A Mixed Predictive Approach To Weather Forecasting Applied To Small Areas

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

Abstract: The work presented in this paper is a mixed solution between numerical and statistical models through a hybrid predictive meteorology model based on a mix of vector autoregression and hierarchical cluster analysis (which we called HP-VAR). The presented model is constituted of three steps: Hierarchical Cluster Analysis (HCA) to study the behavior of stations and remove atypical observations based on layer altitudes, Physical equations to generate virtual stations and Time Series Analysis (TSA), in particular the Vector Auto Regressive (VAR) model, to make predictions. The work proposed in this article shows the efficiency of the model for the prediction of two meteorological parameters: TMIN and TMAX.

Index Terms: Forecasting Model; Hierarchical Cluster Analysis (HCA); Physical equations; Time Series Analysis (TSA); Vector Auto Regressive model (VAR); World Meteorological Organization (WMO); Numerical Model (NM) .

1. INTRODUCTION

Lebanon is a country of a very limited area (10452km^2), the Lebanese Civil Aviation Weather Service (Met service) at Beirut Rafik Hariri International Airport (BRHIA) uses the forecasting tool ARPEGE version 0.5, implemented by "Meteo France". This tool uses primitive equations of motion on the horizontal grid (5 km over France) and vertical layers (105 vertical levels). However, in Lebanon, these grid characteristics make the forecast results issued by ARPEGE erroneous. This problem led to the forecasters of the weather service to test two other well-known Numerical Model (NM): the Global Forecasting System (GFS) and the European Centre for Medium-Range Weather Forecasts (ECMWF). However, like ARPEGE, the GFS uses primitive equations of motion over a horizontal spacing of 13 km and 64 vertical levels [1] while the ECMWF uses primitive equations of motion on a horizontal spacing of 9 km and vertical layers of 137 vertical levels [1], [2]; In fact, ARPEGE, GFS and ECMWF are well known for their forecasting accuracy power and are used by several countries in their meteorological departments. However, the topography of Lebanon makes the forecasting mission complicated to these numerical models. The statistical models like SARIMA [3], ARIMA [4], and VAR [5] were extensively used in this field to offer an alternative to numerical models. The forecasting models based on the TSA approach show a limitation compared to the NM while forecasting a certain meteorological phenomena that occurs on a large scale for several kilometers on the horizon such as a cold front, warm front and tornado. On the other hand NM are well known by their complexities of the methodology during the constructions compare to TSA based forecasting models. The originality of the proposed HP-VAR has covered the limitation of the TSA based models and has reduced the complexities of the NMs; It has been done by combining three approaches:

- *W. Abdallah; E-mail: wissam.abdallah@etud.univ-angers.fr*
- *N. Abdallah; E-mail: nassib.abdallah@uco.fr*
- *J.M. Marion; E-mail: marion@uco.fr*
- *M. Oueidat; E-mail: mohoueidat@yahoo.com*
- *P. Chauvet; E-mail: Pierre.chauvet@uco.fr*

HCA and physical equations that have been considered as basic steps to create an efficient spatio-temporal data series. Furthermore, the statistical analysis (TSA) have been applied to analyze these spatio-temporal data series has offered very promising prospects. The first section deals with existing work related to this line of research. The second section details the methodology adopted to build the HP-VAR forecasting model. The third section discusses the results of the experiment. Finally, the last section concludes the work and gives perspectives.

2. RELATED WORK

Researchers worked on many approaches regarding the construction of a Weather Forecasting Models (WFM). Indeed, World Meteorological Organization (WMO) always claims for integrating the global efforts needed to enhance the accuracy in weather forecasts [6]. Weather prediction models are divided into three categories: numerical prediction models, time series models and models based on machine learning techniques. The first model adopted for weather forecasting was the numerical prediction model (NPM). NPM covers the entire globe with a mesh dimension, chaotic equations, traditional approximation, and system coordination that differ from one model to another. It makes it possible to predict large-scale phenomena such as: depressions, anticyclones, convection, etc. Moreover, Manfred et al. [7], in 2014 have mentioned that ECMWF model was implemented in 1973, the following Numerical Models (NM) models were ARPEGE implemented in 1998 and then a Consortium for Small Scale Modelling (COSMO) in 1999. Actually, authors have highlighted the evolution of the power of these kinds of NM in terms of increasing the accuracy of their outcome and reducing the execution time. On the other hand, authors have presented a set of chaotically equations in addition to some traditional approximations that have been taken into consideration to reduce their complexities. In the same context, Eugenia et al. [8] and DAVIES et al. [9] have confirmed the power of such models to forecast the dangerous meteorological phenomena that spread over an area of thousands of kilometers and propagated over a large scale, such as tornadoes, turbulence, convection, etc. On the other hand, authors have shown some coordinated systems which govern the construction of grid dimensions, based on the constructions of NM. Authors conclude that as well as the dimensional blocks decrease, the forecasting becomes more and more accurate. However, this lets the calculation required by the model take much longer since the interpolation between neighboring points per blocks

made by the MN on a certain grid increases. Recently, in 2017 [10], Rodrigues et al. highlight the effectiveness of the TSA approach as promising solutions that may be adopted beside the NM to enhance the accuracy of forecasting meteorological parameters, whatever, short- or long-term period predictions. Time Series Analysis (TSA) has become a very adopted approach in the field of forecasting, as it is applied in many fields such as economics, weather forecasting, environment, engineering, medicine [10]. TSA prompts to predict future weather situations by gathering historical data set at regular intervals called series [4]. These data are integrated at certain intervals (hourly, daily, weekly, monthly, and yearly, etc.). Some of TSA techniques, ARIMA and SARIMA, offer some advantages in terms of accuracy and reliability results in weather prediction [10]. The prediction models started in the 1930s, with the first univariate models Auto Regressive (AR) models, followed by Moving Average (MA) and then Auto-Regressive Integrated Moving Average model (ARIMA). In this kind of models, the predictions are based on the history data set series of a variable [11] whereas the univariate approach eliminates information that could be used from other data set. Rodrigues et al. [10], in 2017, published a study on rainfall prediction using the ARMA model and multiple linear regression. The averages of accuracy of their models are 82% and 87.3% respectively. Later in 2018, Thi-Thu-Hong et al. [3] had achieved a comparative study to predict the humidity based on seasonal autoregressive integrated moving average models (SARIMA method) and Artificial Neural Network (ANN) weather forecasting methods. SARIMA showed an average accuracy of 92% which is more efficient than that obtained with the ANN delivered that gave an average precision of 91%. Recently, Abdallah et al. [12], used SARIMA model to predict temperature where the model showed an average accuracy (91%) higher than that of the numerical model ARPEGE (68%) for the five-day forecast of January 2017. To overpass the limitations of Keynesian Macro econometric models, Sims in 1980 [5], introduces a multivariate model [13] used for forecasting procedures based on a Bayesian method for estimating Vectorial Auto Regression (VAR). Recently, in 2020, Abdallah et al. [14], used VAR for short-term weather forecasting model. The model gave an average of about 96.67% of precision between real values and three forecasted parameters: maximum humidity, temperature minimum and precipitation. Authors used VAR model to construct a weather forecasting model, it was expressed as follows:

$$X_t = \alpha + A_1 * X_{t-1} + \ldots + A_k * X_{t-k} + B * Y_t + \varepsilon_t \quad (1)$$

