Ecological Questions 34(2023)1 http://dx.doi.org/10.12775/EQ.2023.003

Viable forecasting monthly weather data using time series methods

Ramzan Soomro[1,*], Saghir Pervaiz Ghauri[2], Azhar Ali Marri[3], Sergij Vambol[4,*], Hoang Thi Dung[5], Nazish Manzoor[6], Shella Bano[7], Sana Shahid[8], Asadullah[9], Ahmed Farooq[9], Yurii Lutsenko[10]

[1] Quaid-e-Millat Government Degree College Liaqatabad Karachi, Pakistan
[2] Faculty of Business Administration, Commerce & Economics, Jinnah University for Women, Karachi, Pakistan
[3] Department of Statistics, University of Baluchistan, Quetta, Pakistan
[4] Life Safety Department, State Biotechnological University, Kharkiv, Ukraine
[5] Vietnam National University of Forestry, Xuan Mai, Ha Noi, Vietnam
[6] Kohat University of Science and Technology, Pakistan
[7] Department of Geology, University of Karachi, Pakistan
[8] Media and Communication Studies Department, Sindh Madressatul Islam University, Karachi 74000, Pakistan
[9] Department of Environmental Sciences, Sindh Madressatul Islam University, Karachi 74000, Pakistan
[10] Institute of Public Administration and Research in Civil Protection, Kyiv, Ukraine

*corresponding author's e-mail: Sergij Vambol (S.V.), sergvambol@gmail.com
Ramzan Soomro (R.S.), soomroramzan26@gmail.com

Received: 10 August 2022 / Accepted: 20 September 2022

Abstract. The main object of the research was to assess the forecast values of the weather parameters by using three-time series methods such as Decomposition of time series, Autoregressive (AR) model with seasonal dummies and Autoregressive moving average (ARMA) /Autoregressive Integrated moving average (ARIMA) model. A recent phenomenon in weather changing has disturbed the world in general and Pakistan in particular. In Pakistan due to climate change, flood and heat stroke have taken many lives. Stationarity was measured through the Augmented Dickey-Fuller test; results showed that some variables are I(0) and some are I(1). The reliability of the forecast results was examined through the goodness of fit test. For finding the best fit model, the performance measures of various models: Root Mean Squire Error, Mean Absolute Error and Mean Absolute Percentage Error were considered. The model in which the above statistics are the minimum was chosen as the appropriate model. After model analysis and validation, it was observed that AR-model with seasonal dummies was found to be the best fit model between the three models. Meanwhile, the forecasting for the period Jan.2018 to Dec.2018 was made based on the best fit model. Given the future forecasting results, the temperature will be normal at selected stations. The wind and rainfall will also be present. Overall, it was suggested that the obtained findings of meteorological stations' weather might be normal for the coming few months over there, and no chance of heatstroke and flood might be expected. Future studies must be carried out to provide the awareness to well-being regarding ecological hazardous to minimize their economic loss through mass media.

Key words: Time series, ADF-test, Decomposition, AR-model, Dummies, ARMA /ARIMA, ACF and PACF.

List of Abbreviations:

R_f= Precipitation,
T_{min}= Minimum Temperature,
T_{max}= Maximum Temperature
R_h= Relative humidity,
W_s= Wind speed
A_p= Atmospheric pressure
_k = Karachi,
_L = Lahore,
_P = Peshawar,
_i = Islamabad,
_Q = Quetta.

1. Introduction

Time series analysis and forecasting have become significant tools in different applications in hydrology and environmental management fields (Gholami & Sahour, 2022; Sihag et al., 2021). The increase in temperature has its impact on the rainfall and socioeconomic sectors (Aggarwal, 2003). Decrease in rainfall directly effect on drinking water, agriculture and the environment (Khan et al., 2021). Human being are facing important challenges in climate change, such as shortage of water, shortage of forests, shortage of food etc. Recently in Pakistan, thousands of people have lost their lives due to heatstroke and flood. In this research work, monthly weather data from Jan.1990 to Dec.2017 from for five stations of Pakistan: Karachi, Lahore, Peshawar, Islamabad and Quetta were modelled for parameters rainfall, maximum temperature, relative humidity and wind speed with three methods: Decomposition of time series, AR-model with seasonal dummies and Autoregressive moving average (ARMA) /Autoregressive Integrated moving average (ARIMA) model (Xu et al., 2022). The data set is divided into two portions, first starting from Jan.1990 to Dec.2016 were considered for training (analysis). The remaining 12 data points ranging from Jan. 2017 to Dec. 2017 were for checking and comparing the forecasted results. The reliability of the forecast results was examined through the goodness of fit test. For finding the best fit model, the performance measures of various models: Root Mean Squire Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were considered. The model which has minimum performance measure was chosen as the appropriate model. Stationarity was measured through the Augmented Dickey-Fuller test. ACF and PACF were also measured on the level and first difference. The future forecasting results for the period Jan.2018 to Dec.2018 were made through the best fit model and shown graphically.

Pakistan is a country blessed with beautiful landscapes such as plains, forests, hills, and plateaus (Mir, 2019). In terms of geographical location, it occupies a significant position in South Asia. This beautiful country's geographic regions encompass three distinct regions: Northern Highlands, the Indus Plains and Baluchistan Plateau. Moreover, Pakistan witnesses regional variations of climate. For example, on the Karakoram mountain ranges in the Northern part, it is snowy throughout the year. They also remain capped with snow due to their high altitude.

