Modelling and Forecasting Climate Time Series with State-Space Model

A. F. Adedotun, T. Latunde, O. A. Odusanya

Abstract

This study modelled and estimated climatic data using the state-space model. The study was specifically to identify the pattern of the trend movement i.e., increase or decrease in the occurrence of the climatic change; to use of Univariate Kalman Filter for the computation of the likelihood function for climatic projections; to modelling the climatic dataset using the state-space model and to assess the forecasting power of the state-space models. The data used for the work includes temperature and rainfall for periods January 1991 to December 2017. The data are tested for normality. Shapiro-Wilk, Anderson-Darling and Kolmogorov-Smirnov test of normality for the climatic data all showed that the variables are not normally distributed. The work spans the use of breaking trend regression model to fit climatic data to estimate the slopes which show much increase in climatic data has been recorded from the initial time data collection until the present. Investigations and diagnostic are carried out by checking for corrections in the residuals and also checking for periodicity in the residuals. The results of this investigation show significant autocorrelation in the residuals indicating the presence of underlying noise terms which is not accounted for. By treating the residual as an autoregressive moving average (ARMA) process whereby we can obtain its spectral density, the result from the parametric spectral estimate shows underlying periodic patterns for monthly data, thus, leads to a discussion on the need to treat climatic data as a structural time series model. We select appropriate models by considering the goodness of fit of the model by comparing the Akaike information criterion (AIC) values. Parameters are estimated and accomplished with some measures of precision.

Keywords: Rainfall, Temperature, Kalman filter, State-space models, Residuals

1. Introduction

Largely, state-space models (SSMs) have been used in several areas of applied statistics. In specifically, there are some desirable properties of the linear state-space models and also, they have vast potential in time series modelling that incorporates latent processes. When a model is found to be in the linear state-space form, the most used algorithm to predict the latent process, the state, is the Kalman filter algorithm. This algorithm is a technique for computing, at each time \( t = 1, 2, \ldots \), the optimal estimator of the state vector based on the existing information until \( t \) and its success lies on the fact that is an online estimation procedure.

This predicting structure, whose performance was in recent times compared to other numerous forecasting methods across thousands of time series [1], adapts to underlying alterations...
in series dynamics and automatically revises forecasts as new observations. In agreement with the above, the study adopts the state-space model to analyse and forecast for the temperature and rainfall data in Nigeria.

The purpose of this study is to model and estimate the climate using the state-space model and the specific objectives are to: identify the pattern of the trend movement i.e. increase or decrease in the occurrence of the climatic change; use of Univariate Kalman Filter for the computation of the likelihood function for climatic projections; modelling the climatic dataset using the state-space model and assess the forecasting power of the state-space models.

Shamshad et al. [2] show a comparison of Artificial Neural Networking Multilayer Perceptron (ANN-MLP) with the Automatic Exponential Smoothing Algorithm (ETS) and the Auto-Regressive Integrated Moving Average (ARIMA) models for forecasting Lahore, Pakistan’s main weather parameter. Models are built by taking into account average monthly maximum and minimum temperature, relative humidity, wind speed and precipitation amount. Data covering thirty years (1987-2016) was used to build the models. ANN-MLP is a mathematical method, and the computational methods are ARIMA and ETS. They divide the thirty years data (1987-2018) data into training (1987 till 2016) and test (2017 till 2018) set to ensure the efficiency and reliability of all these models along with the performance criteria of the estimates. The research explained in brief how the various methods of learning can be used to formulate ANN-MLP. Deciding the most appropriate model and network configuration is based on their performance forecast. MAE (Moving Average Error), RMSE (Root-Mean-Square error), ME (mean error), MASE (mean absolute scaled error) suggest better results for ANN-MLP. Afsar et al. [3] investigated variability in temperature and precipitation in the Gilgit-Baltistan region. They used regression and stochastic models to show temperature and rainfall predictions. We observed precipitation prolonged with temperature increasing. A decrease in the amount of precipitation is observed from 2007 to 2011 with an increase in the monthly average maximum temperature. They considered AR(1) ideally suited to temperature forecasting.

In the study of Faisal and Ghaflar [4] on the Thiessen Polygon technique to test Pakistan’s 56 stations for 50 years (1961-2010) weighted rainfall area (AWR). Month-to-month precipitation records of fifty-six stations, storm measurements for the fifty-year season (1961-2010) and a standard precipitation size relationship were used by the Theissen system. Yusof et al. [5] used an amount of rainfall to be categorised into seven categories (extremely wet to extremely dry) to analyze dry and wet events using Peninsular Malaysia data. They used precipitation index (SPI) standardizes to model the best fit distribution to reflect the rainfall. In comparison with Gamma and Weibull distributions, the lognormal distribution is found to be better matched to the daily rainfall in the area.