Where: t: is a random disturbance with a zero mean and is generally assumed to follow a normal distribution and is uncorrelated over time [15]. Yt: denotes exogenous variables, including the term trend. Xt−k: denotes endogenous variables lagged by k periods and α: a constant term. Ai: denotes square matrices coefficients [15]. Machine learning can be used to process immediate comparisons between observations and historical weather forecasts. Machine learning is useful for weather models in terms of better assessing its prediction inaccuracies [16]. Esteves et al. [17], in 2018, used neural network-based algorithm for Precipitation, in short periods between 3 and 7 days for many climate seasons. The accuracy of the model is 56%, 71%, 78% and 62% for autumn, winter, summer and spring respectively. Scher et al. [18], in 2018, worked on predicting weather forecast using convolutional neural networks (CNN). Based on the results they concluded that, trained on past historical data, the correlation between the predicted accuracy of the proposed network and the real historical data is 33% for predicting 3 days ahead. Whereas, training on model error has a lower correlation (27%) between the predicted accuracy of the suggested network and the real historical data to forecast 6 days ahead. In 2019, Zabbah et al. [19] have applied artificial neural network techniques to forecast precipitation. Their model integrated neural networks with linear and non-linear methods. They reached an average accuracy of 86% of the five-layer (5-10-80-1) neural network. In fact, HCA method is a powerful solution used to generate clusters having data set with similar features or behaviors. Angel Arroyo et al. [20] in 2017, have applied HCA to classify samples belong to different climatic zones in Spain to highlight their climate behavior in each zone. In the same field, Shobha N at al. [21], in 2017, used HCA technique to extract patterns like minimum (15 to 17°C) and maximum (28 to 29°C) air temperature, relative humidity (79 to 96%), rainfall (0 mm) and the evaporation (5.22 to 7.2 mm). The obtained results were a generation of clusters that are highly considered in the decision making for sustainable agriculture. Admir Targino et al. [22] in 2019, used HCA on hourly ozone (O3) data collected in the period 2014–2017 at 26 sites in the states of Parana, Sao Paulo and Brazil. HCA technique enabled them to reduce the data set to five homogeneous clusters with respect to seasonal, monthly, weekly and diurnal concentrations. Ibrahim Gada et al. [23] in 2020, proposed a comparative study concerning forecasting and classification models on NCDC weather data. They compared many classification methods and concluded that the classifiers such as AdaBoost, decision tree CART and XGBoost models have showed better classification accuracy when compared to other methods. Moreover, regarding the regression task, they concluded that the linear regression method performs better in terms of R^2 metric. This work combining the three approaches: HCA, TSA and physical equations to build a robust and efficient weather forecasting model.

3 METHODOLOGY AND PROPOSED STRATEGY

Figure (1) represents the architecture of the HP-VAR model to forecast meteorological parameters belong to Real and Virtual Data Matrices Stations (RDMS & VDMS).

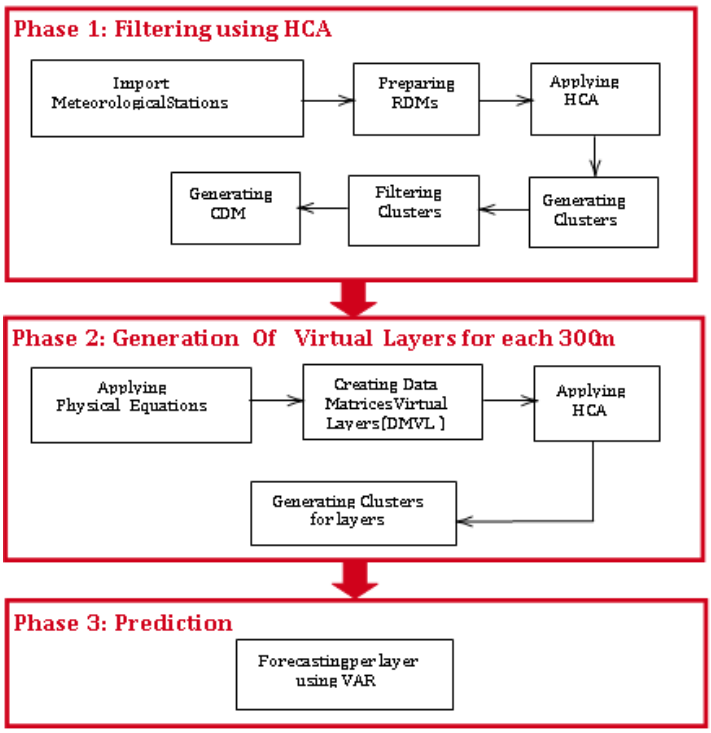

Fig. 1: *Architecture of the proposed system*

The construction of the proposed HP-VAR model was based on a set of daily weather data from January 1, 2006 till December 31, 2015 and provided forecast Tmin and Tmax meteorological parameters for four months belong to four seasons of the year 2016. The architecture shown in figure 1 incorporates the methodology of HP-VAR model. HP-VAR model has been done after achieving the following sequence of processes: (1) Preparing the Real Data Matrices Stations (RDMS) for the existing climate stations, (2) Applying HCA on RDMS and filter the generated clusters, (3) Generating Costal Data Matrices (CDM) from RDMS, (4) Using physical equations on CDMs to create Virtual Layers Data Matrices (VLDM), (5) Implementing HCA on VLDM and (6) forecasting per layer using VAR.

3.1 Preparing the RDMS for the existing climate stations

The meteorological department has 10 climate stations distributed on Lebanon's territory, each station consists of 6 meteorological parameters [12] and characterized by three parameters which are the code, name and altitude. Code: the code of the station identified by the International Civil Aviation Organization (ICAO), Name: the name of the climate station, and Altitude: the elevation of the station above the sea level in meters. These stations are: (40103, Tripoli, 5m), (40120, Sour, 5m), (40100, Beirut airport, 26m), (40132, Qubayyat, 540m), (40141, Quroan, 855m), (40101, Zahleh, 900m), (40136, Jezzine, 955m), (40135, Baysour, 978m), (40134, DeirElAhmar, 1080m), (40133, Quartaba, 1150m). There are 3 climate stations selected from the 10 climate stations that are classified as coastal stations: Beirut, Sour and Tripoli, since they are implemented at altitudes very close to sea level. The meteorological parameters collected from the 10 meteorological stations are: minimum temperature values (Tmin) in degrees Celsius, maximum temperature values (Tmax) in degrees Celsius, minimum relative humidity values (Umin) in percent, maximum relative humidity values (Umax) in percent, dew point values (TD) in degrees Celsius, atmospheric pressure values (PS) in hectopascal (hPa). The collected data to these parameters are extended form 01/January/2006 to 31/December/2015 (10 years, n=3652 observations). The data set series belong to the 10 climate stations have been concatenated on a unique matrix (RDMS).

3.2 Applying HCA on RDMS and filter the generated clusters

Parameters data set for RDMS were injected as an entry for HCA method. Therefore, the first results were poor since certain clusters delivered by the method HCA have showed a mixed of observations belongs to climate stations having a different climate features as shown in table 1.

Table 1: *Clusters generated by HCA method after being injected by RDMS without the altitude parameter*

Climate Stations	Beirut	Tripoli	Sour	Qubayyat	Zahleh	Baysour	Jezzine	Quroan	Quartaba	DeirElAhmar
Number of Total Observations per climate station	3652	3652	3652	3652	3652	3652	3652	3652	3652	3652
Number of observation per cluster C1	1220	1180	1321	1810	1097	1420	1340	1810	1642	1852
Distribution of observation per C1	31%	29%	33%	45%	27%	35%	33%	45%	41%	46 %
Number of observation per cluster C2	1498	1702	1870	1600	2140	1987	2045	1687	1514	984
Distribution of observation per C2	37%	43%	46%	40%	54%	50%	51%	42%	38%	25 %
Number of observation per cluster C3	1299	1135	826	607	780	610	632	520	861	1181
Distribution of observation per C3	32%	28%	21%	15%	19%	15%	16%	13%	21%	29 %

Table 1, shows the creation of three clusters C1, C2 and C3.

HCA offers three clusters, each cluster has its own unique percentage distribution of observations from the total observations. For example, C1 contains 31% of the observations related to Beirut station, which is classified as coastal station, and constitutes of 35% of the observation related to Baysour station, which is classified as the mountain station. Yet, the two stations belong to different climate features. This poor result was always observed even though we tried to inject some other features such as the longitudinal and latitudinal geographical reference in the data set. Finally, when we have tried to introduce the altitude parameters to each station which showed that generated clusters contain observations that belong to a particular meteorological station having the same climate features. HP-VAR model confirms that the altitude is an essential parameter that would be added on RDMS that contain the parameters data set belong to the 10 climate stations. Therefore, the result showed clusters with a big proportion of observations which belong to a meteorological stations having the same climate feature as shown in table 2.

Table 2: *Clusters generated by HCA method after being injected by RDMS on which the altitude parameter has been added*

Climate stations and altitude (meters)	Beirut 26 m	Tripoli 5 m	Sour 5 m	Qubayyat 540 m	Zahleh 900 m	Baysour 978 m	Jezzine 955 m	Quroan 855m	Quartaba 1150 m	DeirElAhmar 1080 m
Number of Total Observations per climate station	3652	3652	3652	3652	3652	3652	3652	3652	3652	3652
Number of observation per cluster C1	3540	3579	3470	-	-	-	-	-	-	-
Distribution of observation per C1	97%	98%	95%	-	-	-	-	-	-	-
Number of observation per cluster C2	112	73	182	-	-	-	-	-	-	-
Distribution of observation per C2	3%	2%	5%	-	-	-	-	-	-	-
Number of observation per cluster C3	-	-	-	3652	-	-	-	-	-	-
Distribution of observation per C3	-	-	-	100%	-	-	-	-	-	-
Number of observation per cluster C4	-	-	-	-	3579	3542	3586	3543	3505	3508

Table 2, shows the creation of five clusters C1, C2, C3, C4 and C5.