On the other hand, the Indus Delta is marked by extreme temperature and long summers. Karachi is the capital of Sindh province and is the largest city of Pakistan situated in the southeastern part of the country near the Arabian Sea and the Sindh province's coastal line in southern Pakistan. It has an important industrial area and port (Port Qasim) along with the Arabian Sea coast. The climate of Karachi is very different in the whole country. Karachi has a relatively mild climate, mostly arid to hyper-arid, because it is located on the coastal area. Lahore is the capital of Punjab province and is located at 31.33°N and 74.19°E. The total area of Lahore is 404 km^2. Croplands cover the area within 40 kilometre of this station. The Ravi River flows on the northern side of Lahore. The climate of Lahore performed fairly cool winters. The spring monsoon either has skipped over the area or has caused it to rain so hard that floods have resulted (Nadeem et al., 2022).

Peshawar is the Khyber Pakhtunkhwa provinces capital, located at the northwest end of Pakistan, and situated at 34.01°N and 71.5°E. The total area of Peshawar is 1257 square kilometres. The climate of Peshawar is referred to as a local steppe climate. There is not much rainfall in Peshawar; the least precipitation occurs in June, and precipitation reaches its peak in March. The temperature ranges between 20°C to 39°C. Islamabad is the capital of Pakistan and is located at 33.43°N and 73.04°E. Its elevation is 507 meters. The total area of Islamabad is 906 square kilometres. It has three artificial reservoirs: Rawal, Simply and Khanpur Dams.

Islamabad has a humid subtropical climate with hot summer and cold winter. Islamabad has mild cold climate and Sub Mountains. Half of the total annual rainfall occurs in two months averaging about 255 mm in July and August and about 50 mm per month in all other remaining months. Quetta is Baluchistan province's capital, situated on mountains with a high elevation from 1680 meters (15510 feet) above sea level and situated at 30°20′N and 76°01′E. Quetta is situated in North West of Baluchistan. The total area of Quetta is 2653 km^2. The climate of Quetta features a continental and Suri arid climate with significant variation between summer and winter temperatures, cold in winter and hot in summer. The principal mode of precipitation in winter is snow. Quetta does not have a monsoon of sustained heavy precipitation; the maximum precipitation happens 113 mm (December 2000). Temperature ranges between -5.0°C to 38°C. Quetta is a high-level altitude city and is well-known as the "Fruit Garden of Pakistan". Sultana et al. (2014) have evaluated two-time series techniques on inflation and economic growth of Pakistan; ARIMA and Decomposition techniques were applied with measuring the mean deviation and sum of squire error as performance measure checking. The results showed that ARIMA is the best model comparing to the decomposition technique. Prema and Rao (2015) used univariate techniques decomposition and ARIMA models for weather data. The model selection based on performance measures RMSE and MAPE, results showed that Decomposition is the best model comparing to other models. Streimikiene et al. (2018), have used time series methods, AR-model with seasonal dummies, ARIMA and Vector Autoregressive (VAR) models for tax revenues in Pakistan. They have used the RMSE, MAE, and MAPE to measure the efficiency of the model and ADF test for measuring the stationarity of the tax variables. Granger causality test were also applied for measuring the relationship between the variables. The conclusion showed that ARIMA method could predict more accurately than other methods. In light of other researchers, which have conducted forecasting of the weather parameters and measured performance of the models from time to time. The selected models for forecasting weather parameters by using the univariate technique are decomposition of time series, AR-model with seasonal dummies and ARMA/ARIMA model, ADF-test, ACF and PACF methods and performance measures: RMSE, MAE and MAPE. The main goal of the present research was to forecast the values of the weather parameters by using three-time series methods: Decomposition of time series, Autoregressive (AR) model with seasonal dummies and Autoregressive moving average (ARMA) /Autoregressive Integrated moving average (ARIMA) model.

2. Methodology

Data used in this research study was obtained from the Climate Data Processing Centre (CDPC) of Pakistan Metrological Department (PMD) Karachi. Monthly data for various metrological parameters were collected: precipitation (mm), maximum temperature (°C), relative humidity (knots) and wind speed for five stations Karachi, Lahore, Peshawar, Islamabad and Quetta, for a period of 28 years 1990-2017.

Decomposition of time series: A time series having the sequential set of data points or a set of vectors, $(X_t) = 0,1,2 \ldots \ldots \ldots p$, with time interval t maybe an hour, week, month and year.

Decomposition is one of the univariate techniques and identifies four components of the basic underlying pattern that characterize the metrological and environmental series. we decompose the time series into four components; the general mathematical representation of the decomposition approach is:

$$Y_t = f(S_t, T_t, C_t, I_t),$$

where Y_t = Time-series value (actual data) at period t

S_t = Seasonal component at period t T_t = Trend component at period t

C_t= Cyclic component at period t I_t = irregular component at period t

Two functional forms of time series used to relate these four components are generally applied for different series, i-e multiplicative and additive models:

$$Y_t = S_t * T_t * C_t * I_t \qquad \text{(Multiplicative rule)}$$
$$Y_t = S_t + T_t + C_t + I_t \qquad \text{(Additive rule)}$$

As the mostly used model is multiplication, was used the multiplication model to forecast short-term values of all metrological parameters.

2.1. AR-Model with seasonal dummies

Autoregressive (AR)-model with seasonal dummies process is as under:

$$Y = f(Y_{t-p})$$

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots\ldots . \text{AR (1)}$$

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t \ldots\ldots\ldots\ldots \text{AR (2)}$$

....