Extreme Pakistan temperature events and rainfall for the period 1965-2009 were examined in Zahid and Rasul work [6] to quantify Pakistan frequency. They used F-test to determine the country minimum and maximum extreme temperature events. We pointed out that all over the country, certain extreme events are increasing. Regarding extreme rainfall events, they used the K-S method at a confidence interval of 95 per cent and concluded that the southern half of Pakistan faces more wet spells due to global warming and climate change. In the same vein, rainfall data in Queensland, Australia including climate indices, monthly rainfall and temperature was surveyed in Abbott and Marohasy [7]. They brought ANN into the area to predict monthly rainfall. They suggested there is scope for improvement in this product design. Analysis of ANN to forecast lasting monthly temperature and rainfall from 76 stations in Turkey at any point for the period 1975-2006 was also carried out in Bilgili and Sahin’s work [8] based on knowledge from neighbouring stations. They divided 76 measuring stations into training sets and test sets. The fitted model was satisfactory because the errors are within reasonable limits.

2. Materials and Methods

2.1. State-Space Model

A state-space representation in control engineering is a mathematical model of a physical system as a collection of input, output and state variables similar to first-order differential equations or difference equations. State variables are variables whose values change over time in a way that depends on the values they have at any given time and often depends on the values of input variables placed externally. The values of the output variables depend on the values of the state variables.

“State-space” is the Euclidean space, where the variables on the axes are the variables of the body. The state of the system within that space can be expressed as a vector. The state-space method is characterized by significant algebraization of general system theory, which makes it possible to use Kronecker vector-matrix structures. The ability of these structures can be efficiently used to study systems with modulation or without modulation. The state-space representation (also called the “time-domain method”) gives a convenient and compact way to model and analyze systems with multiple inputs and outputs. With p and q outputs, we would otherwise have to write down $q \times p$ Laplace transforms to encode all the information about a system.

2.2. State-Space Representation

In the time domain, a system can be described in general by a set of linear differential and algebraic equations (i.e., state-
space model):
\[
\frac{dx}{dt} = Ax + Bu
\]
\[
y = Cx + Du
\]  
(1)  
(2)

Where \(x\) is the vector of the state variable \(u\) is the vector of inputs (manipulated variables) \(y\) is the vector of outputs (controlled/measured variables), and \(A, B, C,\) and \(D\) are constant matrices of appropriate dimensions.

Taking Laplace transform of Equations (1) and (2) using zero initial conditions, and re-arranging, the system transfer function matrix \(G(s)\), the relationship between the inputs \(u\) and the outputs relates the state model equations (1) and (2) as:
\[
G(s) = \frac{y(s)}{u(s)} = C(sI - A)^{-1}B + D
\]  
(3)

The representation of the (SSM) was developed based on graph theory principles. Two major types are used:

1. The signal-flow-graph (SFG) type.
2. The matrix representation.

The first type is easier to use and hence has been adopted by many researchers [9-10], and it has also been adopted in this research for the same reason.

Each SFG consists of directed branches interconnected at nodes. In the state-space model, the nodes represent the variables (signals), and the branches connecting the nodes indicate that these nodes are related. Each branch is assigned a numerical value or a function, which quantifies the relationship between the two variables in terms of a gain factor or a transfer function.

2.3. Identification and estimation of a state-space model process

Before the analysis can be processed, identification of the state-space model would be carried out in the following ways:

The measurement equation has the form
\[
Y_t = Z_t \alpha_t + d_t + \epsilon_t = 1 \ldots T
\]  
(4)

To accommodate data properties as temporal correlation and the periodic behaviour, a periodic state-space model is proposed. This model is defined as
\[
Y_{s,n} = [S (n - 1) + s] X_{s,n} + D_{s,n} + \epsilon_{s,n}
\]  
(5)

\[
X_{s,n} = \mu_s + S(X_{s-1,n} - \mu_{s-1}) + \epsilon_{s,n}
\]  
(6)

where
\(s\) is the season of the year with \(s = 1, 2, \ldots, S\);
\(n\) is the year with \(n = 1, 2, \ldots, N\);
\(Y_{s,n}\) represents the time series observation in the \(S^n\) season of the \(n^{\text{th}}\) year;
\([S (n - 1) + s]^{\text{th}}\) is the observation of the time series;
\([1, 2, 3, 4, \ldots, S]\) are the unknown parameters representing the fixed effects in the model;
\(D_{s,n}\) is a \(1 \times S\) matrix of known values, a design matrix;
the error process \((\epsilon_{s,n})\) is a white noise disturbance, which is assumed to have \(\text{var} (\epsilon_{s,n}) = \frac{\sigma^2}{c}\)
\(\mu_s\) is the mean of the process \((X_{s,n})\) for the \(s^{\text{th}}\) season, \(s\) is the autoregressive parameter for season \(s\) and \(\epsilon_{s,n}\) is the white noise disturbance;

The means, standard deviations and variances of the different series gotten for each sector would be examined where the series with a lesser variance being the most efficient. Their skewness and kurtosis will also be determined. Also, the existence of stationarity, unit root and long memory properties would be determined.

