.0000 → FF
	0.0000 → RR
UMAX	0.0000 → TMAX
	0.0000 ←
	0.0000 ← UMAX
	0.0000 → UMIN
	0.0001 ←
RR	0.0000 → FF
	0.0000 ← TD
	0.0000 ← TMAX
	0.0000 → TMIN
	0.0000 ←
TD	0.0000 → UMAX
	0.0000 ←
	0.0000 ← UMIN
	0.0001 → FF
	0.0000 ←
UMAX	0.0000 → RR
	0.0000 ←
	0.0000 → TMAX
	0.0000 → UMIN
	0.006 ←

**Table 7** (continued)

Dependant variable	<i>P</i> _value Granger causality Test // Independant variable
UMIN	0.0001 → <i>TD</i>
	0.0000 ←
	0.0026 → <i>TMIN</i>
	0.0000 ←
	0.0000 → <i>TMAX</i>
	0.0000 ←
	0.006 → <i>UMAX</i>
	0.0000 ←

of 96% provided by the SARIMA model and 90% provided by the numerical model ARPEGE(0.5).

As to the enumeration of existing results, Table 11 summarizes all the quantitative results of the methods cited in the Related Works section for the prediction of the various meteorological parameters, as well as the results of the three models implemented in this work. In this table, three approaches to data analysis for weather forecasting are identified: the Bayesian approach in the work cited in [12], the autoregression approach in the works cited in [9, 12] and the artificial neural network approach in the works cited in [14] and [12]. As this

**Table 8** Results of accuracy issued from VAR, ARMA and ARPEGE prediction models for the first five days of January 2017 (Beirut Meteorological Station)

January_2017	Tmin real value	VAR result	ARMA result	ARPEGE result	Accuracy VAR (%)	Accuracy ARMA (%)	Accuracy ARPEGE (%)
01/01/2017	10.5	10.52	10.47	7.56	99.8	99.7	72
02/01/2017	9.7	9.82	9.46	6.32	98.7	97.5	65.1
03/01/2017	10	9.85	9.72	6.14	98.5	97.2	61.4
04/01/2017	9.7	9.3	7.21	5.8	95.7	74.3	59.7
05/01/2017	12.2	13.0	13.89	9.84	93	86.1	80.6
Mean accuracy					97.14	90.9	67.8

**Table 9** Results of accuracy issued from VAR, ARMA and ARPEGE prediction models for the first five days of January 2017 (Beirut Meteorological Station)

January_2017	Umax real value (%)	VAR result (%)	ARMA result (%)	ARPEGE result (%)	Accuracy VAR (%)	Accuracy ARMA (%)	Accuracy ARPEGE (%)
01/01/2017	86	88	90	92	97	95	93
02/01/2017	92	95	86	81	96	93	88
03/01/2017	87	82	95	78	94	91	89
04/01/2017	86	91	97	74	94	87	86
05/01/2017	67	62	58	77	93	86	85
Mean accuracy					95	90	88

**Table 10** Results of accuracy issued from VAR, ARMA and ARPEGE prediction models for the first five days of January 2017 (Beirut Meteorological Station)

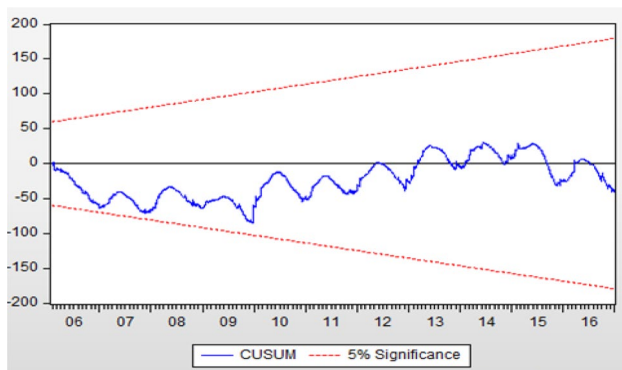
January_2017	RR real value	VAR result	ARMA result	ARPEGE result	Accuracy VAR (%)	Accuracy ARMA (%)	Accuracy ARPEGE (%)
01/01/2017	2.6	2.57	2.68	2.48	98	97	92
02/01/2017	17.0	17.34	16.6	15.3	98	97	90
03/01/2017	12.2	12.0	11.8	10.87	98	96	89
04/01/2017	1.2	1.245	1.14	1.32	96	95	90
Mean accuracy					98	96	90

work focuses on the time series approach using SARIMA autoregression techniques on the one hand and the VAR



**Table 11** This table summarizes the results of many methods that have been used to predict meteorological parameters

References	Model	Meteorological parameter	Data period	Region	Mean accuracy (%)
Rodrigues [9]	ARMA and Multiple Linear Regression	Rainfall	monthly rainfall Data From 2001 To 2013	India	82 and 87
Zabbah et al. [14]	Artificial Neural Network (ANN)	Rainfall	Daily rainfall Data From 1996 To 2016	Iran	86
Thi-Thu-Hong et al. [12]	SARIMA	Humidity	Monthly Data From 1961 To 2015	Vietnam	92
	FFNN				91
	DTWBI				91
	SES				87
	Seasonal-naive (Snaive)				86
Wissam et al. (Proposed work), 2020	VAR	TMIN and TMAX	Daily Data from 01/01/2006 To 1/12/2016	Lebanon	97 and 95
		RR			97
		UMIN and UMAX			95 and 95
		DD			94
		FF			87
		TD			91
		PS			91



**Fig. 2** CUSUM test that shows the model stability

technique which has never been used in weather forecasting on the other hand, a comparative study using the database used in this work would be very interesting for future work.

