

ANALYSIS OF KANO METEOROLOGICAL DATA USING TIME SERIES ANALYSIS AND EMPIRICAL ORTHOGONAL FUNCTIONS

R. Aliyu¹, B.I. Tijjani², U.M Gana², S Bala², S. B Sharafa,³ S Uba⁴, S
Auwalu⁵, S.U.Yerima⁶, U Y Abdulkarim⁷, M. Idris²

¹Department of Physics, Kano University of Science and Technology Wudil. Nigeria

²Department of Physics, Bayero University, Kano, Kano State, Nigeria

³Department of Physics, Usmanu Danfodiyo University Sokoto, Sokoto, State, Nigeria

⁴Department of Physics, Yusuf Maitama Sule University, Kano. Nigeria

⁵Department of Physics, Sa'adatu Rimi College of Education, Kano. Nigeria

⁶Nigerian Meteorological Agency (NIMET), Nnamdi Azikwe International Airport Abuja

⁷Department of Physics, Federal Collage of Education Kano, Kano State, Nigeria

Contact email: rakiyaaliyu238@yahoo.com, rukyna8@gmail.com

ABSTRACT

In this work, two different statistical techniques were used to analyses the meteorological data of Kano State. The meteorological parameters utilized are daily Solar Radiations, Sunshine Hours, wind speed, Maximum and Minimum Temperatures, Rainfall, Cloud Cover, and Relative Humidity, data spanning for thirty-one years (1980 to 2010). The statistical analysis involves the use of time series analysis and empirical orthogonal functions. In the time series analysis (TS) all the parameters are assumed independent. The second is the Empirical Orthogonal Transformation (EOF) in which the data were analyzed using unrotated and orthogonal transformations and six components were extracted. From component matrices, it is discovered that there are two distinct season as Rainy and Dry seasons. The rainy season has two components comprising heavy rain and light rain, while the dry season comprises of three different types of seasons. The period of heavy rain is around 3.3 months and period of light rain is 1.6 month. These gives a total of 4.9 months for rainy season and 7.1 months for dry season. The values of the eigen values are consistent with what is observed in real-time of seven (7.1) months of dry season and (4.7) months of rainy season. Now combining the two results, it can be said that the TS analysis show that all the parameters have seasonal variation, and the EOF now described the two seasons quantitatively

Keywords: meteorological data, regression analysis, time series analysis, empirical orthogonal transformations, oblique transformations.

1 INTRODUCTION

Climate is seen as the state of the atmosphere of an area over a long period of time. It is the synthesis of weather over a long period of time at a given area (Ayoade, 2011). Climate can be defined, in mathematical terms, as the aggregation of all long-term statistical properties of the atmospheric state or, as defined by Lorenz (1970), Climate is regarded as the aggregation of (random) daily weather. It is, therefore, the long-term statistics of weather. Climate variations are also the result of exceedingly complex non-linear interactions between very many degrees of freedom or modes.

A time series is a time-oriented or chronological sequence of observations on a variable of interest (Montgomery et al., 2008). Mostly, these observations are collected at equally spaced, discrete time intervals. Time series models have become popular in recent years since the publication of a book by Box and Jenkins (1970), and the subsequent development of computer software for applying these models (Bell, 1984).

The most basic assumption in any time series analysis modeling is that some aspects of the past pattern will continue to remain in the future.

The objective of time series analysis is generally to understand and identify the stochastic process that produced the observed series and then to forecast future values of a series from past values alone (Akgun, 2003).

Empirical Orthogonal Function (EOF) analysis, also called principal component analysis (PCA) (Fukuoka, 1951; Lorenz, 1956), is among the most widely and extensively used methods in dimensionality reduction and patterns extraction in atmospheric science. Both names are commonly used and refer to the same set of procedures. Typically, the EOFs are found by computing the eigen values and eigenvectors of a spatially weighted anomaly covariance matrix of a field. The derived eigen values provide a measure of the percent variance explained by each mode.

EOF analysis and PCA have become the standard statistical techniques in the geophysical sciences of meteorology, climatology, and oceanography (e.g., Preisendorfer, 1988; Emery and Thomson, 1997), particularly in climate research (Peixoto and Oort, 1992; von Storch and Zwiers, 1999).

The aim of this paper is to analyze the meteorological data of Kano using time series analysis and empirical orthogonal transformations. The purpose of the time series is to determine the number of seasons and for the empirical orthogonal transformation is to determine the characteristics of the seasons and their durations.

2 THEORY

While the theory on mathematically oriented time series analysis is vast and mostly difficult to non-statistician, the focus of this presentation is directed at data analysis. Some basic properties of time series processes and models are presented. These focus mostly on how to visualize and describe time series data, on how to fit models to data correctly, on how to generate forecasts and on how to adequately draw

conclusions from the output that was produced.