2.4. Univariate Kalman Filter for the computation of the likelihood function for climatic projection

Parameter estimation that specifies the state-space model is very important in analyzing various components of the time series model. The idea is that
\[
\Theta = \{ \mu_0, \Sigma_0, \varphi, Q, R, Y, \Gamma \}
\]

used to represent the vector of unknown parameters containing the elements of the initial mean \(\mu_0\), the covariance \(\Sigma_0\), the transition matrix, \(\varphi\), the state and observation covariance matrices, \(Q\) and \(R\) are inputs \(Y\) and \(\Gamma\) are estimated using the maximum likelihood estimation. The maximum likelihood, for a time series model where the observations \(y_1, y_2, \ldots, y_n\) are not independent, is defined as a conditional probability density function to write the joint density function
\[
L(Y : \theta) = \prod_{t=1}^{n} p(y_t | Y_{t-1})
\]  
(7)

Where \(p(y_t | Y_{t-1})\) is the distribution of \(y_t\) conditional on the information set at time \(t - 1\) that is \(Y_{t-1} = \{ y_1, y_2, \ldots, y_{t-1} \}\). The maximum likelihood is used under the assumption that the initial state is normal \(X_0 \sim N(\mu_0 : \Sigma_0)\) and the error \(V_1, V_2, \ldots, V_n\) and \(W_1, W_2, \ldots, W_n\) are jointly normal and uncorrelated vector variables. \(W_t \sim N(0 : Q)\) and \(V_t \sim N(0 : \sigma^2)\).

Hence, the likelihood is derived by using the innovations
\[
\epsilon_t = y_t - A_t X_{t-1} - \Gamma_{\mu t}
\]  
(8)

Which are independent normal where \(E(\epsilon_t) = 0\) and the covariance matrix
\[
\sum_0 = A_t P_{t-1} A_t + R
\]  
(9)

Where the dependence of the innovation on the parameter \(\theta\) has been emphasized using the Kalman filter for given \(\theta\), this can be done as follow:

1. Select initial and starting values for the parameter \((\theta^0)\)
2. For $\theta^*$ compute the likelihood $L(y : \theta^*)$ using the Kalman filter
3. Apply a numerical optimization algorithm to $L(y : \theta^*)$
4. Repeat this process for $n$ steps until the value of $\theta$ corresponding to the maximum likelihood is found.

2.5. Forecasting with Kalman filter

An m-period-ahead forecast of the state vector can be calculated from the equation below:

$$\hat{\xi}_{t+k} = F^k \xi_t + F^{k-1} V_{t+1} + F^{k-2} V_{t+2} + \ldots + F^1 V_{t+k-1} + V_{t+k} \text{ for } k = 1, 2,$$

Lead to equation

$$\hat{\xi}_{t+k} = E(\xi_{t+k} | y_t, x_t, y_{t-1}, \ldots, x_{t-1}, y_{t-1}, \ldots, x_1) = F^k \hat{\xi}_t$$  \hspace{1cm} (11)

The error of this forecast can be found by subtracting $\hat{\xi}_{t+k}$ from $\hat{\xi}_{t+k}$$

$$\xi_{t+k} - \hat{\xi}_{t+k} = F^k (\xi_t - \hat{\xi}_t) + F^{k-1} V_{t+1} + F^{k-2} V_{t+2} + \ldots + F^1 V_{t+k-1} + V_{t+k}$$

from which it follows that the mean squared error of the forecast is

$$P_{t+k | t} = E[(\xi_{t+k} - \hat{\xi}_{t+k})(\xi_{t+k} - \hat{\xi}_{t+k})']$$  \hspace{1cm} (13)

$$= F^k P_{t | t} (F^k)' + F^{k-1} Q (F^{k-1})' + F^{k-2} Q (F^{k-2})' + \ldots + F Q F' + Q$$  \hspace{1cm} (14)