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_p Y_{t-p} + \varepsilon_t \quad \ldots\ldots\ldots \qquad \text{AR (P)}$$

Dummy variables, which have an essential role in explaining these models and measuring seasonal effects and seasonal variation in analysing time-series data, i-e quarterly and monthly data. Deterministic seasonality S_t can be written as the function of dummy variables. Let S be the seasonal frequency then:

- S = 4   If data is quarterly
- S = 12   If data is monthly

For monthly time series, dummy variables are as

$$S_t = \begin{cases} \gamma_1 & If\ t{=}January \\ \gamma_2 & If\ t{=}February \\ \gamma_3 & If\ t{=}March \\ \dots\dots & \dots\dots \\ \gamma_{12} & If\ t{=}December \end{cases}$$

$$= \sum_{i=1}^{s} \gamma_i D_{it}$$

2.2 Linear function of dummy variables

2.2.1 Augmented Dickey-Fuller (ADF) Test

A series is statistically stationary if its statistical properties mean, variance and covariance remain constant over time (no trend exists). Suppose we have a series Y_t which is AR (1)

$Y_t = \phi Y_{t-1} + \epsilon_t$ Where $\boldsymbol{\epsilon_t}$ is white noise, If $|\phi| < 1$ then series is stationary.

Dickey and Fuller suggested different hypotheses. They extended their test procedure by suggesting an augmented version of the test that includes extra lagged terms of the dependent variable to eliminate autocorrelation. Every process contains both AR and MA processes, but the ADF test is considered only an AR process. Test based on the regression.

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \epsilon_t$$

2.2.3 Autoregressive Moving average ARMA / Autoregressive Integrated Moving average ARIMA)

Box-Jenkins Model, ARMA (p,q) model is combination of two models Autoregressive AR(p) and Moving average MA (q) models was very effective and potent model for the time series data, can be used when data series is stationary at level or I(0). This means that the mean, variance and covariance of the series are constant over time by using lag polynomials; general ARMA (p,q) model is:

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

where Y_t is the variable which is stationary at level, p is autoregressive processes, q is moving average processes, ε_t are the error term, and φ & θ are coefficients of both AR & MA factors.

If the series Y_t is non-stationary at level, we again check at the difference and if ΔY_t (1st difference of Y_t), in this situation the model ARIMA may be used, in general ARIMA (p, d, q) model is defined as:

$$\Delta^d Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i \, \Delta^d Y_{t-i} + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j}$$

where, p = order of autoregressive, q = order of moving average and d = integrated difference.

If probability $p > 0.05$ at level, therefore series is non-stationary at level, we again check at the difference, in this situation ARIMA may be used, ARIMA (p, d, q) model using lag polynomials,

$$\varphi(L)(1-L)^d Y_t = \varphi(L)\epsilon_t \quad \text{i-e}$$

$$(1 - \textstyle\sum_{i=1}^{p} \varphi_i L^i)(1-L)^d Y_t) = (1 + \sum_{j=1}^{q} \theta_j L^j)\varepsilon_t),$$

where, p = order of autoregressive, q = order of moving average and d = integrated difference.

3. Result and Discussion

3.1 Result of descriptive statistics

The data in (Table 1) revealed that the descriptive statistics of the various parameters of five locations for the data 1990-2017. For Karachi, rainfall ranging from 0.0 to 270.40 with smaller the mean value and higher standard deviation (16.93±38.11), the maximum temperature ranging from 24.4 to 37.7 with mean value and standard deviation (32.42±3.13), of relative humidity, ranging from 24.0 to 78.0 with mean value and standard deviation (48.13±14.07), and wind speed, ranging from 1.10 to 14.10 with mean value and standard deviation (7.72±2.62).

For Lahore, rainfall ranging from 0.0 to 640.0 with smaller the mean value and higher standard deviation (56.07±85.84), maximum temperature was observed ranging from 15.20 to 41.80 with mean value and standard deviation (30.49±6.83), relative humidity, ranging from 17.0 to 76.0 with mean value and standard deviation (46.01±13.20), and wind speed, ranging from 0.10 to 5.20 with mean value and standard deviation (2.33±1.18).

For Peshawar, rainfall ranging from 0.0 to 409.0 with smaller the mean value and higher standard deviation (45.48±51.50), maximum temperature ranging from 15.5 to 42.70 with mean value and standard deviation (29.58±7.41), relative humidity, ranging from 20.0 to 69.70 with mean value (45.93±10.03) of wind speed, ranging from 0.0 to 15.30 with mean value and standard deviation (6.14±3.45).

For Islamabad, rainfall ranging from 0.0 to 743.30 with smaller the mean value and higher standard deviation (107.85±129.38), maximum temperature ranging from 14.80 to 41.10 with mean value and standard deviation

= 28.90 and sd=6.92, relative humidity, ranging from 18.0 to 71.0 with mean=44.03 and sd=11.89 and wind speed, ranging from 0.20 to 18.40 with mean= 8.90 and sd=4.21.

For Quetta, rainfall ranging from 0.0 to 167.6 with smaller the mean= 20.21and higher the sd = 32.51, maximum temperature ranging from 6.70 to 37.90 with mean= 25.31 and sd=8.89, relative humidity, ranging from 8.0 to 72.0 with mean=30.48 and sd=12.46 and wind speed, ranging from 1.8 to 21.40 with mean= 13.24 and sd=3.40. Gerretsadikan and Sharma (2011) examined the seasonal ARIMA model for the Mekele Station of Tigray region (Ethiopia); the selected parameter was rainfall for 35 years data. Box-L Jung test statistic was used for the magnitude of the auto-correction of stationarity. The accuracy of results was checked by MAE, MAPE, ME, MSE and Thieles. It was concluded that SARIMA observed better model for future forecasting.