Each cluster contains a high percentage of distribution of observations from the total observations (3652) that belong to the stations having the climate features. In fact, C1 contains approximately 97% as a mean of the observations distribution among the three stations: Beirut, Tripoli and Sour. The distance among these stations is less than 300m on the altitude axis on the atmosphere, which means meteorologically, the stations have the same climate features [24], [7] (the altitude step size=300m has been validated experimentally). Concerning the cluster C2, it contains approximately 3% as a mean of the observations distribution belong to the same stations, these values have been considered as irregular values. The HCA in the model permits the detection of these values to render them into efficient values. The atypical values were replaced by the mean of the months series values to which they belong. Moreover, the same analysis has been applied on the C3, C4 and C5. After the processes of filtering, HCA was re-applied and delivered three clusters C1, C2 and C3. Coastal cluster C1 contains all the observations related to coastal stations: Beirut, Tripoli and Sour; C2 contains the observations for Qubayyat and C3 contains observations for Zahleh, Baysour, Jezzine, Quroan, Quartaba and DeirElAhmar.

3.3 Generating Costal Data Matrices (CDM) from RDMS In fact, CDM was built after applying the mean among the coastal climate stations Sour, Beirut and Tripoli matrices per observation since the observations belong to these stations are found in the same clusters. These matrices constituted of 6 meteorological parameters Tmin, Tmax, Umin, Umax, TD and PS. Finally, CDM dimension were 3652x6, where 3652 is the number of observations and 6 is the number of the parameters. When the CDM was created, the physical equations have been applied to create the 4 virtual data matrices.

3.4 Using physical equations on CDMs to create Virtual Layers Data Matrices (VLDM)

The construction of $VLDM_{300\text{-}600}$, $VLDM_{600\text{-}900}$ and $VLDM_{900\text{-}1200}$ was first based on the construction of the CDM on which the following physical equations have been applied. Certain parameters, for example, the temperature starts to show a

significant variation with the altitude as 2 ◦Celsius in each 300 meters on the altitude axis. This rule is valid just for the 1000 meters above the sea level for certain meteorological parameters [24], [7]. Moreover, the physical equations have been tested several times on the initial data matrix CDM. Upon the several experiments each of altitude step-size 100m, 200m and then 300m to create several virtual data matrices for various levels of layers, we conducted that the 300m step-size raise was the best for the atmosphere altitude axis due to the negligible change for the 100m and 200m step-size experiments. In fact, when the experiment was first started by selecting the altitude step-size value of 100m, we applied the proposed equations on CDM, and generated virtual matrices that belong to layers (L100) $VLDM_{100\text{-}200}$. The data matrix that belongs to CDM and $VLDM_{100\_200}$ was integrated and re-injected into HCA method. The experiment has showed a generation of cluster that contained all the observations that belong to CDM and of the cluster $VLDM_{100\text{-}200}$ which lead to the conclusion that the generated cluster contains the observations having the same feature without any meteorological significant changes. Similarly, the same negative results were observed in the 200m step-size altitude. However, the experiment had confirmed that at the altitude step-size of 300m succeeded in our proposed model while applying the meteorological equations on CDM.

1. Temperature TT

Equation 2 takes the temperature data set series from CDM as a first parameter with the altitude step-size between the previous layer and the current layer (600meters – 300 meters=300meters) generating a new series of temperature which belongs to $VLDM_{300\text{-}600}$. Respectively, the new temperature data set series is obtained at each step upon applying equation 2 with the modification of the altitude step-size of 300m. Eventually, the temperatures for data series CDM is appended with that belong to $VLDM_{300\text{-}600}$, $VLDM_{600\text{-}900}$ and $VLDM_{900\text{-}1200}$ to generate global temperature data set series constituted of 14608 observations (3652*4 where 3652 is the number of observations per virtual matrices and 4 is the number of virtual matrices).

$$T_c = T_{c-1} + L_c * (H_c - H_{c-1})\ [7], [8] \quad (2)$$

Where
T_c : the temperature data set for the current layer
T_{c-1} : the temperature data set for the previous layer L_c : the value of temperature laps rate
H_{c-1}: altitude of the previous layer above the sea level
H_c : altitude of the current C above the level

2. Atmospheric Pressure at climate station level PS

For and altitude step-size of 300 meters, the equation 3 takes a couple of temperature data set series from $VLDM_{300\text{-}600}$ as a current layer accompanied by the pressure data series getting from CDM of the previous layer $VLDM_{sea\text{-}300}$ as two parameters. Just as the temperature data set series calculation in the previous paragraph, we applied the similar respective steps for obtaining the new pressure data set series upon applying equation 3 for each step-size for the respective previous and current layers. Eventually, the pressure data set series for CDM is appended with that belong the layers $V\ LDM_{300\text{-}600}$, $V\ LDM_{600\text{-}900}$ and $V\ LDM_{900\text{-}1200}$ to have global pressure data set series constituted of 14608 observations (3652*4 where 3652 is the number of observations per virtual matrices and 4 is the number of virtual matrices).

$$Pc = Pc{-}1 * [Tc/(Tc + Lc * (Hc - Hc{-}1))]^{g_0 * M/R * Lc-1}[25] \ (3)$$

Where
P_c : the pressure at the current layer
P_{c-1} : the pressure at the previous layer
T_c : the temperature at the current layer
L_c : the value of temperature laps rate
H_{c-1}: altitude of the previous layer above the sea level H_c : altitude of the current layer above the level g_0 : Constant equal to the gravitational accelerator M : Constant equal to the molar mass of earth surface
R : Constant equal to the universal gas constant

3. Relative humidity RH

Equation 6 shows that the humidity data set series was done after calculating the essential dependents parameters P_w (equation 4) and P_{ws} (equation 5). Based on CDM, we will present the method of calculation of the humidity data set series of humidity that belong to the ($V\ LDM_{300\text{-}600}$) with the respective similar calculations for the each upcoming layer.

$$log_{10}(p_w) = A - (B/C + T_c)[24] \quad (4)$$

Where
P_w : partial water vapor pressure
T_c : the temperature at the current layer
A,B,C: constants

Equation 4 takes the temperature data set series as an entry to calculate partial water vapor pressure (P_w) data series relate to $V\ LDM_{300\text{-}600}$.

$$Pws = A * 10^{[m * Tc/Tc+Tn]}[25] \quad (5)$$

Where
P_{ws} : saturated water vapor pressure
A,m,T_n: constants

For the same layer $VLDM_{300\text{-}600}$, the equation 6 was established to generate saturated water vapor pressure data series P_{ws} related to $V\ LDM_{600\text{-}900}$.

$$RH = (P_w * 100)/P_{ws}\ [25] \quad (6)$$

Where
RH : relative humidity
P_w : partial water vapor pressure
P_{ws} : saturated Water vapor pressure

Upon obtaining and using the 2 series data set: partial water vapor pressure (P_w) and saturated water vapor pressure P_{ws} respectively form the equations 4 and 5. They are used to generate the relative humidity data series after applying the equation 6 belong to $VLDM_{300\text{-}600}$. Following the same steps as before, we have generated the other humidity data series belong to $V\ LDM_{600\text{-}900}$ and $V\ LDM_{900\text{-}1200}$. Finally, the 4 series related to the CDM, $VLDM_{300\text{-}600}$, $V\ LDM_{600\text{-}900}$ and $VLDM_{900\text{-}1200}$ are appended in one global series for obtaining the humidity parameter.