These results can also be used to describe m-period-ahead forecasts of the observed vector $y_{t+m}$, provided that $x_t$ is deterministic. Applying the law of iterated expectations to

$$E(y_{t+k} | \hat{\xi}_t, \hat{\xi}_{t-1}, \ldots, \hat{\xi}_t)$$

$$E((Ax_{t+k}, + H' \xi_{t+k} + W_{t+k}) | \hat{\xi}_t, \hat{\xi}_{t-1}, \ldots, \hat{\xi}_{t}, \hat{\xi}_{t-1}, \ldots, x_{t}, y_{t-1}, \ldots)$$  \hspace{1cm} (15)

results in

$$\hat{y}_{t+k} = E(y_{t+k} | \hat{\xi}_t, \hat{\xi}_{t-1}, \ldots, y_t, y_{t-1}, \ldots)$$  \hspace{1cm} (16)

The error of this forecast is

$$y_{t+k} - \hat{y}_{t+k} = (A' x_{t+k} + H' \xi_{t+k} + W_{t+k}) - (A' \hat{x}_{t+k} + H' \hat{\xi}_{t+k})$$  \hspace{1cm} (17)

$$= H' (\xi_{t+m} - \hat{\xi}_{t+m}) + W_{t+m}$$  \hspace{1cm} (18)

With mean squared error

$$E[(y_{t+k} - \hat{y}_{t+k} | y_{t+k} - \hat{y}_{t+k})'] = H' P_{t+k | t} H + R$$  \hspace{1cm} (19)

3. Presentation of Results and Discussion

Figures 1-2 represent the time plot of temperature and rainfall;

![Figure 1. The time plot of the monthly temperature series shows the underlying trend and possible seasonal and cyclic patterns as well](image)

![Figure 2. The time plot of the monthly rainfall series shows the underlying trend and possible seasonal and cyclic patterns as well](image)

3.1. Fourier Analysis

Table 1 shows the values of the various components of the spectral analysis for temperature. The numbers in parentheses, $(d, D, s, M, T)$, are defined as follows: $d$ is the regular differencing order, $D$ is the seasonal differencing order, $s$ is the number of seasons (ignorad if $D$ is 0), $M$ is 1 if the mean is subtracted, 0 otherwise, $T$ is 1 if the trend is subtracted, 0 otherwise. (0, 0, 12, 1, 0) indicates that there is no regular differencing. The seasonal differencing is zero, while the number of seasons is zero. The value indicates that the mean is subtracted, while the trend is not subtracted.

Figures 3-6 show that there exists an underlying periodic component in the residuals obtained by fitting the smoothed and filtered data to the observation. We can see that the first peak looking at the monthly series corresponding to the fitted data is at a frequency $\omega = 0.389$, corresponding to a period of 50 months. And the second peak is at a frequency of $\omega = 0.35$, with a period of approximately 12 months.

Therefore, the need to extend the structural model to contain a Seasonal and cyclic component is important.
Table 1. Fourier Analysis of temp (0,0,12,1,0)