Table 1. Descriptive Statistics for various parameters for five stations.

Variable	Minimum	Maximum	Mean	Standard deviation
Rf_K	0.000	270.400	16.928	38.111
Tmax_K	24.400	37.700	32.415	3.125
Rh_K	24.000	78.000	48.130	14.070
Ws_K	1.100	14.100	7.724	2.620
Rf_L	0.000	640.000	56.070	85.840
Tmax_L	15.200	41.800	30.490	6.830
Rh_L	17.000	76.000	46.010	13.200
Ws_L	0.100	5.200	2.330	1.180
Rf_P	0.000	409.000	45.480	51.500
Tmax_P	15.500	42.700	29.580	7.410
Rh_P	20.000	69.700	45.930	10.030
Ws_P	0.000	15.300	6.140	3.450
Rf_i	0.000	743.300	107.850	129.380
Tmax_i	14.800	41.100	28.900	6.920
Rh_i	18.000	71.000	44.030	11.890
Ws_i	0.200	18.400	8.900	4.210
Rf_Q	0.000	167.600	20.210	32.510
Tmax_Q	6.700	37.900	25.310	8.890
Rh_Q	8.000	72.000	30.480	12.460
Ws_Q	1.800	21.400	13.240	3.400

Source: Authors Calculations

3.2 Unit root identification by ADF-test

The data in (Table 2) showing the results of the stationarity for four parameters of five locations at 5% level of significance. Three parameters, rainfall, maximum temperature and relative humidity for Karachi, Lahore and Peshawar, rainfall for Islamabad and four parameters rainfall, maximum temperature, relative humidity and wind speed for Quetta have a p-value less than 0.05, reject the null hypothesis of unit root, concluded that these parameters are stationary at level, we use ARMA for these parameters, whereas wind speed for Karachi, Lahore and Peshawar and three variables maximum temperature, relative humidity and wind speed for Islamabad have p-value greater than 0.05, again these variables are checked at the first difference, now p-values become less than 0.05, these variables are concluded as stationary at the first difference, we use ARIMA for these parameters. Afrifa-Yamoah et al. (2016) have used decomposition and SARIMA models on weather data using only rainfall. ADF-test was applied for stationarity measure and built ACF and PACF with AIC criteria. Data divided into two portions analyzation and comparison of the forecast results. Findings showed that the SARIMA model could predict accurate values and helpful in long-term prediction.

Table 2. Stationarity of data series (ADF-test).

	ADF- test Statistics			
Variable	At level		At first difference	
	Value	P-Value	Value	P-Value
Rf_K	-14.387	0.000		
Tmax_K	-2.996	0.036		
Rh_K	-4.441	0.000		
Ws_K	-2.113	0.240	-13.599	0.000
Rf_L	-3.692	0.005		
Tmax_L	-3.140	0.025		
Rh_L	-3.656	0.005		
Ws_L	-1.930	0.318	-17.964	0.000
Rf_P	-15.474	0.000		
Tmax_P	-2.891	0.048		
Rh_P	-3.422	0.011		
Ws_P	-1.992	0.290	-4.604	0.000
Rf_i	-3.343	0.014		
Tmax_i	-2.401	0.142	-20.726	0.000
Rh_i	-2.668	0.081	-16.925	0.000
Ws_i	-2.104	0.243	-12.563	0.000
Rf_Q	-13.415	0.000		
Tmax_Q	-3.304	0.016		
Rh_Q	-7.909	0.000		
Ws_Q	-3.185	0.022		

Source: Authors Calculations

3.3 Chi-Square test results

The data in (Table 3) indicated the Chi-Squire test was applied to measure the comparative performance with null hypothesis that there are no differences in preferring each technique at 5% level of significance with (n-1) degree of freedom. All the parameters have minimum values that accept parameter rainfall for all stations, which does not follow the test. Sultana and Hasan (2015) evaluated the time series techniques decomposition and SARIMA model for weather data using only parameter temperature. The ADF-test was used for unit root identification to diagnose the forecasting model. Autocorrelation and partial autocorrelation functions, and serial correlation were measured. Results showed that SARIMA found to be better model for forecasting time series data.

Table 3. Showing Chi- square values, order of AR, AARMA and ARIMA model.

Parameter	ARMA/ARIMA-model		AR-model		Parameter	ARMA/ARIMA-model		AR-model	
	Order	Chi-Sq	Order	Chi-Sq		Order	Chi-Sq	Order	Chi-Sq
R_f_K	(11,12)	321.56	AR(8)	538.10	R_h_P	(12,9)	21.90	AR(10)	15.11
T_{max}_K	(15,13)	0.37	AR(11)	0.31	W_S_P	(14,1,7)	28.78	AR(11)	36.37
R_h_K	(14,14)	11.03	AR(10)	11.55	R_f_i	(13,11)	490.45	AR(9)	404.44
W_S_K	(13,1,11)	53.78	AR(9)	7.69	T_{max}_i	(10,1,11)	4.62	AR(11)	1.09
R_f_L	(7,4)	1325.12	AR(9)	474.62	R_h_i	(10,1,10)	28.83	AR(10)	11.88
T_{max}_L	(11,7)	2.85	AR(11)	1.04	W_S_i	(8,1,8)	4.70	AR(10)	2.93
R_h_L	(11,8)	17.06	AR(11)	11.09	R_f_Q	(12,9)	157.46	AR(11)	244.74
W_S_L	(10,1,8)	7.23	AR(9)	2.03	T_{max}_Q	(10,7)	0.82	AR(11)	0.45
R_f_P	(12,9)	319.06	AR(8)	212.96	R_h_Q	(11,6)	24.40	AR(10)	28.16
T_{max}_P	(14,14)	1.34	AR(11)	1.44	W_S_Q	(12,10)	2.33	AR(10)	1.29