2.1 TIME SERIES COMPONENTS AND DECOMPOSITION

An important step in analyzing TS data is to consider types of data patterns, so that the models most appropriate to those patterns can be utilized. Four types of time series components can be distinguished. They are:

- (i) Horizontal – when data values fluctuate around a constant value.
- (ii) Trend – when there is a long-term increase or decrease in the data.
- (iii) Seasonal – when a series is influenced by season factor and reoccurs on regular periodic basis.
- (iv) Cyclic – when the data exhibit rises and fall that are not of a fixed period.

Many data series include combinations of the preceding patterns. After separating out the existing pattern in any time series data, the pattern that remains unidentifiable form the 'random' or 'error' component. Time plot (data plotted over time) and seasonal plot (data plotted against individual seasons in which the data were observed) help in visualizing these patterns while exploring the data. A crude, yet practical, way of decomposing the original data (ignoring cyclic pattern) is to go for a seasonal decomposition either by assuming an additive or multiplicative model viz.

$$Y_t = T_t + S_t + E_t \quad (2.1)$$

$$\text{or } Y_t = T_t \cdot S_t \cdot E_t, \quad (2.2)$$

where, Y_t = the original TS data, T_t = Trend component, S_t = Seasonal component and E_t = Error / Irregular component.

Equation (2.1) is called an additive seasonal model if for a given time series $Y_1; \dots; Y_t$ there is the assumption that these data are realizations of random variables Y_t that are themselves sums of four components. Equation (2.2) is called a multiplicative seasonal model; this is appropriate for a time series in which the amplitude of the seasonal pattern is proportional to the average level of the series, i.e., a time series displaying multiplicative seasonality. In other words, if the magnitude of a TS varies with the level

of the series, then one has to go for a multiplicative model, otherwise called an additive model.

This decomposition may enable one to study the TS components separately or will allow analysts to de-trend or to do seasonal adjustments, if needed, for further analysis.

2.2 ARIMA MODELS

ARIMA is an abbreviation of Autoregressive Integrated Moving Average introduced by Box and Jenkins (Box et.al., 1994). As such, some authors refer to this modeling approach as Box and Jenkins model. Box-Jenkins model is a stationary time series model. Time series that generated from zero-mean, finite variance, and uncorrelated variable are called 'white noise'. Many useful models can be constructed from them.

Most physical processes exhibit inertia and do not change that quickly. This, combined with the sampling frequency, often makes consecutive observations correlate. Such correlation between consecutive observations is called autocorrelation. When the data are auto correlated, most of the standard modeling methods, based on the assumption of independent observations may become misleading or sometimes even useless. We, therefore, need to consider alternative methods that consider the serial dependence in the data. This can be easily achieved by employing time series models such as the autoregressive integrated moving average (ARIMA) models.

ARIMA, which is often called method of Box-Jenkins's time series, has appreciable accuracy for short-term forecasting, but less accuracy for long-term forecasting. Usually, it will tend to become flat for a sufficiently long period. ARIMA model ignores the independent variable completely and uses past and present values of dependent variable to produce accurate short-term forecasting (Hendranata, 2003). ARIMA is suitable when the observation of time series is statistically related to the dependent. The purpose of this model is to determine good

statistical relationships between the variables that are being predicted and the historical value of these variables, so that forecasting can be performed with the model (Hendranata, 2003).

The ARIMA modeling is essentially an exploratory data-oriented approach that has the flexibility of fitting an appropriate model which is adapted from the structure of the datum itself. The stochastic nature of the time series can be approximately modeled with the aid of autocorrelation function and partial autocorrelation function from which information such as trend, random variables, periodic components, cyclic patterns, and serial correlation can be discovered. As a result, forecasts of the future values of the series with some degree of accuracy can be readily obtained (Ho and Xie, 1998).

Although ARIMA modeling is sophisticated in theory, but with the advent of computer technology, the iterative model building process and, hence, accurate forecast, can be aided and made simpler by the ease of many user-friendly statistical software packages such as SAS, SPSS, Statgraphics, Statistica and Minitab. An iterative three stage process, through model identification, parameter estimation and diagnostic check, is required to determine the adequacy of the proposed model (Ho and Xie, 1998).

ARIMA contains three components, namely: AutoRegressive (AR), Integrated (I) and Moving Average (MA) parts. The AR part describes the relationship between present and past observations. The MA part is assumed that the current value of the series is a weighted sum of past white noise terms, and also represents the autocorrelation structure of error. The I part represents the differencing level of the series to eliminate non-stationary (Hasmida, 2009). It is usually denoted by $(p, d, q)(P, D, Q)$ where p denotes order of autoregressive component, d denotes order of differencing, q denotes order of moving average and (P, D, Q) denotes corresponding seasonal component.