### 5 Conclusion

This paper presents a forecast model based on vector autoregression as an efficient and robust solution in the field of weather forecasting and as an alternative to numerical models to strengthen the decision of the forecasters in the meteorological services in territories with small areas such as Lebanon.

Indeed, some numerical models based on the calculation of chaotic physical equations on defined grids give biased forecast results.

The models discussed in this work have been applied to the set of meteorological parameters in the database available from the BRHIA meteorological service. The measurement of forecast accuracy showed the power of the VAR model compared to other forecast models such as SARIMA and ARPEGE in forecasting parameters such as minimum temperature (TMIN), maximum temperature (TMAX), minimum humidity (UMIN), maximum humidity (UMAX), wind direction (DD), wind speed (FF), precipitation (RR), dew point (TD), station pressure (PS).

VAR models are capable of providing more accurate forecasts than physics-based models, which are much more complex.

In perspective, in the Results and Discussion section, this paper presents the approaches used for prediction. A comparative study will be planned in order to propose a new approach that uses or combines the different techniques to improve the results obtained.

**Acknowledgements** The authors thank the two anonymous reviewers for providing helpful comments on earlier drafts of the manuscript.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Antonsen I, Bell S, Benyon R, Boese N, Vuillermoz E (2015) A new challenge for meteorological measurements: the meteomet project—metrology for meteorology. In: AIP conference proceedings
2. Malathi D, Sudarsan J, Arun C, Sheikh F, Karthicky S (2020) Analysis of data mining techniques for weather prediction. *Indian J Sci Technol*, 9
3. Reikard G, Haupt SE, Jensen T (2017) Forecasting ground-level irradiance over short horizons: time series, meteorological, and time-varying parameter models. *Renew Energy* 112:474–485
4. Kalnay Eugenia (2003) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge
5. Benard P, Vivoda J, Masek J, Smolikova P, Yessad K, Smith C, Brozkova R, Geleyn J (2009) Dynamical kernel of the Aladin-NH spectral limited-area model: revised formulation and sensitivity experiments. *Q J R Meteorol Soc*
6. Brousseau Pierre, Seity Yann, Ricard Didier, Leger Julien (2016) Improvement of the forecast of convective activity from the AROME-France system. *Q J R Meteorol Soc* 142:2231–2243
7. Sims C (1980) Macroeconomics and reality. *Econometrica* 48:1–49
8. Doan T, Litterman R, Sims C (1984) Forecasting and conditional projection using realistic prior distributions. *Econom Rev* 3:1–100
9. Rodrigues Jesleena, Deshpande Arti (2017) Prediction of rainfall for all the states of India using auto-regressive integrated moving average model and multiple linear regression. In: International conference on computing, communication, control and automation ICCUBE, IEEE
10. Chang D-F, Lin YH, Nyeu F-Y (2019) Detecting the issues of population aging by using ARIMA model. *ICIC Express Lett* 10:39–45
11. Kaul S (2017) Review of existing data mining techniques used for weather prediction. *Int J Res Appl Sci Eng Technol*, ISSN 5
12. Phan TTH (2018) Émilie POISSON Caillaud, André Bigand, comparative study on univariate forecasting methods for meteorological time series. In: 26th European signal processing conference EUSIPCO, IEEE
13. Kunjumon C, Nair S, Preetha DR (2018) Survey on weather forecasting using data mining. In: International conference on emerging devices and smart systems ICEDSS, IEEE
14. Zabbah I, Roshani A, Khafage R (2019) Prediction of monthly rainfall using artificial neural network mixture approach. *J Earth Space Phys* 44:115–126
15. Abdallah W, Abdallah N, Marion J-M, Oueidat M, Chauvet P (2020) A hybrid methodology for short term temperature forecasting. *Int J Intell Sci* 10:65–81
16. Chavan G, Momin B (2017) An integrated approach for weather forecasting over internet of things. In: International conference on I-SMAC, IEEE
17. Du Y, Hu P, Lin H, Xiang K, Zheng H, Deng J, Li M, Yang X (2017) Cointegration analysis between electricity consumption and economic growth in Fujian Province. In: Conference on energy internet and energy system integration EI2, IEEE
18. Ma X, Sun S, Liu XC, Ding C, Chen Z, Wang Y (2018) A time-varying parameters vector auto-regression model to disentangle the time varying effects between drivers responses and tolling on high occupancy toll facilities. *Transp Res Part C* 88:208–226
19. McKinney W, Perktold J, Seabold S (2011) Time series analysis in python with statsmodels. In: Proceedings of the 10th Python in science conference SCIPY
20. Lin W-S, Lin C-Y, Wu C-H (2018) Using time series analysis to predict affecting factors of thallium myocardial perfusion scan service usage frequency. In: International conference on applied system innovation, IEEE

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