3.4 Performance measure results

As shown in (Table 4), the comparison of the performance measures for various parameters for different locations with three methods viz: Decomposition of time series, AR-model with seasonal dummies and ARMA/ARIMA methods. The data in (Table 5) showing that AR-model with seasonal dummies is the most frequently (08 times) minimum measures model compared to other models. Mahsin et al. (2012) investigated the modelling of meteorological parameters rainfall, humidity, temperature and drought. They have used ARIMA and SARIMA models. The finding showed that the ARIMA and SARIMA models were adequate and reliable for another forecasting. Momani and Nail (2009) revealed the worked and modelled the Box-Jenkins (ARIMA) model for Amman airport station of Jordan for parameter rainfall. The ACF and PACF were also made based on unit root identification and also measured seasonal ARIMA models. The conclusion showed that individual parameter rainfall does not give such results. However, the model's performance can be improved in forecasting by taking the peek values. Furthermore, Shamsnia et al. (2011) have studied and modelled the stochastic methods on weather data with time series techniques, univariate and multivariate, correlated and non-correlated methods. The finding showed that the ARIMA model gave more accurate results. Indeed, Bari et al. (2015) studied and modelled the long-term prediction of rainfall for station Sylhet by using ARIMA model. The validity of the model may be checked through residuals. It was concluded that the ARIMA model has more applications in predicting floods, harvesting and crop management. Azad et al. (2022) revealed that the SARIMA–ANN hybrid model outperformed the remaining models seeing all concert criteria for reservoir RWL forecast, furthermore it was verifies that the SARIMA–ANN hybrid model can be a viable choice for the correct forecast of reservoir water level.

Table 4. Comparison of the performance measures by three methods.

Variable	Decomposition			AR-with seasonal dummies			ARMA / ARIMA		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Rf_K	123.83	71.86	64.43	7.91	7.90	0.04	5.32	5.31	0.01
Tmax_K	0.90	0.81	0.03	2.00	1.81	5.81	2.12	1.87	5.94
Rh_K	5.85	3.75	0.98	3.64	3.43	9.49	4.15	3.81	10.36
Ws_K	2.17	1.85	0.69	0.89	0.71	13.07	2.95	2.41	172.56
Rf_L	60.52	40.72	36.38	14.92	14.92	3711.86	20.26	14.59	7160.41
Tmax_L	1.58	1.26	0.09	0.95	0.69	2.98	2.38	1.92	7.16
Rh_L	6.34	5.08	0.92	2.83	2.03	3.57	3.65	3.60	6.87
Ws_L	0.58	0.49	0.16	0.53	0.53	83.58	0.84	0.73	89.13
Rf_P	33.16	28.59	26.67	28.81	28.77	0.00	21.54	21.48	0.00
Tmax_P	0.12	1.54	0.12	1.02	0.75	3.32	1.34	1.07	4.68
Rh_P	12.13	8.40	6.23	11.08	10.75	27.56	7.96	7.76	19.87
Ws_P	4.51	4.02	2.70	1.70	1.66	118.08	1.26	1.12	39.77
Rf_i	70.21	59.30	45.29	38.93	38.77	0.00	13.31	10.72	0.01
Tmax_i	1.52	1.37	0.09	2.21	1.58	6.73	1.85	1.41	46.72
Rh_i	5.91	4.62	0.82	5.26	5.17	13.31	2.88	2.63	271.22
Ws_i	1.84	1.72	0.40	0.64	0.63	9.39	0.66	0.64	305.91
Rf_Q	122.95	62.44	57.23	16.35	14.93	0.00	18.38	18.17	0.01
Tmax_Q	0.96	0.86	0.04	2.44	1.97	10.44	2.67	2.10	11.12
Rh_Q	8.34	7.25	2.28	12.28	11.93	60.21	9.78	9.59	48.51
Ws_Q	1.22	0.98	0.12	1.05	0.96	8.51	1.35	1.22	11.31

Source: Authors Calculations

Table 5. Frequency table of the performance measure.

Name of Model	Frequency
1) Decomposition of time series	6
2) AR-model with seasonal dummies	8
3) ARMA / ARIMA model	6

Source: Authors calculations

3.5 Sample forecasting results

From this analysis, it is observed that AR-model with seasonal dummies is the best fit model between other models. The sample results of this method revealed the comparison of original and forecast values for 2017 by Auto Regressive model for two variables, which is our best model, graph shows that the forecast values are near to the original values, thus we can say that our method is capable for future forecast, i-e 2018 (Figure1).

As shown in (Figure 2) the obtained data highlighted the forecast values for 2018 through best fit model (AR model), taken some variables for each station as sample. Variables relative humidity and wind speed for station Karachi, maximum temperature and wind speed for station Lahore, wind speed and atmospheric pressure for Islamabad, atmospheric pressure for Peshawar and wind speed for Quetta.