AR(p) model expresses that the current value of time series as a linear combination of p previous values or Autoregressive models are based on the idea that the current value of the series, Y_t , can be explained by p past values, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ and a white noise term (random shock) where p determines the number of steps into the past needed to forecast the current value. Bell (1984) expresses the current value of time series of AR(p) model as:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (2.3)$$

where $\phi_1, \dots, \phi_p$ are AR(p) parameters, the ε_t is the random shock in normal distribution with zero mean and variance at time t , and p is the order of AR(p).

By introducing the backshift operator B , which defines ($BY_t = Y_{t-1}$), equation (2.3) can be written as:

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = \varepsilon_t \quad (2.4)$$

$$\text{Or } \phi_p(B) Y_t = \varepsilon_t \quad \text{where } \phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p) \quad (2.5)$$

MA(q) model expresses the current value of a time series as a linear combination of a current and q previous values of a white noise process. The (purely) moving average (MA) model is (Bell, 1984):

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.6)$$

$$\text{Or } Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad \text{or} \quad Y_t = \theta_q(B) \varepsilon_t \quad (2.7)$$

where q is the order of MA(q), and $\theta_q(B)$ are parameters of MA(q).

To increase flexibility when fitting actual time series, both autoregressive and moving average operators are combined to give the ARMA (p,q) model (Bell, 1984): coefficients are MA(q) model parameters.

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.8)$$

Which can be written as:

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad \text{or } \phi_p(B) Y_t = \theta_q(B) \varepsilon_t \quad (2.9)$$

The mixed type of series which are explained both by its own lagged values and by lagged noise terms, is called Autoregressive Moving-Average models of

order (p,q). This systematic class of stationary time series models carries great importance and usefulness especially in real-life situations. If the process is stationary, a suitable ARMA model can be used to represent the data. If it is nonstationary, differencing is applied to make the model become stationary and this leads to ARIMA model (Akgun, 2003).

The first and most important condition of the series Y_t in equation (2.8) or (2.9) is that it must be stationary. In practice Y_t may well be nonstationary, but has to be made stationary by first difference,

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - B) Y_t \quad (2.10)$$

If equation (3.10) ie $(1-B)Y_t$ is nonstationary, then there will be the need to take the second difference,

$$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - B)^2 Y_t \quad (2.11)$$

In general, we may need to take the d^{th} difference $(1-B)^d Y_t$ (although rarely is d larger than 2). Substituting $(1-B)^d Y_t$ for Y_t in (3.9) yields the ARIMA (p,d,q) model (Bell, 1984):

$$\phi_p(B) (1 - B)^d Y_t = \theta_q(B) \varepsilon_t \quad (2.12)$$

where d is the order of differencing.

When a time series exhibits potential seasonality indexed by s , using a multiplied seasonal ARIMA(p,d,q)(P,D,Q)^s model is advantageous. The seasonal time series is transformed into a stationary time series with non-periodic trend components. A multiplied seasonal ARIMA model can be expressed as (Lee and Ko, 2011):

$$\phi_p(B) \Phi_P(B^s) \nabla^d \nabla_s^D Y_t = \theta_q(B) \Theta_Q(B^s) \varepsilon_t$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 \dots -$$

$$\phi_p B^p, \theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \dots - \Phi_P B^{Ps},$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} \dots - \Theta_Q B^{Qs}$$

B is the backshift operator (i.e. $BY_t = Y_{t-1}$, $B^2 Y_t = Y_{t-2}$, $B^{12} Y_t = Y_{t-12}$ and so on) s , is the seasonal lag, and ε_t is a sequence of independent normal error variables with mean zero and variance σ^2 , D is the order of seasonal differencing, $\Phi_P(B^s)$ and $\Theta_Q(B^s)$

are the seasonal AR(p) and MA(q) operators respectively.

2.3 EXPONENTIAL SMOOTHING

Exponential smoothing is a procedure for continually revising a forecast in the light of more recent experience. Exponential Smoothing assigns exponentially decreasing weights as the observation gets older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

2.3.1 Single exponential smoothing

This is also known as simple exponential smoothing. Simple smoothing is used for short-range forecasting, usually just one month into the future. The model assumes that the data fluctuates around a reasonably stable mean or level (no trend or consistent pattern of growth).

The specific formula for simple exponential smoothing is:

$$L_t = \alpha Y_t + (1 - \alpha)L_{t-1} \quad (2.14)$$

where α is a smoothing constant between 0 and 1.

When applied recursively to each successive observation in the series, each new smoothed value (forecast) is computed as the weighted average of the current observation and the previous smoothed observation; the previous smoothed observation was computed in turn from the previous observed value and the smoothed value before the previous observation, and so on.

Thus, in effect, each smoothed value is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter (α). If it is equal to 1 (one) then the previous observations are ignored entirely; if it is equal to 0 (zero), then the current observation is ignored entirely, and the smoothed value consists entirely of the previous smoothed value (which in turn is computed from the smoothed observation before it, and so on; thus, all smoothed values will be equal to the initial smoothed

value L_0). In-between values will produce intermediate results.