| Frequency | Wavelength | Period | Cosine(a’s) | Sine(b’s) | Spectrum |
|-----------|------------|--------|-------------|-----------|----------|
| 0.2010619 | 31.25      | 17392.62 | 21.48521    | -14.59315 | 9522.42  |
| 0.2764601 | 22.72727   | 2180.31 | 0.8657146   | -9.155004 | 7795.009 |
| 0.3518584 | 17.85714   | 3812.099 | 11.42719    | -4.155941 | 4270.688 |
| 0.4272566 | 14.70588   | 6819.656 | -14.50404   | -7.357564 | 15927.96 |
| 0.5026549 | 12.5       | 37152.14 | -6.556379   | 37.38934  | 17404.13 |
| 0.5780531 | 10.86957   | 8240.605 | -13.99402   | -11.12565 | 15948.32 |
| 0.6534513 | 9.615385   | 2452.211 | -7.719598   | 5.959619  | 3608.136 |
| 0.7288495 | 8.620689   | 131.5911 | 2.229034    | -0.3676695| 2495.332 |
| 0.8042477 | 7.8125     | 4902.192 | -12.52921   | -5.757686 | 6569.818 |
| 0.8796459 | 7.142857   | 14675.67 | 4.816339    | 23.36665  | 7431.676 |
| 0.9550442 | 6.578948   | 2717.165 | 13.99337    | -11.12565 | 15948.32 |
| 1.030442  | 6.097561   | 28936.85 | 16.04339    | -25.16181 | 12334.13 |
| 1.105841  | 5.681818   | 5348.37 | 3.409208    | -13.99337 | 11674.39 |
| 1.181239  | 5.319149   | 737.9613 | 4.387738    | 5851.817  |
| 1.256637  | 5          | 11469.12 | 25.16181    | 12334.13  |
| 1.332035  | 4.716981   | 3892.745 | 11.29772    | 5620.186  |
| 1.407434  | 4.464286   | 1498.693 | 7.624041    | 3149.379  |
| 1.482832  | 4.237288   | 4056.698 | -3.564355   | 4142.489  |
| 1.55823   | 4.032258   | 3174.855 | 4.774079    | 4193.151  |
| 1.633628  | 3.846154   | 5347.899 | 3.591002    | 4181.128  |
| 1.709026  | 3.676471   | 4020.631 | 16.04339    | 5366.608  |
| 1.784425  | 3.521127   | 3058.936 | 7.860816    | 2993.722  |
| 1.859823  | 3.378378   | 1901.597 | 8.331114    | 1728.495  |
| 1.935221  | 3.246753   | 224.9497 | 10.01725    | 1347.075  |
| 2.010619  | 3.125      | 1914.677 | 2.08475     | 1418.189  |
| 2.086018  | 3.012048   | 2114.941 | -8.523886   | 4419.094  |
| 2.161416  | 2.906977   | 9227.663 | -16.48055   | 4187.763  |
| 2.236814  | 2.808989   | 1220.684 | 0.3565736   | 6609.169  |
| 2.312212  | 2.717391   | 9379.16 | 16.41459    | 5221.976  |
| 2.38761   | 2.631579   | 5066.084 | 8.594324    | 5238.45   |
| 2.463009  | 2.55102    | 1270.106 | 4.720753    | 2286.001  |
| 2.538407  | 2.475248   | 521.812 | 0.8846894   | 4.410879  | 1232.752 |
| 2.613805  | 2.403846   | 1906.337 | -8.432693   | 1.681413  | 1483.803 |
| 2.689203  | 2.336449   | 2023.261 | 6.147003    | 1774.629  |
| 2.764601  | 2.272727   | 1394.288 | -0.8783695  | 7.3011    | 2174.107 |
| 2.84      | 2.212389   | 3104.772 | -10.88803   | -1.367346 | 4679.055 |
| 2.915398  | 2.155172   | 9538.105 | -18.3482    | 5.768869  | 4346.808 |
| 2.990796  | 2.10084    | 397.5458 | -3.743783   | 1.184457  | 3586.218 |
| 3.066195  | 2.04918    | 823.0026 | -5.125026   | 2.377886  | 668.0311 |
| 3.141593  | 2         | 783.5449 | -5.5127     | -2.484305E-12 | 809.85   |
Table 2. Fourier Analysis of rainfall (0,0,12,1,0)

