Statistical Modelling of Intensity Modes of Rainfall Using Principal Component Itemization: A Case Study of Kano State

Ibrahim Abubakar Sadiq[1,2*] and Nazir Muhammad Isah[1]

[1]*Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.*
[2]*Department of Mathematics and Statistics, Mewar University, Chittorgarh, Rajasthern, India.*

Authors' contributions

This work was carried out in collaboration between both authors. Author IAS designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors NMI and IAS managed the analyses of the study. Author NMI managed the literature searches. Both authors read and approved the final manuscript.

Review Article

Abstract

Kano State is experiencing greater weather extremes, changes in rainfall patterns, analysis of heat and cold waves, and increasing droughts and floods (Kano meteorological agency). As a result, there is a need for the provision of the necessary weather advisories and early warnings to planners, decision-makers, and operators of the various rainfall-sensitive socio-economic sectors. However, this study is aiming to realize some hidden variables of Kano State total monthly rainfall dataset from the onset to cessation period of rain from the month of April to October over a 105 years (1911-2015) for classification into the intensity of the rain of the area under study, also to determine the linear model for the changing patterns of rainfall in Kano State and to identify some of the adverse impacts on socio-economic sectors and transport infrastructures. Thus, the appearances of the rainfall figure are established for the study region with the operation of Principal Component Analysis (PCA), application least square method. The leading three (3) PCs, gives account for about 61% of the entire disparity, is described. The revision displays and describe PC1 as associated with the heavy intensity rainfall, whereas PC2 is connected to the moderate-intensity rainfall and finally PC_3 is linked to the light intensity rainfall of the region under study. By the scores of our PCs, uniform rainfall zones are established over the region of enquiring to which the yearly performance of rainfall is discussed. Statistically, all three models for the various mode of rainfall intensity are significant, which serves as the annual pattern of rainfall in the study area.

_____

Corresponding author: E-mail: abubakarsadiq463@gmail.com;

1 Introduction

In Kano State, limited analyses have been conducted on how changes in rainfall would affect the infrastructures necessary to fill current mobility needs. The targets of this study are to determine the linear model for the changing patterns of rainfall in Kano state using long-term data, identifying some of the adverse impacts on socio-economic sectors, transport infrastructures [1].

Extreme and unusual weather events, resulting in loss of life and property, and disruption of socio-economic activities, are being experienced all over the world. The increasing frequency and intensity of these events constitute a major challenge to socio-economic development, particularly in developing countries. Nigeria is not immune to this global phenomenon. Timely weather and climate information are therefore vital tools for planning in vibrant parts of the economy which are subtle to meteorological conditions [2].

The destructive impacts of great and uncommon weather on key divisions of the economy are numerous and varied. In the aviation sector, severe weather conditions lead to flight delays and cancellations, resulting in disruption of business activities. In Nigeria, the rainy season is characterized by thunderstorms, strong winds, line squalls, and turbulence, especially at the onset and cessation phases of the season. Agriculture in Nigeria is predominantly rain-fed. Consequently, farmers suffer tremendous crop failure whenever there are significant changes in the rainfall pattern. Quite often, their traditional methods of determining when to commence farming fail on account of uncertainties brought about by climate change. Dam managers face the challenge of ensuring optimum performance of hydrological stations as a result of water shortage or surplus following drought or above normal rainfall, increasing the risk of flooding, soil erosion, drought, etc. In various parts of the country are disaster risks which disaster managers have to deal with. The outbreak and spread of some diseases such as cerebrospinal meningitis, malaria, and respiratory tract infections are affected by weather conditions [3].

Human activity, particularly deforestation and the burning of fossil fuels, is driving climate change by increasing atmospheric concentrations of carbon dioxide and other greenhouse gases (GHGs). As a result, Kano State is experiencing greater weather extremes, changes in rainfall patterns, heat and cold waves, and increasing droughts and floods (Kano meteorological agency). These phenomena harm the environment and people's lives and livelihoods. Marginalized groups in the poorest regions are particularly affected, even as they are least responsible for these fluctuations (UNDP, 2019). It is a provision of the necessary weather advisories and early warnings to planners, decision-makers and operators of the various rainfall-sensitive socio-economic sectors [4,5]. The main plan of this study is to enhance our understanding of the dynamics of the underlying system (Principal Component Analysis), by interpreting the components as representing physically significant 'modes' of rainfall variability and therefore, to determine the linear model for the changing patterns of rainfall in Kano state using long-term data. The study serves as a provision of the necessary weather advisories and early warnings to planners, decision-makers, and operators of the various rainfall-sensitive socio-economic sectors of Kano State and Nigeria at large such as the aviation sector, Agriculture sector, hydrological stations, disaster risks centers, and health sectors [6].