2.3.2 Double Exponential Smoothing

This method is used when the data show a trend. Exponential smoothing with a trend works much like simple smoothing except those two components must be updated each period - level and trend. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. The specific formulas for double exponential smoothing are:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad 0 < \alpha < 1 \quad (2.15)$$

$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \quad 0 < \gamma < 1 \quad (2.16)$$

Note that the current value of the series is used to calculate its smoothed value replacement in double exponential smoothing.

2.3.3 Triple Exponential Smoothing

This method is used when the data show level, trend, and seasonality. To handle seasonality, we must add a third parameter. We now introduce a third equation to take care of seasonality. The resulting set of equations is called the "Holt-Winters" (HW) method, after the names of the inventors.

$$\text{Estimate of the level } L_t = \alpha(Y_t/S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.17)$$

$$\text{Estimate of the growth rate (or trend)} \quad T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \quad (2.18)$$

$$\text{Estimate of the seasonal factor } S_t = \delta(Y_t/L_t) + (1 - \delta)S_{t-s} \quad (2.19)$$

where α , γ , and δ are smoothing constants between 0 and 1, s = number of seasons in a year ($s = 12$ for monthly data, and $s = 4$ for quarterly data).

2.4 EMPIRICAL ORTHOGONAL TRANSFORMATIONS

The data to be used for the EOF analysis are transformed into a two-dimensional data matrix by the software, X (say) as follows:

$$X = M \begin{bmatrix} & N \\ & \end{bmatrix} = X_{i,j} \text{ where } i = 1, M; j = 1, N \quad (2.20)$$

where M and N are the dimensions of the data matrix enclosed by the square brackets, and the subscript notation $X_{i,j}$ to indicate the same matrix. The transpose of the matrix is obtained by reversing the order of the indices to make it an NxM matrix.

$$X^T = N \begin{bmatrix} & M \\ & \end{bmatrix} = X_{j,i} \text{ where } i = 1, M; j = 1, N \quad (2.21)$$

In multiplying a matrix time, itself we generally need to transpose it once to form an inner product, which results in two possible “dispersions” matrices, in this case are called correlation/covariance matrices. The first can be obtained as:

$$C = XX^T = M \begin{bmatrix} & N \\ & \end{bmatrix} \begin{bmatrix} & M \\ & \end{bmatrix} N = X_{i,j}X_{j,i} = X_{i,i} = \begin{bmatrix} & M \\ & \end{bmatrix} M \quad (2.22)$$

The other dispersion matrix in which the roles of the structure and sampling variables are reversed.

$$C = X^TX = N \begin{bmatrix} & M \\ & \end{bmatrix} \begin{bmatrix} & N \\ & \end{bmatrix} M = X_{j,i}X_{i,j} = X_{j,j} = \begin{bmatrix} & N \\ & \end{bmatrix} N \quad (2.23)$$

In this projection of a matrix onto itself, one of the dimensions gets removed and we are left with a measure of the dispersion of the structure with itself across the removed dimension (or the sampling dimension). If the sampling dimension is time, then the resulting dispersion matrix is the matrix of the covariance of the spatial locations with each other, as determined by their variations in time. These dispersion matrices are in fact covariance/correlation matrices. In the second case the covariance at different times is obtained by projecting on the sample of different spatial points. Either of these dispersion matrices may be scientifically meaningful, depending on the problem under consideration. These matrices generated can be either positive definite or positive semi-definite covariance/correlation matrices and are usually symmetrical.

EOF (or PCA) analysis consists of an eigenvalue analysis of any one of these dispersion matrices. Any symmetric matrix C can be decomposed through a

diagonalization, or eigen analysis using the followings:

$$Ce_i = \lambda_i e_i \quad (2.24)$$

$$CE = E\Lambda \quad (2.25)$$

where E is the matrix with the eigenvectors e_i as its columns, and Λ is the matrix with the eigenvalues λ_i , along its diagonal and zeros elsewhere.

The eigenfunctions of the Hermitian matrix C form an orthonormal basis to represent C and hence called empirical functions or empirical modes. The eigenvalue λ_i is a measure of the percentage variability represented by i^{th} EOF.

The set of eigenvectors, e_i , and associated eigenvalues, λ_i , represent a coordinate transformation into a coordinate space where the matrix C becomes diagonal. Because the covariance/correlation matrix is diagonal in this new coordinate space, the variations in these new directions are uncorrelated with each other, at least for the sample that has been used to construct the original covariance/correlation matrix.

The eigen vectors define directions in the initial coordinate space along which the maximum possible variance can be explained, and in which variance in one direction is orthogonal to the variance explained by other directions defined by the other eigenvectors. The eigen values indicate how much variance is explained by each eigenvector. If you arrange the eigenvector/eigen value pairs with the biggest eigen values first, then you may be able to explain a large amount of the variance in the original data set with relative few coordinate directions, or characteristic structures in the original structure space.