4. Dew Point DP

The equation 7 and 8 let us conclude that the DP and RH parameters are dependent one of each other.

$$P_w = P_{ws} * RH/100 \text{ [25]} \quad (7)$$

Where
P_w : partial water vapor pressure
P_{ws} : saturated water vapor pressure
RH : relative humidity
For the layer $V\ LDM_{300\text{-}600}$, the equation 7 was built to get the partial water vapor pressure data series P_{ws}. In fact, this was done after applying the equation 7 on the two series P_w and RH.

$$Td = T_n/[(m/log(Pw/A) - 1] \text{ [7], [25]} \quad (8)$$

Where
T_d : dew point
P_w : partial water vapor pressure
T_c : the temperature at the current layer
A,m,T_n: constants

When we have the partial Water vapor pressure data series P_{ws}, specifically for $V\ LDM_{300\text{-}600}$, we apply the equation 8 which generates the DP data series for $V\ LDM_{300\text{-}600}$. Following the same steps we have got the 2 other TD data series belong to $V\ LDM_{600\text{-}900}$ and $VLDM_{900\text{-}1200}$. Finally, the 4 data series were appended in a one global DP series data set. Upon concatenating all global data series belonging to the parameters: temperatures, relative humidity, dew point, atmospheric pressure at level stations and altitude, we constructed data matrices belong to virtual layers from 300m to 1200m $VLDM_{300\text{-}600}$, $VLDM_{600\text{-}900}$ and $V\ LDM_{900\text{-}1200}$. These matrices have been appended all together with CDM into one matrix which were injected into the HCA. All observations belong to the two matrices CDM and $V\ LDM_{100\text{-}200}$ have been classified on a same cluster C1. That means meteorologically, the observations have not shown a significant change while being raised up 100m on the altitude axis in the atmosphere [20].

3.5 Implementing HCA on VLDM:

HCA method has been injected by CDM aggregated with the three matrices $V\ LDM_{300\text{-}600}$, $V\ LDM_{600\text{-}900}$, and $V\ LDM_{900\text{-}1200}$ and has generated 4 clusters that classified the 14608 observations with a significant proportions per layer. The results obtained from HCA shows the generation of 4 clusters. The first cluster contains 3652 observations that belong to altitude step-size sea level to 300m ($CL_{sea\text{-}300}$), the second cluster contains 3652 belong to the altitude step-size 300m to 600m $CL_{300\text{-}600}$, the third cluster contains 3652 belong to the altitude step-size 600m to 900m $CL_{600\text{-}900}$ and the forth cluster contains 3652 belong to the altitude step-size 900m to 1200 m $CL_{900\text{-}1200}$. That means meteorically, each of these four cluster's observations have the same meteorological features, but their behavior differ among each other [20]. The observations belong to CDM were classified on the same cluster $CL_{sea\text{-}300}$ [20]; On the other hand the result shows the similarities of the observations in this cluster (no irregular values). This analysis could be generalized to the other three clusters. The conclusion of this result let us proceed to next part of the propose (HP-VAR) which is building a VARs model based on the three generated clusters $CL_{300\text{-}600}$, $CL_{600\text{-}900}$ and $CL_{900\text{-}1200}$. The stability of the outcome of these VARs model while forecasting the parameters Tmin and Tmax will be validated later by applying the VARs model on the RDMS to forecast the same parameters in the same period.

3.6 Forecasting per layer using VAR

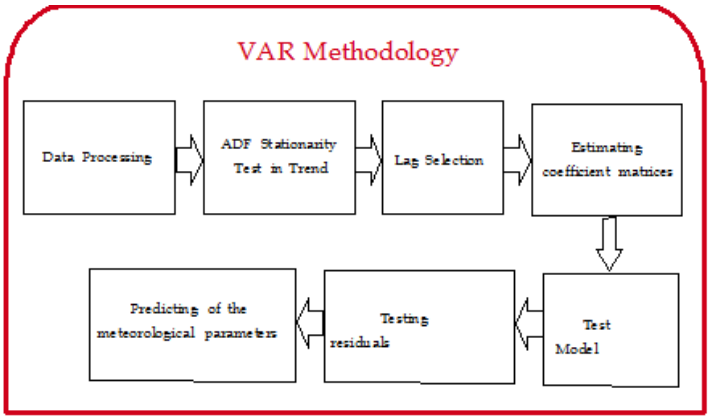

Fig. 2: *Architecture of VAR model methodology*

The final study of the proposed work focus on:
(i) Step 1: Forecast using VAR model on HP-VAR clusters: $CL_{sea\text{-}300}$, $CL_{300\text{-}600}$, $CL_{600\text{-}900}$ and $CL_{900\text{-}1200}$
(ii) Step 2: Forecast using VAR model on the 10 real climate stations.
(iii) Comparison and discussion of the results.

Figure 2 illustrates the sequence of the process that must be done to apply the VAR model clusters and VAR climate stations per layer.

3.6.1 Data pre-processing this process is applied on data series meteorological parameters that belong to the 10 climate stations. In fact, this process consists of replacing the null data related to each meteorological data set series with the mean values of the series itself for a month in which it was found in a specific year.

3.6.2 Testing stationarity Augmented Dickey Fuller(ADF) Statistical Test was applied on each endogenous variable for the clusters $CL_{sea\text{-}300}$, $CL_{300\text{-}600}$, $CL_{600\text{-}900}$ and $CL_{900\text{-}1200}$ in addition to each endogenous variable belong to the 10 climate stations as well; this test was started by assuming null hypothesis that a unit root is presented in a time series parameters. The result of this test given by VAR applied on each parameters data set is p values, these p values were always less than 0.05 which means we reject null hypothesis and confirm the stationary of the series parameters [10], [13].

Table 3: *ADF result given by VAR model on Beirut climate station*

Parameter	t-Statistics	Prob*
TMIN	-3.3229	0.0009
TMAX	-4.1106	0.0009
UMIN	-10.9737	0.0000
UMAX	-11.9537	0.0000
TD	-4.105	0.0010
PS	-6.0889	0.0000

Table 3 shows that all meteorological series is stationary since p values are less than 0.05. This result is well noticed for other VAR models applied on all other 9 stations and the 4 clusters.

3.6.3 Selecting Lag Length: the selection of the lag was based on results obtained from applying VAR models on 10 climate stations and on the 4 clusters. Such obtained results were through choosing the minimum values between criteria such as Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Final Prediction Error (FPE) and the Hannan-Quinn Information Criterion (HQIC) [10], [11]. The results show the chosen order (Lag) for the model VAR on the meteorological stations Tripoli, Sour, Beirut and $CL_{sea\text{-}300}$ is 3 whereas order of the VAR model on Qubayyat and $CL_{300\text{-}600}$ is 2; in addition, order of VAR on Quroan, Zahleh and $CL_{600\text{-}900}$ is 2; finally, order of VAR on Jezzine, Baysour, DeirAhmar, Quartaba and $CL_{900\text{-}1200}$ is 4.

3.6.4 Estimating coefficient matrices: One of interesting step of VARs model is to establish the equations for certain dependent meteorological parameters measurements relative to other endogenous variables [15]. The endogenous variables are: T_{min}, T_{max}, U_{min}, U_{max}, TD and PS. These equations are useful to generate the forecast values of T_{min} and T_{max} parameters for four months of January, April, August and November of the year 2016. Each equation belongs to the meteorological parameters represents a model defined by the lag (to where we should go back in time on the series) and by a certain matrix of coefficients that differs from other models.

3.6.5 Evaluating of stability of the VAR model The CUSUM test was a required additive step in the VAR methodology that should be applied to test the stability of the all VARs models in time [26], [27]. As shown in figure 3, the CUSUM test shows that the model is stable.

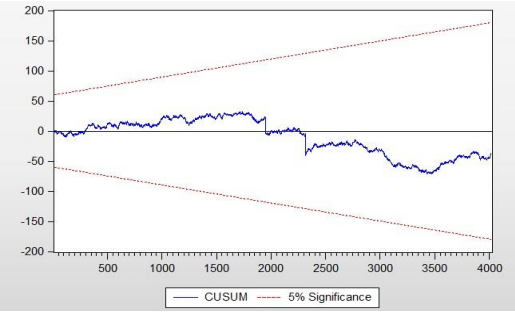

Fig. 3: *CUSUM test that shows the model stability*

The CUSUM test in the figure 3 shows an example of VAR model belongs to Tripoli meteorological parameters. Knowing that this result is similarly noticed for the other climate stations and virtual layers clusters.

3.6.6 Testing for residuals autocorrelation The Breusch-Godfrey statistical test was an essential process to prove the coherent of meteorological variables VAR models belong to all climate stations (10) and belong to the 4 clusters. Actually, Breusch-Godfrey tests started by defining null hypothesis H0 that suppose there is no serial correlation of any order up to 'p' in the residual series, where p, represents the order for each meteorological parameters model; after the verification of the p values which were all greater than 0.05, we conclude and accept H0 [10], [12] which confirms the residuals generated for VAR models are no correlated.

3.6.7 Predicting of the meteorological parameters The proposed HP-VAR model associate for T_{min} and T_{max} meteorological parameters related to climate stations and belong to layers clusters become ready to forecast 30 days for four months belong to four seasons of the year 2016. Knowing that these models are satisfied all the required steps are highlighted in the architecture shown in the figure 2.