| Frequency | Wavelength | Period  | Cosine(a’s) | Sine(b’s) | Spectrum       |
|-----------|------------|---------|-------------|-----------|----------------|
| 0.2010619 | 31.25      | 3505336 | -155.2216   | 334.4563  | 1.340589E+07   |
| 0.2764601 | 22.72727   | 2.376145E+07 | -473.745   | 834.9586  | 1.081362E+07   |
| 0.3518584 | 17.85714   | 5174079 | 204.4676    | 398.5852  | 1.242262E+07   |
| 0.4272566 | 14.70588   | 8332332 | 372.0547    | 429.8204  | 8652136        |
| 0.5026549 | 12.5       | 1.245E+07 | -522.3141   | 458.3256  | 1.045092E+07   |
| 0.5780531 | 10.86957   | 1.057043E+07 | 496.6129   | -404.1668 | 2.557089E+07   |
| 0.6534513 | 9.615385   | 5.369223E+07 | 347.9105   | -1400.506 | 2.371912E+07   |
| 0.7288495 | 8.620689   | 6894689 | 310.2988    | -413.6736 | 2.459778E+07   |
| 0.8042477 | 7.8125     | 1.320643E+07 | 271.7906   | -662.0743 | 1.099229E+07   |
| 0.8796459 | 7.142857   | 1.287576E+07 | -420.4702  | -567.9723 | 9935063        |
| 0.9550442 | 6.578948   | 3723000 | -203.048    | -321.1984 | 1.345724E+07   |
| 1.105841  | 5.681818   | 7312064 | 418.9944    | 328.6986  | 1.113207E+07   |
| 1.181239  | 5.319149   | 2311208 | -185.0102   | 235.3968  | 5046145        |
| 1.256637  | 5          | 5515162 | 108.9928    | -449.4738 | 3876995        |
| 1.332035  | 4.716981   | 3804614 | -342.3183   | 174.3     | 4267821        |
| 1.407434  | 4.464286   | 3483686 | -122.2476   | -346.6563 | 5960738        |
| 1.482832  | 4.237288   | 1.059391E+07 | -237.493   | -595.3847 | 2.304692E+07   |
| 1.55823   | 4.032258   | 5.506316E+07 | -1014.657  | -1051.713 | 2.313139E+07   |
| 1.633628  | 3.846154   | 3737106 | 182.7847    | -333.9667 | 1.996097E+07   |
| 1.709026  | 3.676471   | 1082651 | 77.12722    | -189.8476 | 1860007        |
| 1.784425  | 3.521127   | 778264.9 | -13.22315   | 173.2346  | 1592170        |
| 1.859823  | 3.378378   | 2915595 | 331.0794    | -58.89026 | 1575621        |
| 1.935221  | 3.246753   | 1033002 | 79.65166    | 183.632   | 2011053        |
| 2.010619  | 3.125      | 2084562 | 141.3623    | 246.7116  | 3220544        |
| 2.086018  | 3.012048   | 6544069 | -469.2179   | 183.4309  | 5135841        |
| 2.161416  | 2.906977   | 6778893 | -507.325    | -74.44065 | 5237774        |
| 2.236814  | 2.808989   | 2390359 | -264.1649   | 151.4172  | 4376703        |
| 2.312212  | 2.717391   | 3960858 | -391.2391   | 23.54103  | 2330373        |
| 2.38761   | 2.631579   | 639903.6 | 54.25442    | -147.9026 | 2294082        |
| 2.463009  | 2.55102    | 2281486 | -274.1098   | 115.5485  | 979001.7       |
| 2.538407  | 2.475248   | 15615.94 | 5.175655    | -24.05989 | 2194616        |
| 2.613805  | 2.403846   | 4286746 | -333.2503   | -234.9598 | 1575710        |
| 2.689203  | 2.336449   | 424768.6 | -121.5072   | -41.36044 | 1572502        |
| 2.764601  | 2.272727   | 626.016  | -3.307111   | -15.22817 | 806171.2       |
| 2.84      | 2.212389   | 1987484 | 64.74786    | 269.9861  | 1018778        |
| 2.915398  | 2.155172   | 1062588 | -102.2991   | 175.3496  | 1285118        |
| 2.990796  | 2.10084    | 805281.5 | 176.6789    | 4.182919  | 2102421        |
| 3.066195  | 2.04918    | 4439392 | 126.9668    | -395.0464 | 2849186        |
| 3.141593  | 2          | 3302885 | -357.9144   | 2.442917E-10 | 4060556    |
Table 2 shows the values of the various components of the spectral analysis for rainfall. The numbers in parentheses, \((d, D, s, M, T)\), are defined as follows: \(d\) is the regular differencing order, \(D\) is the seasonal differencing order, \(s\) is the number of seasons (ignored if \(D\) is 0), \(M\) is 1 if the mean is subtracted, 0 otherwise, \(T\) is 1 if the trend is subtracted, 0 otherwise. \((0, 0, 12, 1, 0)\) indicates that there is no regular differencing. The seasonal differencing is zero, while the number of seasons is zero. The value indicates that the mean is subtracted, while the trend is not subtracted.

Figures 7-10 show that there exists an underlying periodic component in the residuals obtained by fitting the smoothed and filtered data to the observation. We can see that the first peak looking at the monthly series corresponding to the fitted data
3.2. State-space model

From Table 3, the AR coefficients with respect to the MLEs are given as 1.271 and -0.274, respectively also, the roots of the characteristic equation \( \phi = 1 - 1.271z + 0.274z^2 = 0 \) are 0.542 and 1.998, respectively. Since these values are greater than 1, the AR component is covariance stationary.

The variance of the permanent component is 0.000000, and the variance of the transitory component is 2.648752906. Hence, the variance of the transitory component is higher than the variance of the permanent component and the ratio of the variance of the permanent component to the variance of the stationary component is 0.000. This shows that the stationary component is almost twice as important as the permanent component for explaining the variation of temperature.

### 3.3. Forecasting

The in-sample performance of the state-space model for forecasting the rainfall and temperature series appears to be more favourable to the model identified by Shittu and Yemitan [11]. This is evident in Tables 5 and 6.

Given the relative accuracy of the model by Shittu and Yemitan [11], the improvement achieved by the Kalman filter method is mainly as a result of the built-in the specification for updating the estimation based on latest available information. Hence, the rainfall and temperature exhibit elements of time-varying or regime-switching characteristics over the study period. As a result, it is an indication and a signal that the climatic data may be best estimated using non-linear time series methodologies.

### 4. Conclusion

This study modelled and estimated climatic data using the state-space model. The study was specifically to identify the pattern of the trend movement, model the dataset using the state-space model and to evaluate the forecasting power of the state-space models.