3.METHODOLOGY

MATERIALS AND METHODS

Study Area

Kano State which corresponds to areal coverage of basement complex covers the area extending between Latitudes $10^{\circ} 3'N$ and $12^{\circ} 4'N$ and Longitude $7^{\circ} 4'E$ and $9^{\circ} 3'E$. It is bounded to the Northwest by Katsina State, to the South by Kaduna State and to the Northeast by Jigawa State, while to the Southeast by Bauchi State (Figure 1).

Kano experiences four distinct seasons, *Rani* (warm and dry), *Damina* (warm and wet), *Kaka* (cool and dry) and *Bazara* (hot and dry) closely associated with the movement of the Inter Tropical Discontinuity (ITD) zone. (Szentes, 1974; Bennett, *et al.*, 1978).

The bulk of the rainfall within a year is concentrated between May and September with August having the highest amount (Mohammed et al, 2015). The annual rainfall showed a mono modal pattern as confirmed by Iloeje, (1981).

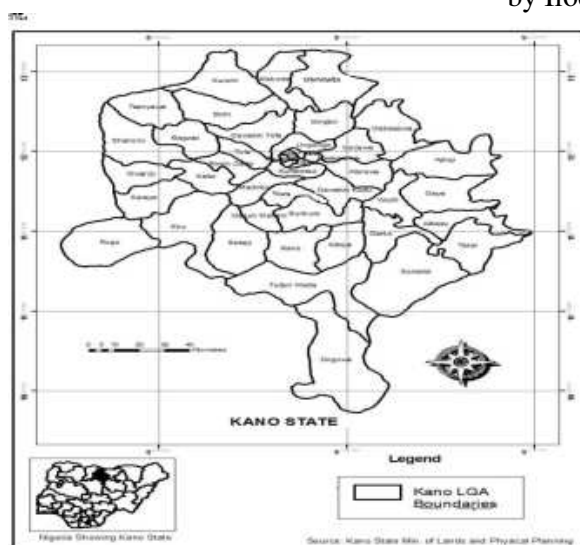


Figure 1: Map of Kano State with LGA Boundaries (Mohammed et al, 2015).

To analyze the data, time series an Expert Modeler of SPSS 16.0 software were used. The Expert Modeler only selects the candidate predictors to find the best model of those predictors that have a statistically significant relationship with the dependent series. It shows how predictors are useful (in terms of how each predictor is significant) and the model developed can be used for making forecast with the predictors.

An important preliminary step in any data analysis is to consider the possibility of a

(nonlinear) data transformation. It is often the case that the scale in which the data naturally arrives to the data analyst is not necessarily the best scale to analyze the data in. The primary goal of the transformation is to identify a scale where the residuals after fitting a model will have homogeneous variability and be independent of the level of the time series.

Another is a simplified data structure. The most common transformation used is the logarithmic transformation.

The modeler will give either exponential smoothing or ARIMA models. It shows

whether the model is additive or multiplicative and whether there is/are transformations.

4 RESULTS AND DISCUSSIONS

4.1 Time Series Modeler

Table 1: Model fit Statistics

Stationary R-squared	0.637272958
R-squared	0.631662181

Table 1 shows the values of R^2 and stationary R^2 for the whole model. The R^2 gives the estimates of the proportion in percentage of the total variation on the series that is explained by the model. It shows that 63.1% of the total variation has been

captured. The stationary R^2 is a measure that compares the stationary part of the model to a simple mean model in percentage. It shows that 63.7% of the stationary part of the model has been captured.

Table 2: Descriptions of the models for all the parameters

Parameters	Model Type	Stationary R-squared	R-squared	Sig.
Solar radiation	Simple Seasonal	0.63010	0.52583	1.63161E-02
Sunshine Hours	Simple Seasonal	0.70797	0.08168	6.84657E-01
Wind speed	Simple Seasonal	0.59951	0.63066	5.91505E-02
Maximum Temperature	Winters' Additive	0.72983	0.68946	6.60954E-03
Minimum Temperature	Simple Seasonal	0.75643	0.58153	9.18526E-01
Rainfall	ARIMA (0,0,3)(0,1,1)	0.43690	0.74621	5.74668E-01
Cloud Cover	Simple Seasonal	0.55744	0.88643	1.13431E-10
Relative Humidity	Winters' Additive	0.68000	0.91150	1.69787E-05

Table 2 shows the descriptions of the models of the parameters. The model types we have are Simple Seasonal, Winter Additive and ARIMA (0,0,3) (0,1,1)¹². The simple seasonal is the model that has level and season as the smoothing parameters, the Winter's additive has its smoothing parameters as level, trend, and season while ARIMA (0,0,3) (0,1,1)¹², has MA of order three, and the seasonal integration of order one and seasonal moving average of order

one. The value of the significance shows that all the models are highly significant.