4. Results and discussions

This section presents the results of the predicted Tmin and Tmax data series from the VAR and HP-VAR models for four months, corresponding respectively to four seasons: January for winter, April for spring, August for summer and November for autumn.

4.1 January 2016: results of forecasts TMIN and TMAX obtained from VAR and HP-VAR: the generation of the TMIN and TMAX series are offered in the table 8 (Appendix).

Table 4: *Accuracy delivered from the results of the VAR and HP-VARS models to forecast TMIN and TMAX for January 2016*

Day	TMIN Real January	TMIN VAR January Accuracy	TMIN HP-VAR January Accuracy	TMAX Real January	TMAX VAR January Accuracy	TMAX HP-VAR January Accuracy
1	11.1	98%	92%	18.2	91%	91 %
2	11.7	95%	94%	20	89%	91 %
3	13.1	83%	90%	19.85	89%	84 %
4	12.6	91%	96%	20.3	96%	83 %
5	13	87%	86%	17.25	88%	96 %
6	11.3	92%	89%	16.32	93%	90 %
7	12.5	84%	71%	18.35	84%	94 %
8	11.7	93%	91%	19.5	83%	93 %
9	11.9	91%	91%	17.1	95%	97 %
10	11.1	79%	95%	17.6	98%	88 %
11	9.7	84%	79%	13.9	78%	87 %
12	10.2	96%	78%	15.9	99%	82 %
13	10.3	86%	96%	15.5	90%	81 %
14	9.6	88%	83%	16.3	90%	83 %
15	8.9	99%	70%	14.3	76%	99 %
16	7.9	66%	73%	17.32	98%	99 %
17	8.9	59%	74%	19.324	98%	93 %
18	8.7	57%	71%	23.58	76%	65 %
19	9.5	50%	55%	24.24	71%	50 %
20	10.8	39%	39%	25.368	55%	40 %
21	12.3	32%	27%	25.541	52%	37 %
22	9.2	39%	46%	23.35	40%	27 %
23	8.4	55%	41%	18.4	39%	23 %
24	12.5	32%	26%	21.7	29%	25 %
25	13	40%	23%	18.6	50%	34 %
26	10.8	58%	25%	18.3	61%	28 %
27	10.4	49%	40%	14	43%	49 %
28	10.4	43%	35%	15.8	29%	46 %
29	9.7	30%	31%	14.9	17%	25 %
30	7.3	29%	45%	14.7	38%	65 %

The model VAR had been constructed on daily data series issued from the real coastal climate station since January 2006 till December 2015. Such results have generated forecast data series for January 2016 whereas HP-VAR had been constructed on a similar daily data series between sea level and 300 m on the altitude axis. This series has been concluded after applying the physical equations on the real data series. Therefore, such results, issued from VAR and HP-VARS models to forecast TMIN and TMAX for January 2016, were very promising especially for around the eighteen days (table 4). The reason of such high precision within such duration is due to the minor difference of Tmin and Tmax among the real data series, VAR data series and HP-VAR data series. A graphical presentation in Figure 4 offers the real data series for TMAX and that generated by VAR and HP-VAR values in a more comparable way. Experiment showed in figure 4 confirms an inversely proportional relation between time and accuracy. For instance, as long as time passes for more than seventeen days, the accuracy to forecast Tmin and Tmax decreases. As observed in the graph, it shows that both curves concerning the real data series and the HP-VAR show high similarity with negligible and minor differences during the initial duration of eighteen days.

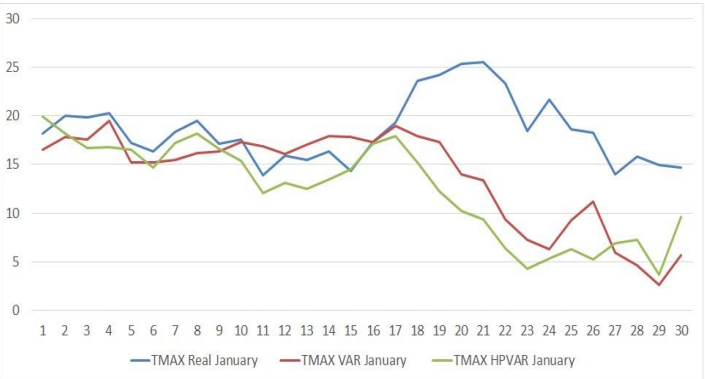

Fig. 4: *Variation of forecasting accuracy issued from VAR and HP-VAR of January 2016.*

4.2 April 2016: the outcomes of the forecasts of TMIN and TMAX developed from VAR and HP-VAR: table 9 (Appendix) shows the issued data for the real, VAR and HP-VAR data series which belong to each of TMIN and TMAX for April 2016 in this study.

Table 5: *Accuracy obtained from VAR and HP-VARS models to forecast TMIN and TMAX for April 2016*

Day	TMIN Real April	TMIN VAR April Accuracy	TMIN HP-VAR April Accuracy	TMAX Real April	TMAX VAR April Accuracy	TMAX HP-VAR April Accuracy
1	13.3	97%	98%	20.3	96%	96 %
2	12.8	91%	93%	17.4	86%	85 %
3	13.7	97%	75%	18.8	95%	73 %
4	12.1	94%	72%	20.8	95%	98 %
5	11.6	95%	82%	18	77%	83 %
6	12.6	68%	71%	19.2	91%	82 %
7	13.1	90%	82%	19.1	80%	90 %
8	11.5	70%	77%	24.9	92%	87 %
9	16.5	88%	84%	26.6	92%	91 %
10	14.7	86%	98%	20.4	99%	92 %
11	14.7	96%	97%	20.3	99%	85 %
12	13.5	99%	88%	20.3	93%	86 %
13	14.3	96%	95%	26	98%	94 %
14	16.7	77%	95%	31.8	89%	97 %
15	17	87%	79%	25.2	87%	83 %
16	12.4	85%	60%	19.4	54%	49 %
17	14	79%	81%	21	60%	51 %
18	14.8	83%	80%	20.5	48%	57 %
19	14.9	79%	79%	22.3	51%	78 %
20	15.8	82%	77%	28.6	77%	80 %
21	15.8	84%	74%	28.6	73%	80 %
22	16	81%	67%	21.2	24%	20 %
23	15.2	58%	44%	22.9	42%	59 %
24	17.9	98%	78%	21.8	48%	55 %
25	14.4	92%	43%	20	54%	44 %
26	14	71%	39%	22.2	50%	46 %
27	16.9	37%	74%	21.4	35%	26 %
28	13.7	56%	8%	21.4	43%	57 %
29	13.5	39%	40%	25.8	79%	67 %
30	19.1	21%	61%	27.24	69%	60 %

The model VAR had been constructed on daily data series issued from the real coastal climate station within a period among January 2006 and March 2016. Such results have generated foretasted data series for April 2016 whereas HP-VAR had been constructed on a similar daily data series between sea level and 300m on the atmosphere. This series has been concluded after applying the physical equations on the real data series. Therefore, such results, issued from VAR and HP-VARS models to forecast TMIN and TMAX for April 2016, were very promising especially for around the first fifteen days (table 5). This phenomenon is observed due to the minimal differences associated to the values of Tmin and Tmax resulted from the real data series, VAR series and HP-VAR data series. A graphical presentation in Figure 5 offers the real data series for TMAX and that generated by VAR and HP-VAR values in a more comparable way.

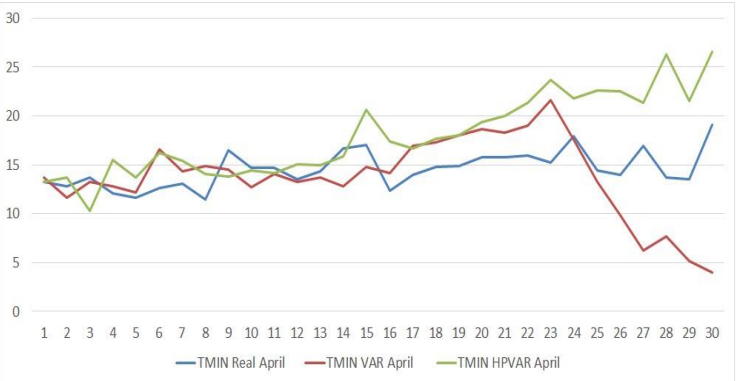

Fig. 5: *Variation of forecasting accuracy issued from VAR and HP-VAR of April 2016.*

Experiment represented in the figure 5 confirms an inversely proportional relation between time and accuracy. For instance, as long as time passes for more than fifteen days, the accuracy to forecast Tmin and Tmax decreases. As observed in the graph, it shows that both curves concerning the real data series and the HP-VAR show high similarity with negligible and minor differences during the initial duration of fifteen days.