The data used for the study include temperature and rainfall for periods January 1991 to December 2017. The data were tested for normality. The study showed that the average temperature is 27.3\(^\circ\)C and standard deviation is 1.87\(^\circ\)C. The maximum temperature is 31.5\(^\circ\)C and minimum temperature is 23.3\(^\circ\)C. The average rainfall is 94.5mm with a standard deviation of 86.3mm. The maximum rainfall is 314.3mm and the minimum rainfall is 0.2mm. Shapiro-Wilk, Anderson-Darling and Kolmogorov-Smirnov test of normality for the climatic data all showed that the variables are not normally distributed. The plot of the monthly temperature series shows the underlying trend and possible seasonal and cyclic patterns as well. The plot of the monthly rainfall series shows the underlying trend and possible seasonal and cyclic patterns as well.

The MLEs for the AR coefficients are 1.311 and -0.456, respectively. The roots of the characteristic equation \( \phi = 1 - 1.311z + 0.456z^2 = 0 \) are 1.779 and 1.137, respectively. Since these values are greater than 1, the AR component is covariance stationary.

The variance of the permanent component is 0.000000, and the variance of the transitory component is 2.648752906. Hence, the variance of the transitory component is higher than the variance of the permanent component and the ratio of the variance of the permanent component to the variance of the stationary component is 0.000. This shows that the stationary component is almost twice as important as the permanent component for explaining the variation of temperature.
### Table 3. State-space Model (Temperature)

| Coefficient | Std. Error | z-Statistic | Prob.   |
|-------------|------------|-------------|---------|
| C(1)        | 0.818740   | 10.32875    | 0.0000  |
| C(2)        | 1.271171   | 22.88260    | 0.0000  |
| C(3)        | -0.274151  | -4.874137   | 0.0000  |

**Final State**

| SV1         | Root MSE   | z-Statistic | Prob.   |
|-------------|------------|-------------|---------|
| 22.24512    | 1.505869   | 14.77229    | 0.0000  |

**Log likelihood** -600.6427

**Akaike info criterion** 3.726190

**Schwarz criterion** 3.761197

**Hannan-Quinn criter.** 3.740162

### Table 4. State-space Model (Rainfall)

| Coefficient | Std. Error | z-Statistic | Prob.   |
|-------------|------------|-------------|---------|
| C(1)        | 7.792711   | 115.8147    | 0.0000  |
| C(2)        | 1.311527   | 26.69830    | 0.0000  |
| C(3)        | -0.456845  | -8.697765   | 0.0000  |

**Final State**

| SV1         | Root MSE   | z-Statistic | Prob.   |
|-------------|------------|-------------|---------|
| -1.907737   | 49.22274   | -0.038757   | 0.9691  |

**Log likelihood** -1723.221

**Akaike info criterion** 10.65568

**Schwarz criterion** 10.69069

**Hannan-Quinn criter.** 10.66966
### Table 5. Forecast for Temperature

| Period | Forecast (SSM) | 95% Limits   |
|--------|----------------|--------------|
|        |                | Lower  | Upper  |
| 2018   | 24.6816        | 21.8312 | 27.5320 |
| 2018   | 25.6077        | 22.2266 | 28.9888 |
| 2018   | 26.1985        | 22.6239 | 29.7730 |
| 2018   | 26.5754        | 22.9250 | 30.2258 |
| 2018   | 26.8158        | 23.1351 | 30.4966 |
| 2018   | 26.9692        | 23.2762 | 30.6623 |
| 2018   | 27.0671        | 23.3690 | 30.7652 |
| 2018   | 27.1295        | 23.4294 | 30.8296 |
| 2018   | 27.1694        | 23.4684 | 30.8703 |
| 2018   | 27.1948        | 23.4935 | 30.8960 |
| 2018   | 27.2110        | 23.5096 | 30.9124 |
| 2018   | 27.2213        | 23.5199 | 30.9228 |
| 2019   | 27.2279        | 23.5264 | 30.9294 |
| 2019   | 27.2321        | 23.5306 | 30.9336 |
| 2019   | 27.2348        | 23.5333 | 30.9363 |
| 2019   | 27.2365        | 23.5350 | 30.9380 |
| 2019   | 27.2376        | 23.5361 | 30.9391 |
| 2019   | 27.2383        | 23.5368 | 30.9398 |
| 2019   | 27.2388        | 23.5373 | 30.9403 |
| 2019   | 27.2391        | 23.5375 | 30.9406 |
| 2019   | 27.2392        | 23.5377 | 30.9407 |
| 2019   | 27.2393        | 23.5378 | 30.9409 |
| 2019   | 27.2394        | 23.5379 | 30.9409 |
| 2019   | 27.2395        | 23.5380 | 30.9410 |