Table 3 shows the estimates of the exponential smoothing parameters of the respective models. From the values in the significance, they show that, the levels for the whole parameters are highly significant. But, the values of the significance for the seasonal and trend parameters showed that is only the seasonal part of the cloud cover is highly significant.

Table 3: Exponential smoothing Model Parameters with no Transformation

Model parameters		Estimate	Sig.
Solar radiation	Alpha (Level)	0.199983941	1.06399E-09
	Delta (Season)	1.00893E-05	0.999586624
Sunshine Hours	Alpha (Level)	0.099842103	0.000190052
	Delta (Season)	4.96923E-05	0.998928041
Wind speed	Alpha (Level)	0.30002223	1.14027E-14
	Delta (Season)	3.421E-05	0.998870144
Maximum Temperature	Alpha (Level)	0.055304732	0.022627348
	Gamma (Trend)	1.07303E-08	0.999993814
	Delta (Season)	1.2158E-05	0.999573468
Minimum Temperature	Alpha (Level)	0.099664934	0.000325355
	Delta (Season)	8.69517E-06	0.999763606
Cloud Cover	Alpha (Level)	0.215882915	1.7501E-13
	Delta (Season)	0.420337665	4.93526E-18
Relative Humidity	Alpha (Level)	0.033282509	0.017925646
	Gamma (Trend)	0.000149338	0.793681886
	Delta (Season)	0.001000000	0.96782277

Table 4: ARIMA Model Parameters for Rainfall with No Transformation

		Estimate	Sig.
MA	Lag 1	-0.223434642	3.59745E-05
	Lag 2	-0.210005518	0.000116268
	Lag 3	0.11453552	0.03190101
Seasonal Difference	1	1	0
MA, Seasonal	Lag 1	0.755751871	3.92398E-58

Table 4 shows the estimates of the model parameters for rainfall. The values of the significance show that all the estimates are highly significant.

4.2 Empirical Orthogonal Transformations

Table 5: Kaiser-Meyer-Olkin Measure of Sampling Adequacy.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.601756673
	Sig.	1.6018E-169

From table 5 the KMO is used to determine the sampling adequacy of the data. From the table it can be seen that the value is 0.60, which is greater than 0.5. This shows that some information can be obtained. The value of the significance parameter shows that the data are also statistically significant.

Table 6: Communalities

Model parameters	Initial	Extraction
Solar radiation	1	0.98552
Sunshine Hours	1	0.99757
Wind speed	1	0.99859
Maximum Temperature	1	0.86435
Minimum Temperature	1	0.87180
Rainfall	1	0.90582
Cloud Cover	1	0.99778
Relative Humidity	1	0.88533

From Table 6, the communalities show the percentages of the information in the individual parameters that are being extracted and captured by the software for

the statistical analysis. From the column of the extracted information captured, it shows that most of the extracted information amount from 85% and upward.

Table 7: Component Matrix and the total variance explained with no rotation

Model parameter	Component					
	1	2	3	4	5	6
Relative Humidity	0.84738	-0.32841	0.21605	0.10282		
Rainfall	0.76991	-0.44455	0.12231			-0.30723
Minimum Temperature	0.73611	0.39067	0.25572	-0.12422	-0.28726	0.11822
Cloud Cover	0.58201		-0.58089		0.42339	0.35644
Maximum Temperature	0.32091	0.82414			-0.21101	0.16743
Solar radiation	0.14395	0.78235	-0.20904	-0.16175	0.30946	-0.43255
Wind speed	-0.10900	0.17988	0.85098		0.44721	0.15812
Sunshine Hours		0.33299		0.93549		
Initial Eigenvalues	2.33307	1.90015	1.23286	0.95125	0.61185	0.47759
% of Variance	29.16337	23.75186	15.41073	11.89060	7.64815	5.96984
Cumulative %	29.16337	52.91523	68.32596	80.21655	87.86470	93.83454
Number of Months	3.49960	2.85022	1.84929	1.42687	0.91778	0.71638
Descriptions	Damina	Bazara	Damina	Kaka	Kaka	Rani

Table 7 shows the component matrix and the total variance explained by each component, the number of months that each component covered, and the corresponding appropriate names. From the table, there are six components extracted. Component 1 is named Damina. This is because the coefficients of rainfall and minimum temperature are high. It last for 3.5 months. It may represent the months of June to September. Component 2 is named Bazara. This is because the coefficient of the maximum temperature is high and the coefficients of rainfall and Relative humidity are negative (which shows, the complete dryness of the weather). It usually lasts for 2.9 months. Component 3 is named also Damina. This is because of the presence of rainfall and minimum temperature. It last for