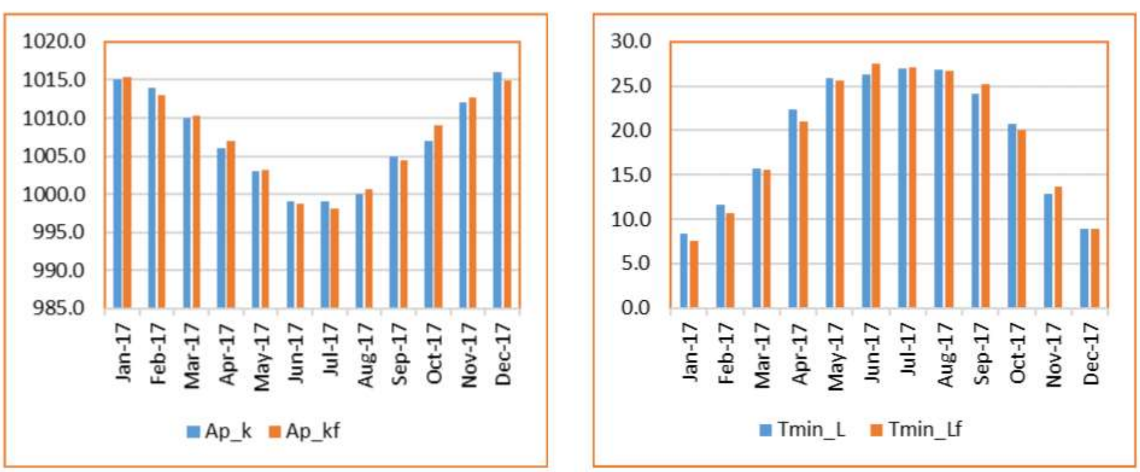
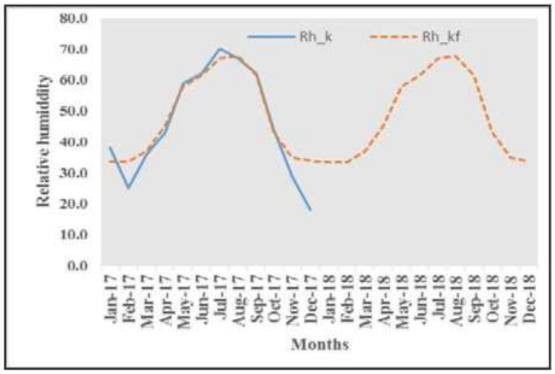
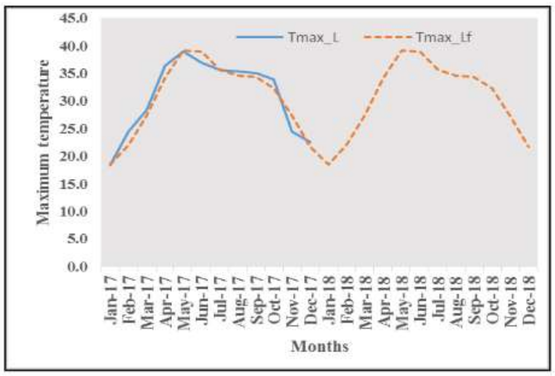


Figure 1. Sample Comparison of the forecast results for 2017 by AR Model.

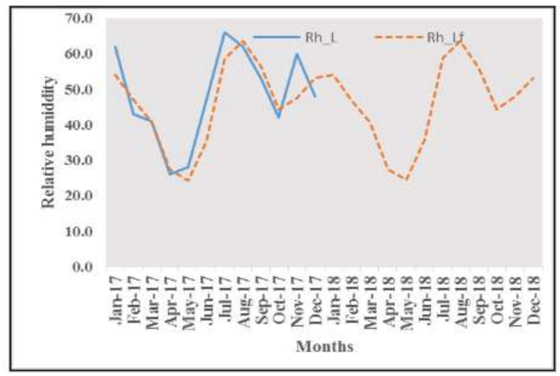


Rh_k
Rh_kf
Relative humiddity
Months

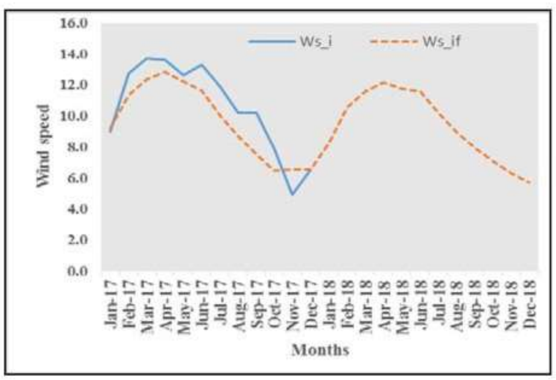
Ws_k
Ws_kf
Wind speed
Months



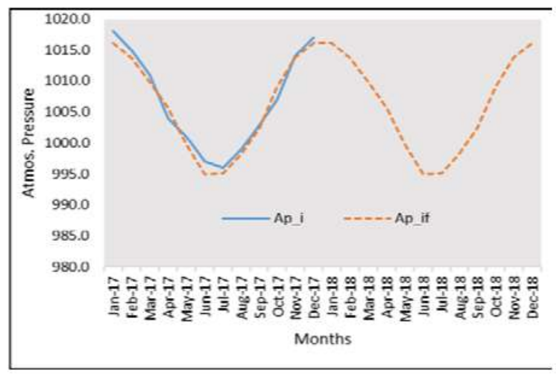
Tmax_L
Tmax_Lf
Maximum temperature
Months