### Table 6. Forecast for Rainfall

| Period | Forecast (SSM) | 95% Limits   |
|--------|----------------|--------------|
|        |                | Lower  | Upper  |
| 2018   | 21.9472        | -83.629 | 127.524 |
| 2018   | 37.1513        | -97.037 | 171.340 |
| 2018   | 49.0790        | -100.014 | 198.172 |
| 2018   | 58.4364        | -99.130 | 216.003 |
| 2018   | 65.7775        | -96.785 | 228.340 |
| 2018   | 71.5366        | -94.025 | 237.099 |
| 2018   | 76.0547        | -91.327 | 243.436 |
| 2018   | 79.5992        | -88.892 | 248.091 |
| 2018   | 82.3799        | -86.791 | 251.551 |
| 2018   | 84.5614        | -85.027 | 254.149 |
| 2018   | 86.2728        | -83.571 | 256.117 |
| 2018   | 87.6154        | -82.386 | 257.617 |
| 2019   | 88.6687        | -81.430 | 258.767 |
| 2019   | 89.4950        | -80.663 | 259.653 |
| 2019   | 90.1433        | -80.051 | 260.338 |
| 2019   | 90.6518        | -79.565 | 260.869 |
| 2019   | 91.0508        | -79.180 | 261.282 |
| 2019   | 91.3638        | -78.876 | 261.603 |
| 2019   | 91.6094        | -78.635 | 261.854 |
| 2019   | 91.8020        | -78.446 | 262.050 |
| 2019   | 91.9531        | -78.297 | 262.203 |
| 2019   | 92.0717        | -78.180 | 262.323 |
| 2019   | 92.1647        | -78.087 | 262.417 |
| 2019   | 92.2377        | -78.015 | 262.490 |
MLEs for the AR coefficients are 1.311 and -0.456, respectively. The roots of the characteristic equation $\phi(z) = 1 - 1.311z + 0.456z^2 = 0$ are 1.779 and 1.137, respectively.

Investigations and diagnostic were carried out by checking for correlations in the residuals and also checking for periodicity in the residuals. The results of this investigation show significant autocorrelation in the residuals indicating the presence of underlying noise terms which is not accounted for. Also, the result from the parametric spectral estimate shows underlying periodic patterns for monthly data, thus, leads to a discussion on the need to treat climatic data as a structural time series model. We selected the appropriate models by considering the goodness of fit of the models and by comparing the AIC values. Parameters were estimated and accomplished with some measures of precision. An important aspect of fitting structural models is the underlying changes in the unobserved trend component.

For the case of state-space models with correlated errors, further work can be done here by taking into consideration correlation between the error terms in state equation when considering structural models of the form trend, seasonal, cycle and an autoregressive process. This type of results is important since they show how well the Kalman filter for correlated errors carries out the filtering of the unobserved component as compared to the case when errors are not correlated.

References

[1] S. Makridakis & M. Hibon, “The M3-Competition: results, conclusions and implications”, International Journal of Forecasting, 16(2000), 451.
[2] B. Shamshad, M. Z Khan & Z. Omar, “Modeling and forecasting weather parameters using ANN-MLP, ARIMA and ETS model: a case study for Lahore, Pakistan”, Journal of Applied Statistics 5 (2019) 388.
[3] S. Afzar, N. Abbas, & B. Jan, “Comparative study of temperature and rainfall fluctuation in Hunza-Nagar District”, Journal of Basic and Applied Sciences 9 (2013) 151.
[4] N. Faisal & A. Ghaffar, “Development of Pakistan’s new area weighted rainfall using Thiessen polygon method”, Pakistan Journal of Metrology 17 (2012) 107.
[5] F. Yusof, I. L. Kane & Z. Yusop, “Structural break or long memory: An empirical survey on daily rainfall data sets across Malaysia”, Hydrology and Earth System Sciences 17 (2013) 1311.
[6] M. Zahid & G. Rasul, “Frequency of extreme temperature and precipitation events in Pakistan 1965-2009”, Science International 23 (2011) 313.
[7] J. Abbot, & J. Marohasy, “Application of artificial neural networks to rainfall in Queensland, Australia”, Advances in Atmospheric Sciences 29 (2012) 717.
[8] M. Bilgili & B. Sahin, “Prediction of long-term monthly temperature and rainfall in Turkey, Energy Sources 1 (2010) 60.
[9] R. D. Johnston, “Steady-state closed-loop structural interaction analysis”, International Journal of Control 6 (1990) 1351.
[10] J. Wang, & K. M. Cameron, “A method of constructing a modified signal-flow-graph”, International Journal of Control 12 (1993) 305.
[11] O. I. Shittu & R. A. Yemitan, “Inflation targeting with dynamic time series modelling - the Nigerian case”, International Journal of Scientific and Engineering Research 9 (2014) 206.