1.8 months. It may represent the low rainfall of May and June. Component 4 is named Kaka. and this is because the coefficient of the temperature is negative (indicating coolness) and no rainfall. It last for 1.4 months. Component 5 is also named Kaka. This is because the coefficient of the temperature is negative (indicating coolness) and no rainfall. This may also be because the coefficients of both maximum and minimum temperatures are negative. It last for almost twenty-eight days. Component 6 is called Rani. the coefficients of both maximum and minimum temperatures are low and there is no rain. It last for almost twenty-one days. From this result, it can be said that the rainy season covers five month (5) and three days, and the dry season five months eleven days (11)

Table 8: Component Matrix and the total variance explained with Varimax rotation

Parameters	Component					
	1	2	3	4	5	6
Rainfall	0.94127					0.11253
Relative Humidity	0.88472	0.20109	-0.16977			0.17261
Maximum Temperature	-0.16138	0.84406	0.32818	0.11970		
Minimum Temperature	0.42551	0.82483				
Solar radiation	-0.11095	0.27433	0.94262			
Sunshine Hours				0.99447		
Wind speed					0.98531	-0.15774
Cloud Cover	0.23144				-0.17612	0.95023
Initial Eigenvalues	1.94180	1.52260	1.03776	1.01225	1.01171	0.98065
% of Variance	24.27247	19.03253	12.97196	12.65309	12.64634	12.25815
Cumulative %	24.27247	43.30500	56.27696	68.93005	81.57639	93.83454
Number of Months	2.91270	2.28390	1.55664	1.51837	1.51756	1.47098
Descriptions	Damina	Bazara	Rani	Rani	Kaka	Damina

Table 8 shows the component matrix and the total variance explained by each component and the number of months that each component covers, and the corresponding appropriate names. Component 1 named Damina. This is because the correlation coefficient of the rainfall is very high and has moderate coefficient for the minimum temperature. It can be seen that it lasts for 3.2 months. It may represent the months of June to September. Component 2 is a Bazara season. This is because the coefficient of the maximum and minimum temperatures are high and the coefficient of rainfall is zero. (This shows that; the weather is completely dry). It usually lasts for 2.8 months. Component 3 is a Rani season. This is because the coefficient of minimum

temperature is zero, the coefficient of maximum temperature is moderate, and the coefficient of the rainfall is zero. It last for 1.6 months. Component 4 is a Rani season. This is because the coefficient of minimum temperature is zero, the coefficient of maximum temperature is low, and the coefficient of the rainfall is zero. It last for 1.5 months. Component 5 is a Kaka season. This is because the coefficients of the temperatures and rainfall are zero. It last for almost one and half months. Component 6 represents Damina. This is because of the presence of rainfall. It last for one and half months. This may represent the small rainfall experience in the months of May and September.

Table 9: Component Matrix and the total variance explained with Quartimax rotation

	Component					
	1	2	3	4	5	6
Rainfall	0.94348					
Relative Humidity	0.89899	0.17615			-0.15441	0.13721
Maximum Temperature	-0.15131	0.86652		0.11015	0.27369	
Minimum Temperature	0.44152	0.81903				
Wind speed			0.98561			-0.15415
Sunshine Hours				0.99368		
Solar radiation	-0.13018	0.33214			0.92118	
Cloud Cover	0.26595		-0.17991			0.93976
Initial Eigenvalues	2.00414	1.58288	1.01285	1.00819	0.95728	0.94143
% of Variance	25.05173	19.78596	12.66063	12.60233	11.96600	11.76789
Cumulative %	25.05173	44.83769	57.49832	70.10065	82.06666	93.83454
Number of Months	3.00621	2.37432	1.51928	1.51228	1.43592	1.41215
Descriptions	Damina	Bazara	Kaka	Rani	Rani	Kaka

Table 9 shows the component matrix and the total variance explained by each component and the number of months that each component covers, and the corresponding appropriate names. Component 1 named Damina season. This is because the correlation coefficient of the rainfall is very high and has moderate coefficient for the minimum temperature. It can be seen that it lasts for 3.0 months. It may represent the rain experience in the months of June to September. Component 2 is a Bazara season. This is because the coefficient of the maximum and minimum temperatures are high and the coefficient of rainfall is zero. (This shows that, the weather is completely dry). It usually lasts for 2.4 months.

Component 3 is a Kaka season. This is because the coefficients of the temperatures and rainfall are zero. It last for almost one and half months. Component 4 is a Rani season. This is because the coefficient of minimum temperature is zero, the coefficient of maximum temperature is very low, and the coefficient of the rainfall is zero. It last for 1.5 months. Component 5 is a Rani season. This is because the coefficient of minimum temperature is zero, the coefficient of maximum temperature is very low, and the coefficient of the rainfall is zero. It last for almost 1.4 months. Component 6 is a Kaka season. This is because the coefficients of the temperatures and rainfall are zero. It last for almost 1.4 months.