4.3 August 2016: results of the forecasts of TMIN and TMAX obtained from VAR and HP-VAR: the generation of forecast data series are presented in table 10 (Appendix) which belong to each of TMIN and TMAX for August 2016.

Table 6: *Accuracy of forecast data series which belong to each of TMIN and TMAX for August 2016 delivered by VAR and HP-VAR*

Day	TMIN Real August	TMIN VAR August Accuracy	TMIN HP-VAR August Accuracy	TMAX Real August	TMAX VAR August Accuracy	TMAX HP-VAR August Accuracy
1	23.4	97%	92%	30	94%	91 %
2	23.5	96%	93%	30.3	96%	92 %
3	22.6	88%	90%	30.5	91%	89 %
4	22.9	95%	90%	30.6	92%	88 %
5	22.4	94%	90%	30.6	90%	91 %
6	22.4	93%	96%	30.3	94%	89 %
7	23.7	93%	88%	28.2	91%	96 %
8	25.1	93%	93%	33.24	92%	91 %
9	24.1	98%	95%	36.5	94%	95 %
10	22.7	93%	87%	38.32	93%	93 %
11	23.9	92%	86%	33.2	90%	86 %
12	24.6	93%	96%	27.98	91%	81 %
13	23.9	97%	99%	30.4	86%	82 %
14	24.3	98%	97%	34.8	90%	80 %
15	23.3	98%	94%	38.36	80%	76 %
16	24.5	92%	90%	30.7	88%	80 %
17	23.9	83%	79%	31.4	84%	79 %
18	24.7	81%	76%	31.4	78%	69 %
19	25.9	91%	79%	31.6	88%	73 %
20	26.8	93%	80%	31.8	89%	80 %
21	25.7	79%	71%	31.5	82%	82 %
22	25	79%	71%	31.8	80%	77 %
23	23.7	68%	58%	31.8	77%	70 %
24	23.3	56%	51%	31.7	86%	83 %
25	24.9	58%	64%	31.5	82%	79 %
26	26.1	76%	66%	31.3	75%	73 %
27	25.7	69%	75%	31	70%	67 %
28	24.2	53%	54%	31.1	67%	64 %
29	25.2	48%	52%	31.4	78%	64 %
30	24.8	54%	42%	31	80%	61 %

The model VAR had been constructed on daily data series issued from the real coastal climate station (Beirut) within a period among January 2006 and July 2016. Such results have generated forecast data series for August 2016 whereas HP-VAR had been constructed on a similar daily data series between sea level and 300m on the atmosphere. This series has been concluded after applying the physical equations on the real data series. Therefore, such results, issued from VAR and HP-VARS models to forecast TMIN and TMAX for August 2016, were very promising especially for around the first twenty two days (table 6). Such Accuracy within the associated duration is because of the high similarity among the real data series, data series generated by VAR and the series generated by HP-VAR regarding Tmin and Tmax. A graphical presentation in Figure 6 offers the real data series for TMAX and that generated by VAR and HP-VAR values in a more comparable way.

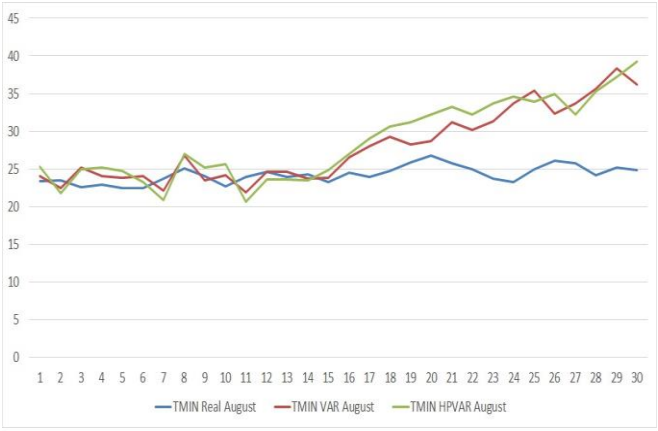

Fig. 6: *Variation of forecasting accuracy issued from VAR and HP-VAR of August 2016.*

The variation of the accuracy with respect to days in figure 6 is the experiment's outcome. It confirms an inversely proportional relation between time and accuracy. For instance, as long as time passes for more than twenty two days, the accuracy to forecast Tmin and Tmax decreases. As observed in the graph, it shows that both curves concerning the real data series and the HP-VAR show high similarity with negligible and minor differences during the initial duration of twenty two days.

4.4 November 2016: Obtained data of forecasts concerning TMIN and TMAX are observed from VAR and HP-VAR: finally, such forecast data series which belong to TMIN and TMAX for November 2016 is showed in table 11 (Appendix).

Table 7: *VAR and HP-VAR data series' precision of forecast which belong to each of TMIN and TMAX for November 2016.*

Day	TMIN Real November	TMIN VAR November	TMIN HP-VAR November	TMAX Real November	TMAX VAR November	TMAX HP-VAR November
1	17.8	19.1	18.8	21.4	22.4	23.51
2	17.5	18.4	17.8	24	26.6	27
3	18	19.4	17.1	24.9	26.4	25.4
4	19	20.4	18	24.6	26.8	25.7
5	17.25	18.3	19.4	17.5	19.25	20.
6	18.02	19.5	17.3	17.5	15.24	14.2
7	18.25	18.65	17.2	20.2	23.25	22.5
8	16.25	15.9	14.37	20.8	21.9	23.2
9	16.98	18.2	18.2	20.9	22.6	23.7
10	15.65	18.1	17.8	21.5	23.5	24.9
11	14.98	17.9	17.6	22.1	21.7	25.3
12	13.368	15.6	17.1	22.7	24.4	25.78
13	12.98	15.7	17.3	22.9	22.9	23.6
14	11.68	13.4	14.62	22.7	20.1	19.2
15	12.35	10.68	9.89	22.4	19.24	17.87
16	12.2	14.1	15.6	22.2	18.47	17.98
17	10.2	11.94	13.25	13.2	12.24	11.2
18	11.98	17.4	15.3	21.5	14.21	11.2
19	9.98	18.3	17.6	22.1	15.4	10.85
20	7.35	14.2	18.9	22.4	11.98	8.65
21	8.65	12.7	17.8	22.6	8.21	11.2
22	9.32	11.8	17.21	22.7	6.258	11.5
23	7.21	13.3	17.25	22.2	5.87	11.8
24	6.98	14.1	18.32	22.3	9.25	12.8
25	8.02	13.3	19.987	22.4	11.2	8.21
26	9.24	13	20.68	22	14.58	7.84
27	9.22	14.1	22.36	22.3	10.1	6.98
28	11.9	14.7	21.984	22.3	8.258	5.98
29	13	15.7	19.98	21.9	9.25	8.65
30	11.6	16.6	23.68	22.8	11.2	8.57

The model VAR had been constructed on daily data series issued from the real coastal climate station (Beirut) within a period among January 2006 and October 2016. Such results have generated forecast data series for November 2016 whereas HP-VAR had been constructed on a similar daily data series between sea level and 300m on the atmosphere. This series has been concluded after applying the physical equations on the real data series. Therefore, such results, issued from VAR and HP-VARS models to forecast TMIN and TMAX for November 2016, were very promising especially for around the first sixteen days (table 7). Similarly for the above seasons, such precision during this duration were computed and shown since the difference among real series, VAR data series and HP-VAR data reasons were found to be insignificant concerning Tmin and Tmax. A graphical presentation in Figure 7 offers the real data series for TMAX and that generated by VAR and HP-VAR values in a more comparable way.