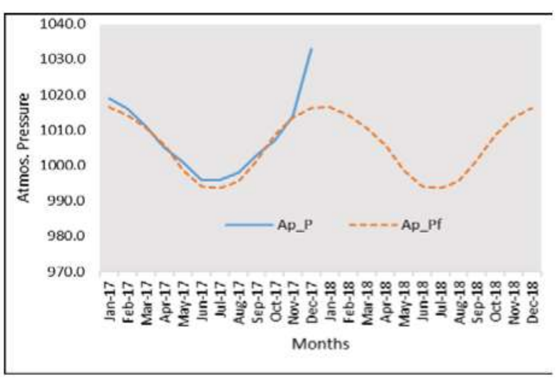
Rh_L
Rh_Lf
Relative humiddity
Months



Ws_i
Ws_if
Wind speed
Months



Atmos. Pressure
Ap_i
Ap_if
Months



Atmos. Pressure
Ap_P
Ap_Pf
Months

Ws_Q
Ws_Qf
Wind speed
Months

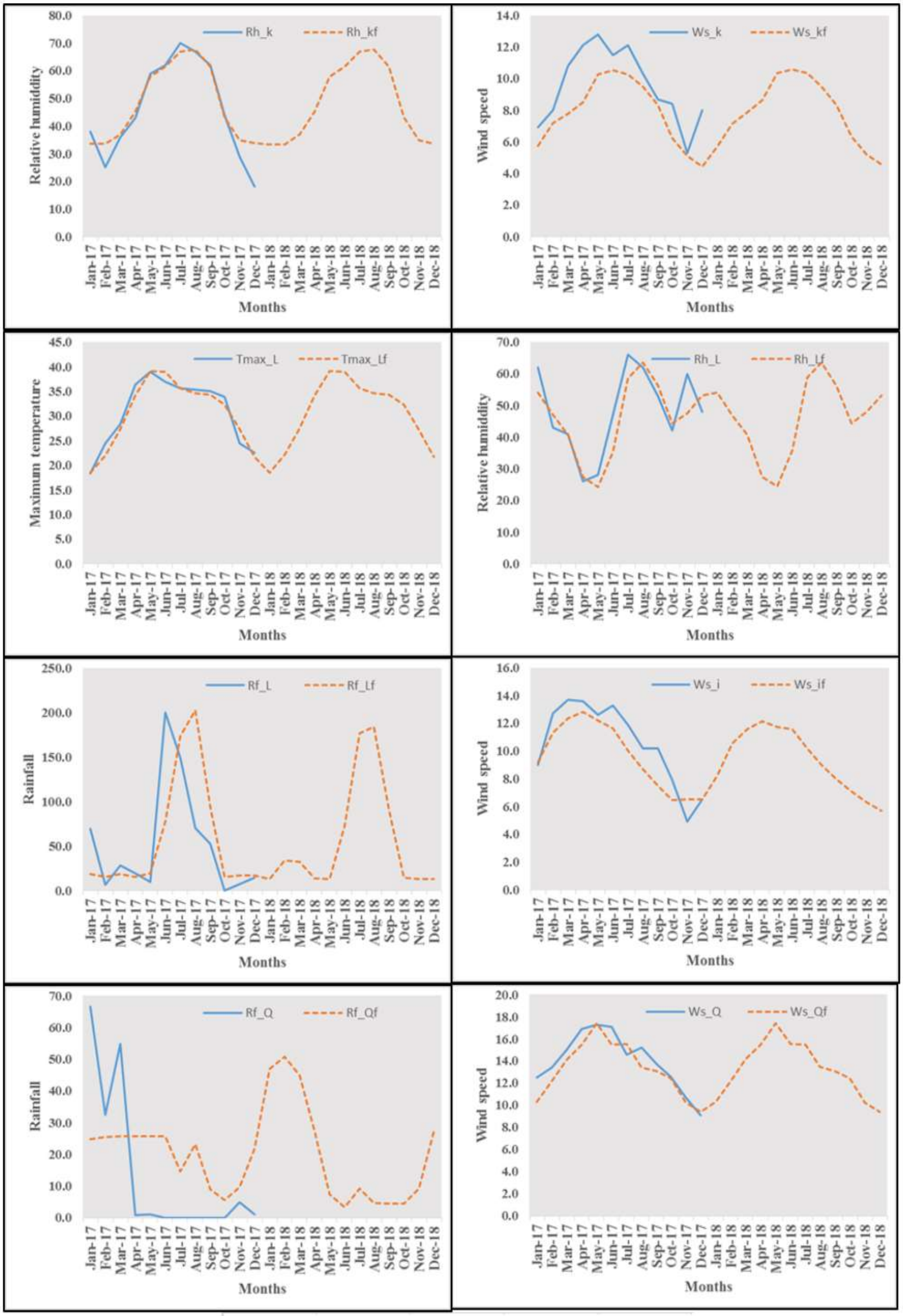

Rh_k
Rh_kf
Relative humiddity
Months
Ws_k
Ws_kf
Wind speed
Months
Tmax_L
Tmax_Lf
Maximum temperature
Months
Rh_L
Rh_Lf
Relative humiddity
Months
Rf_L
Rf_Lf
Rainfall
Months
Ws_i
Ws_if
Wind speed
Months
Rf_Q
Rf_Qf
Rainfall
Months
Ws_Q
Ws_Qf
Wind speed
Months

Figure 2. Forecast results for 2018 by AR-model for all variables for five locations.