Table 10: Components Matrix and the total variance explained with Equamax rotation

	Component					
	1	2	3	4	5	6
Rainfall	0.93003			0.18198		
Relative Humidity	0.85126	0.24290	-0.19719	0.24216		
Minimum Temperature	0.39256	0.83323				
Maximum Temperature	-0.17956	0.80776	0.39618		0.13581	
Solar radiation		0.19946	0.96450			
Cloud Cover	0.16205			0.96563		-0.17579
Sunshine Hours					0.99562	
Wind speed				-0.15602		0.98600
Initial Eigenvalues	1.80817	1.45261	1.14765	1.06363	1.01972	1.01497
% of Variance	22.60214	18.15767	14.34563	13.29541	12.74653	12.68716
Cumulative %	22.60214	40.75981	55.10544	68.40085	81.14738	93.83454
Number of Months	2.71226	2.17892	1.72148	1.59545	1.52958	1.52246
Descriptions	Damina	Bazara	Rani	Damina	Rani	Kaka

Table 10 shows the component matrix and the total variance explained by each component and the number of months that each component covers, with their corresponding appropriate names. Component 1 is named rainy season (*Damina*). This is because rainfall and relative humidity have the highest correlation coefficients. It can last for 2.7 months. This may correspond to the months of June to August. Component 2 is a Bazara season. This is because the coefficient of the maximum and minimum temperatures are high and the coefficient of rainfall is zero. (This shows that the weather is completely dry). It usually lasts for 2.2 months. Component 3 is a Rani season. This is because, the coefficient of maximum temperature is very low. It last for 1.7 months. Component 4 is a Damina season. This is because of the presence of rainfall. It last for 1.6 months. This may represent the small rainfall experience in the months of

May and September. Component 5 is dry season. This is because the coefficient of minimum temperature is zero, the coefficient of maximum temperature is low, and the coefficient of the rainfall is zero. It last for almost 1.5 month. Component 6 is named Kaka. This is because the coefficients of the temperatures and rainfall are zero. It last for almost one and half months.

CONCLUSION

By combining the analysis of the results, it can be said that from the results of time series analysis the area has two prominent seasons they are rainy and dry seasons. The applications of empirical orthogonal functions show that, the seasons are further subdivided into four seasons as *Rani* (warm and dry), *Damina* (warm and wet), *Kaka* (cool and dry) and *Bazara* (hot and dry). The analysis of the rainy season shows that the rainy season is concentrated within five months (May to September).

REFERENCES

- Akgun, B. (2003). *Identification of Periodic Autoregressive Moving Average Models*. Middle East Technical University.
- Ayoade J. O. (2011) Climate change and the environment invited contribution to the Conference on Climate Change and Sustainable Development

- holding in University of Ilorin, Ilorin.;1-5
- Bell, W. R. (1984). *An Introduction to Forecasting with Time Series Models*. Insurance: Mathematics and Economics 3, pp. 241-255.
- Bennett, J.G., A.A. Hutcheon, J. Ibanga, W.J. Rackham, and J. Valette (1978)
- Box G.E.P., Jenkins G., (1970). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*, Pearson Education, Delhi
- Emery W. J, and Thomson R. E. (1997). *Data Analysis Methods in Physical Oceanography*. Pergamon: Oxford.
- Fukuoka A. 1951. *A Study of 10-day Forecast (A Synthetic Report)*, Vol. XXII. The Geophysical Magazine: Tokyo; 177–218.
- Hasmida, H. (2009). *Water Quality Trend at The Upper Part of Johor River in Relation to Rainfall and Runoff Pattern*. Universiti Teknologi Malaysia.
- Hendranata, A. (2003). *ARIMA (Autoregressive Moving Average)*. Manajemen Keuangan Sektor Publik FEUI.
- Ho, S. L. and Xie, M. (1998). *The Use of ARIMA Models for Reliability Forecasting and Analysis*. Computers ind. Engng, Vol. 35, Nos 1-2, pp. 213-216.
- Iloeje, N. P. (1981). *A New Geography of Nigeria*, UK. Longman.
- Lorenz EN. 1956. Empirical Orthogonal Functions and Statistical Weather Prediction. Technical report, Statistical Forecast Project Report 1, Dep of Meteor, MIT: 49
- Lorenz EN. 1970. Climate change as a mathematical problem. *Journal of Applied Meteorology* 9: 325–329.
- Lee, C. and Ko, C. (2011). *Short-term Load Forecasting Using Lifting Scheme and ARIMA Models*. Expert Systems with Applications, Vol. 38, pp. 5902-5911.
- Mohammed, M. U., Abdulhamid, A., Badamasi, M. M. and Ahmed, M. (2015). Rainfall Dynamics and Climate Change in Kano, Nigeria. *Journal of Scientific Research & Reports*, vol. 7, No. 5, pp. 386-395.
- Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons, Inc.
- Preisendorfer RW. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Elsevier: Amsterdam.
- Szentes, G. (1974) – *Hydro-geological Data of NorthEast Kano State*. Water Resources and Engineering Construction Agency (WRECA), Kano State.
- Von Storch H. and Zwiers F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press: Cambridge.