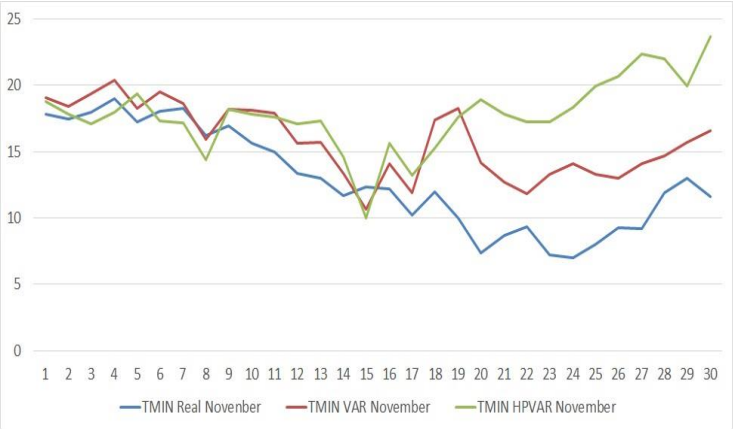

Fig. 7: *Variation of forecasting accuracy issued from VAR and HP-VAR of November 2016.*

The experiment showed in figure 7 confirms an inversely proportional relation between time and accuracy. For instance, as long as time passes for more than sixteen days, the accuracy to forecast Tmin and Tmax decreases. As observed in the graph, it shows that both curves concerning the real data series and the HP-VAR show high similarity with negligible and minor differences during the initial duration of sixteen days. The proposed HP-VAR model showed us significant and remarkable results. Just like the initial experiment was done of altitude from sea level to 300 m, it has also shown us a similar outreach through the remaining corresponding altitude levels of the atmosphere for [300,600], [600,900] and [900,1200] concerning tables, graphs, accuracy and obtained results.

4. Conclusion

The proposed mixed HP-VAR model was a robust weather forecasting solution for small areas tested at the BRHIA weather service. It was considered as a relevant solution for the historical data which has been present for around eleven years and that had been collected by several climate stations. HP-VAR was able to deliver a very promising model to forecast meteorological parameters on a thirty days period with high precision on the first 17 days. It also had a reasonable precision on the upcoming days for temperature forecasting. In addition, the HP-VAR mixed model was not only considered as a solution for historical data but also confirmed that it can support the prediction of numerical models with minimal cost and higher accuracy. This was clearly observed since HP-VAR is affordable compared to NPM in terms of building or upgrading such models. For its construction, HP-VAR was based on a mixed of meteorological physical equations, time series and HCA approaches. This model is less complicated than the numerical models, especially regarding the complexity of chaotic equations and the construction of mesh dimensions. The main advantages of the HP-VAR model are the stability and high forecasting accuracy of the meteorological parameters by taking into consideration a set of layers (4 layers: sea-level -1200m) selected based on the altitude axis of small area regions. The limitations of the proposed model were the inability to predict wind speed and direction, as no meteorological equation can evaluate their variation with altitude. A model based on deep learning can be a solution to this problem in our ongoing work.

Conflict of interest

There is no conflict of interest

References

[1] M. DeMaria, J. Rhome, R. Pasch, J. Clark, Technical Summary of the National Hurricane Center Track and Intensity Models, National Oceanic and Atmospheric Administration, 2009.

[2] ECMWF, Dynamics and Numerical Procedures, European Centre for Medium-Range Weather Forecasts (ECNWF), June 2018.

[3] T. Phan, Caillault, A. Bigand, Comparative study on univariate forecasting methods for meteorological time series, 26th European Signal Processing Conference EUSIPCO, IEEE, (September 2018) – doi:10.23919/EUSIPCO.2018.8553576.

[4] D. Chang, Y. Lin, F. Nyeu, Detecting the issues of population aging by using arima model, ICIC Express Letters, (2019 January) –doi:10.24507/icicelb.10.01.39.

[5] C. Sims, Macroeconomics and reality, Econometrica, (48) (1980) ,1–49.

[6] I. Antonsen, S. Bell, R. Benyon, N. Boese, D. delCampo, M. Dobre, J. Drnovˇsek, A. Elkatmis, E.Georgin, E. Grudniewicz, M. Heinonen, C. Rathlou, J. Johansson, P. Klason, R. Knorova, C. Melvad, J. Merrison, K. Miga la, M. Podesta, H. Saathoff, D. Smorgon, F. Sparasci, R. Strnad, A.Grzebyk, A. Merlone, G. Lopardo, E. Vuillermoz, A new challenge for meteorological measurements: The meteomet project – metrology for meteorology, AIP Conference Proceedings, (July 2015) –doi:10.1063/1.4821419.

[7] M. Wendisch, M. Salzmann, Numerical weather prediction, University of Leipzig Leipzig Institute for Meteorology (LIM) Stephanstraße 3, 04103 Leipzig, (July 2014) –.

[8] E. Kalnay, Atmospheric modeling, data assimilation and predictability, Cambridge University Press, (2003) –.

[9] P. Benard, J. Vivoda, J. Masek, P. Smolikova, K. Yessad, C. Smith, R.Brozkova, J. Geleyn, Dynamical kernel of the aladin-nh spectral limited-area model: revised formulation and sensitivity experiments,, Q. J. R. Meteorol. Soc. , (2009) –.

[10] J. Rodrigues, A. Deshpande, Prediction of rainfall for all the states of india using autoregressive integrated moving average model and multiple linear regression, International Conference on Computing, Communication, Control and Automation ICCUBEA, IEEE, (2017) –.

[11] G. Reikard, S. Haupt, T. Jensen, Forecasting ground-level irradiance over 12 short horizons, National Bureau Of Economic Research, (May 2017) – doi:10.1016/j.renene. 2017.05.019.

[12] W. Abdallah, N. Abdallah, J.-M. Marion, M. Oueidat, P. Chauvet, A hybrid methodology for short term temperature forecasting, international journal of intelligence science, International Journal of Intelligence Science 10, (January 2020) ,65–81.

[13] T. Doan, R. Christopher, A. Sims, Forecasting and conditional projection using realistic prior distributions, National Bureau Of Economic Research, (1202) (September 1983)

[14] W. Abdallah, N. Abdallah, J.-M. Marion, M. Oueidat, P. Chauvet, A vector autoregressive methodology for short-

term weather forecasting: tests for lebanon, SN Applied Sciences, springer, (August 2020) – doi:https://doi.org/10.1007/s42452-020-03292-y.

[15] W. McKinney, J. Perktold, S. Seabold, Time series analysis in python with statsmodels, Proc. Of The 10th Python In Science Conf Scipy, (2011) –.

[16] M. Zaytar, C. ElAmrani, Sequence to sequence weather forecasting with long short-term memory recurrent neural networks,, International Journal of Computer Applications, 143 (11) (June 2016) 0975–8887.

[17] J. Esteves, G. Rolim, A. S. Ferraudo, Rainfall prediction methodology with binary multilayer perceptron neural networks, Climate Dynamics, Springer-Verlag GmbH Germany, part of Springer Nature, (52) (2018) 2319—2331. doi:https://doi.org/10.1007/ s00382-018-4252-x.

[18] S. Scher, G. Messori, Predicting weather forecast uncertainty with machine learning, Quarterly Journal of the Royal Meteorological Society, (144) (2018 November) 2830 —2841. doi:10.1002/qj.3410.

[19] I. Roshani, A. R. Khafage, Prediction of monthly rainfall using artificial neural network mixture approach, Journal of the Earth and Space Physics, 44 (4) (2019 November) 115–126. doi:10.22059/JESPHYS.2018.244511.1006941.

[20] A. Arroyo, Herrero, V.Tricio, E. Corchado, Analysis of meteorological conditions in spain by means of clustering techniques,, Journal of Applied Logic, 24 (2017) 76–89.

[21] N. Shobha, T. Asha, Monitoring weather based meteorological data: Clustering approach for analysis, International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), (2017) –.

[22] A. Targino, R. Harrison, P. Krecl, P. Glantz, C. Lima, D. Beddows, Surface ozone climatology of south eastern brazil and the impact of biomass burning events, Journal of Environmental Management, (252) (2019) –.

[23] D. H. Ibrahim Gada, A comparative study of prediction and classification models on ncdc weather data, International Journal of Computers and Applications (2020) –.

[24] A. Rabah, Antoine vapor pressure correlation: Generalization and prediction of coefficients of normal alkanes, Journal of Chemical Engineering and Materials Science, 5 (4) (August 2014) 37—46. doi:10.5897/JCEMS2014.0182.

[25] O. Vaisala, Humidity conversion formulas, VAISALA: www.vaisala.com, (2013) –.

[26] G. Chavan, B. Momin, An integrated approach for weather forecasting over internet of things, International Conference on I-SMAC, IEEE, (2017) –.

[27] Y. Du, P. Hu, H. Lin, K. Xiang, H. Zheng, J. Deng, M. Li, X. Yang, Cointegration analysis between electricity consumption and economic growth in fujian province, Conference on Energy Internet and Energy System Integration EI2, IEEE, (2017) –.