4. Conclusions

Recently in Pakistan, due to heatstroke and flood, thousands of people have lost their lives; knowledge of fluctuations in the selected period and forecasting of them in planning is needed. This study depends on and provides reliable and accurate econometric tools that Statisticians use in forecasting the weather. This paper modeled the secondary data (1990-2017) of four parameters of Pakistan's five locations by different models. Different models were used to forecast the values of selected parameters; the final model was tested by comparing the performance measures: RMSE, MAE and MAPE. The results showed that AR-model with seasonal dummies has the minimum least performance measures thus; it was observed that AR-model with seasonal dummies is the best fit model among these models. This model can be used to forecast for studied parameters except rainfall having more significant the standard deviation and smaller the mean. This is due to the non-occurrence of the rainfall in these locations. Given the future forecasting results, it was concluded that temperature might be normal at selected stations. The wind and rainfall might also be present. Further, these stations' weather may possibly typical for the present situation and over there no chance of heatstroke and flood might be expected.

4.1 Recommendation

The media plays a significant role in the dissemination of information about weather forecasting. It is recommended to disseminate weather reports and data via mass media so that individuals and institutions can use them to lessen physical and biological losses, losses brought on by weather, and to improve social benefits such as the preservation of life and property, the promotion of public health and safety, and the advancement of economic well-being and quality of life. Social media may also be utilised as an extra means of swiftly sharing information on watches or warnings for severe weather. Social media platforms are excellent for informing the public about extreme weather.

References

Afrifa-Yamoah E., Saeed B.I. & Karim A., 2016, Sarima modelling and forecasting of monthly rainfall in the Brong Ahafo Region of Ghana. World Environment 6(1): 1–9.

Aggarwal P.K., 2003, Impact of climate change on Indian agriculture. Journal of Plant Biology-new Delhi 30(2): 189–198.

Azad A.S., Sokkalingam R., Daud H., Adhikary S.K., Khurshid H., Mazlan S.N.A. & Rabbani, M.B.A., 2022. Water Level Prediction through Hybrid SARIMA and ANN Models Based on Time Series Analysis: Red Hills Reservoir Case Study. Sustainability, 14(3): 1843.

Bari S.H., Rahman M.T., Hussain M.M. & Ray S., 2015, Forecasting monthly precipitation in Sylhet city using ARIMA model. Civil and Environmental Research 7 (1): 69–77.

Gerretsadikan A. & Sharma M.K., 2011, "Modeling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia)", Statistics and Application 9 (1&2): 31–53.

Gholami V. & Sahour H., 2022, Simulation of rainfall-runoff process using an artificial neural network (ANN) and field plots data. Theoretical and Applied Climatology 147(1): 87–98.

Khan R.A., El Morabet R., Mallick J., Azam M., Vambol V., Vambol S., & Sydorenko V., 2021, Rainfall Prediction using Artificial Neural Network in Semi-Arid mountainous region, Saudi Arabia. Ecological Questions 32(4): 127–133.

Mahsin M.D., Akhter Y. & Begum M., 2012. Modeling rainfall in Dhaka division of Bangladesh using time series analysis. Journal of mathematical modelling and application 1(5): 67–73.

Mir B.I., 2019, An Overview of Corporate Social Responsibility in the Hospitality and Tourism Sector of Pakistan. UW Journal of Management Sciences 3(2): 1–14.

Momani P.E.N.M. & Naill P.E., 2009, Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. American Journal of Environmental Sciences 5 (5): 599.

Nadeem M.U., Waheed Z. & Ghaffar A.M., 2022, Application of HEC–HMS for flood forecasting in Hazara catchment Pakistan, south Asia. International Journal of Hydrology 6(1): 7–12.

Prema V. & Rao K.U., 2015, Time series decomposition model for accurate wind speed forecast. Renewables: Wind, Water, and Solar 2 (1): 1–11.

Shamsnia S.A., Shahidi N., Liaghat A., Sarraf A. & Vahdat S.F., 2011, Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). In International conference on environment and industrial innovation 12(1): 282–285.

Sihag P., Sadikhani M. R., Vambol V., Vambol S., Prabhakar A.K. & Sharma N., 2021, Comparative study for deriving stagedischarge–sediment concentration relationships using soft computing techniques. Journal of Achievements in Materials and Manufacturing Engineering 104(2): 57–76.

Streimikiene D., Rizwan Raheem A., Vveinhardt J., Pervaiz Ghauri S. & Zahid S., 2018, Forecasting tax revenues using time series techniques–a case of Pakistan. Economic research-Ekonomska istraživanja 31(1): 722–754.

Sultana K., Rahim A., Moin N., Aman S. & Ghauri S.P., 2014, Forecasting inflation and economic growth of Pakistan by using two time series methods. International Journal of Business and Economics Research 2 (6): 174–178.

Sultana N. & Hasan M.M., 2015, Forecasting Temperature in the Coastal Area of Bay of Bengal-An Application of Box-Jenkins Seasonal ARIMA Model. Statistics 7(8): 149–159.

Xu W., Peng H., Zeng X., Zhou F., Tian X. & Peng X., 2022, A Hybrid Modeling Method Based on Linear AR and Nonlinear DBN-AR Model for Time Series Forecasting. Neural Processing Letters 54(1): 1–20.