7.1 Appendices

Table 8: *Result issued from VAR and HP-VARS models to forecast TMIN and TMAX for January 2016*

Day	TMIN Real January	TMIN VAR January	TMIN HP-VAR January	TMAX Real January	TMAX VAR January	TMAX HP-VAR January
1	11.1	10.87	10.2	18.2	16.5	19.9
2	11.7	11.1	12.4	20	17.8	18.2
3	13.1	10.9	11.8	19.85	17.6	16.7
4	12.6	11.5	12.1	20.3	19.5	16.8
5	13	11.3	11.2	17.25	15.2	16.5
6	11.3	10.4	12.5	16.32	15.2	14.7
7	12.5	10.5	8.9	18.35	15.5	17.2
8	11.7	10.9	10.6	19.5	16.2	18.2
9	11.9	12.98	10.8	17.1	16.3	16.6
10	11.1	8.8	10.6	17.6	17.3	15.9
11	9.7	8.1	7.65	13.9	16.9	12.1
12	10.2	10.6	7.98	15.9	16.1	13.1
13	10.3	11.7	9.84	15.5	17	12.5
14	9.6	8.4	7.98	16.3	17.9	13.5
15	8.9	8.8	6.2	14.3	17.8	14.5
16	7.9	5.21	5.8	17.32	17.3	17.1
17	8.9	5.21	6.6	19.324	19	17.9
18	8.7	4.98	6.21	23.58	17.9	15.2
19	9.5	4.714	5.23	24.24	17.3	12.2
20	10.8	4.21	4.21	25.368	14	10.2
21	12.3	3.954	3.28	25.541	13.4	9.3
22	9.2	3.547	4.21	23.35	9.35	6.3
23	8.4	4.657	3.451	18.4	7.25	4.2
24	12.5	3.987	3.2	21.7	6.25	5.3
25	13	5.21	2.98	18.6	9.24	6.2
26	10.8	6.24	2.6	18.3	11.2	5.2

27	10.4	5.124	4.2	14	5.98	6.8
28	10.4	4.458	3.6	15.8	4.587	7.2
29	9.7	2.87	2.9	14.9	2.58	3.6
30	7.3	2.14	3.3	14.7	5.65	9.6

Table 9: *Result issued from VAR and HP-VARS models to forecast TMIN and TMAX for April 2016*

Day	TMIN Real April	TMIN VAR April	TMIN HP-VAR April	TMAX Real April	TMAX VAR April	TMAX HP-VAR April
1	13.3	13.7	13.5	20.3	19.4	19.5
2	12.8	11.6	13.7	17.4	19.9	20
3	13.7	13.3	10.3	18.8	19.7	23.8
4	12.1	12.8	15.5	20.8	19.7	20.4
5	11.6	12.2	13.7	18	22.1	21
6	12.6	16.6	16.2	19.2	21	22.58
7	13.1	14.354	15.4	19.1	22.9	21
8	11.5	14.9	14.1	24.9	22.8	21.7
9	16.5	14.5	13.8	26.6	28.7	29.1
10	14.7	12.7	14.4	20.4	20.7	22
11	14.7	14.1	14.2	20.3	20.6	23.4
12	13.5	13.3	15.1	20.3	21.7	23.1
13	14.3	13.7	15	26	25.365	27.5
14	16.7	12.8	15.84	31.8	28.36	32.6
15	17	14.8	20.6	25.2	21.9	20.84
16	12.4	14.2	17.4	19.4	28.36	29.36
17	14	16.987	16.657	21	29.36	31.25
18	14.8	17.325	17.698	20.5	31.25	29.36
19	14.9	17.984	18.0258	22.3	33.25	27.25
20	15.8	18.65	19.365	28.6	35.25	34.25
21	15.8	18.254	19.987	28.6	36.25	34.25
22	16	18.98	21.32	21.2	37.21	38.21
23	15.2	21.65	23.658	22.9	36.2	32.25
24	17.9	17.54	21.8	21.8	33.21	31.54
25	14.4	13.25	22.65	20	29.21	31.25
26	14	9.87	22.54	22.2	33.25	34.21
27	16.9	6.21	21.324	21.4	35.25	37.24
28	13.7	7.65	26.325	21.4	33.54	30.58
29	13.5	5.21	21.54	25.8	31.25	34.25
30	19.1	4.02	26.54	27.24	35.68	38.25

Table 10: *Result issued from VAR and HP-VARS models to forecast TMIN and TMAX for August 2016*

Day	TMIN Real August	TMIN VAR August	TMIN HP-VAR August	TMAX Real August	TMAX VAR August	TMAX HP-VAR August
1	23.4	24.1	25.3	30	31.8	32.58
2	23.5	22.5	21.8	30.3	31.6	32.58
3	22.6	25.2	24.9	30.5	33.2	33.98
4	22.9	24.1	25.2	30.6	28.24	26.84
5	22.4	23.8	24.7	30.6	33.8	27.89
6	22.4	24.01	23.2	30.3	32.14	33.68
7	23.7	22.1	20.9	28.2	25.54	29.35
8	25.1	26.8	26.95	33.24	30.7	36.2
9	24.1	23.5	25.2	36.5	38.69	34.58
10	22.7	24.2	25.64.2	38.32	40.85	41.025
11	23.9	21.87	20.6	33.2	36.5	28.4
12	24.6	24.6	23.6	27.98	25.5	33.2

13	23.9	24.6	23.6	30.4	26.24	24.87
14	24.3	23.7	23.5	34.8	31.2	27.98
15	23.3	23.8	24.8	38.36	30.5	28.98
16	24.5	26.5	26.98	30.7	34.25	36.98
17	23.9	27.9845	28.987	31.4	36.5	38.124
18	24.7	29.324	30.65	31.4	38.24	41.25
19	25.9	28.245	31.25	31.6	35.258	40.21
20	26.8	28.65	32.21	31.8	35.214	38.21
21	25.7	31.2	33.25	31.5	37.254	37.214
22	25	30.21	32.21	31.8	38.214	39.254
23	23.7	31.258	33.65	31.8	38.987	41.24
24	23.3	33.65	34.65	31.7	36.214	36.98
25	24.9	35.45	33.98	31.5	37.21	38.21
26	26.1	32.365	34.9	31.3	39.25	39.65
27	25.7	33.68	32.21	31	40.21	41.254
28	24.2	35.65	35.25	31.1	41.24	42.25
29	25.2	38.32	37.21	31.4	38.21	42.698
30	24.8	36.25	39.22	31	37.214	42.999

Table 11: *Result issued from VAR and HP-VARS models to forecast TMIN and TMAX for November 2016*

Day	TMIN Real November	TMIN VAR November	TMIN HP-VAR November	TMAX Real November	TMAX VAR November	TMAX HP-VAR November
1	17.8	19.1	18.8	21.4	22.4	23.51
2	17.5	18.4	17.8	24	26.6	27
3	18	19.4	17.1	24.9	26.4	25.4
4	19	20.4	18	24.6	26.8	25.7
5	17.25	18.3	19.4	17.5	19.25	20.
6	18.02	19.5	17.3	17.5	15.24	14.2
7	18.25	18.65	17.2	20.2	23.25	22.5
8	16.25	15.9	14.37	20.8	21.9	23.2
9	16.98	18.2	18.2	20.9	22.6	23.7
10	15.65	18.1	17.8	21.5	23.5	24.9
11	14.98	17.9	17.6	22.1	21.7	25.3
12	13.368	15.6	17.1	22.7	24.4	25.78
13	12.98	15.7	17.3	22.9	22.9	23.6
14	11.68	13.4	14.62	22.7	20.1	19.2
15	12.35	10.68	9.89	22.4	19.24	17.87
16	12.2	14.1	15.6	22.2	18.47	17.98
17	10.2	11.94	13.25	13.2	12.24	11.2
18	11.98	17.4	15.3	21.5	14.21	11.2
19	9.98	18.3	17.6	22.1	15.4	10.85
20	7.35	14.2	18.9	22.4	11.98	8.65
21	8.65	12.7	17.8	22.6	8.21	11.2
22	9.32	11.8	17.21	22.7	6.258	11.5
23	7.21	13.3	17.25	22.2	5.87	11.8
24	6.98	14.1	18.32	22.3	9.25	12.8
25	8.02	13.3	19.987	22.4	11.2	8.21
26	9.24	13	20.68	22	14.58	7.84
27	9.22	14.1	22.36	22.3	10.1	6.98
28	11.9	14.7	21.984	22.3	8.258	5.98
29	13	15.7	19.98	21.9	9.25	8.65
30	11.6	16.6	23.68	22.8	11.2	8.57