2 Materials and Methods

The PCA discovers a set of sizes and dimensions in a smaller group of the position defined by the set of variables. Such coordinates are characterized as battle-axes points. They are at right angles (orthogonal) to one another. Let's say, supposing you study triplet's variables which are designated in trilateral dimensional spaces. The respective variable represents the individual axis. At this moment assume that the data fall close to the dual dimensional plane inside the trilateral dimensions. The PCA of the data has to expose two factors that can give an interpretation for the dual dimensions. An individual may spin the axes of this side by side level plane through controlling the 90° point of view angle among them, objectively as the knife-edges of a

jet propeller spin yet preserve equal angles between themselves. Thus, the expectation is that revolving the axes could advance individual skill to understand the importance and deduction of every component.

Several different forms of whirling have been proposed. The greatest of them remained technologically advanced for use in factor analysis. The varimax and quartimax where provided by NCSS as two options for orthogonal rotation [7,8].

The Varimax gyration is the greatest widespread orthogonal gyration method. In this method, the battle-axes are revolved to make best use of the entirety of the variances of the squared loadings inside every pilaster of the loadings matrix. Exhausting the possibilities giving to this assumption militaries the loadings to be one of vast or slight. The expectation is that by spinning the factors, one will find new factors that are respectively highly associated with simply rare of the original variables. It makes simpler the understanding of the factor to concern about these dual or three variables. An alternative way of declaring the objective of varimax gyration is that it groups the variables into clusters, where every cluster is essentially a novel factor. Meanwhile varimax pursues to maximize a particular principle, it yields a unique way out. This has supplementary to its validation. Let denote the matrix $B = \{b_{ij}\}$ be the revolved factors. However, the aim varimax gyration aims to take full advantage of the measure:

$$Q_1 = \sum_{J=1}^{K} \left(p \sum_{i=1}^{p} b_{ij}^4 - \sum_{i=1}^{p} b_{ij}^2 \Big/ p \right)$$

(1)

Equation (1) offers the raw varimax gyration. This gyration has the weakness of not dispersal of the variance precisely consistently between the novel factors. As an alternative, it has a habit of forming one big factor shadowed by various lesser ones. In advancing this process we employs normalized varimax gyration and the maximized quantity is given below.

$$Q_N = \sum_{j=1}^{k} \left(p \sum_{i=1}^{p} \left(\frac{b_{ij}}{h_i} \right)^4 - \sum_{i=1}^{p} \left(\frac{b_{ij}}{h_i} \right)^2 \Big/ p^2 \right)$$

(2)

Given h_i represent the square root of communality of ith variable. The Quartimax gyration is comparable to varimax gyration, for only that the rackets of matrix *B* are maximized comparatively than pilasters of matrix *B*. The mentioned gyration is more probable to yield a universal factor than what the varimax produced willingly. Regularly, the outcomes are to a certain extent comparable. Below are measures to which inflate for quartimax.

$$Q_N = \sum_{j=1}^{k} \left(\sum_{i=1}^{p} \left(\frac{b_{ij}}{h_i} \right)^4 \Big/ p \right)$$

(3)

Equation (3) is called Bartlett's sphericity test employing for challenging the null hypothesis (H_0) of which the matrix of correlation matrix is a matrix of identity. By execution of the process, if the probability value is larger than the 0.05 alpha level, an individual must abstain from performing a PCA on such facts. This assessment is legally effective for a large samples size viz. begin from $N > 100$. This test apply statistical Chi-square distribution with $p[p-1]/2$ degrees of freedom. Recall that this analysis is only obtainable when one examines the matrix of correlation.

$$\chi^2 = \frac{(11+2p-6N)}{6} Log_e |R|$$

(4)

Rainfall data for Kano State from the period 1911 to 2015 (105 years) were used in this study. These are secondary data that were sourced from the archive of the Nigerian Meteorological Agency (NIMET), Kano office division. We emphasize the approaches of applying the PCA on the available data by using relevant software. From this data, scree plots were generated using the SPSS software.

3 Results and Discussion

Table 1. Descriptive statistics

Months (variables)	Mean	Std. deviation	Analysis N
April	12.8210	22.40302	105
May	60.5686	46.82682	105
June	127.7771	69.21149	105
July	229.1581	107.53372	105
August	311.6590	118.60680	105
September	146.3381	82.17665	105
October	14.0124	19.30465	105

Interpretation:

Table 1 presents the descriptive measure of variables under scrutiny which give insight for PCA. This measures include Standard Deviations, Mean and total number of observation.

Table 2. Correlation matrix[a]

Months (variables)		April	May	June	July	August	September	October
Correlation	April	1.000	.073	.069	.013	-.038	-.046	.197
	May	.073	1.000	.214	.218	-.046	-.016	-.101
	June	.069	.214	1.000	.311	.181	.026	.050
	July	.013	.218	.311	1.000	.186	.364	.051
	August	-.038	-.046	.181	.186	1.000	.360	.062
	September	-.046	-.016	.026	.364	.360	1.000	.216
	October	.197	-.101	.050	.051	.062	.216	1.000

a. Determinant = .527

Interpretation:

Table 2 is the Correlation Matrix. It is a table that provides the correlations among the actual scrutiny variables. Before running a PCA, there is a need to inspect the correlations among the variables under investigation. This relationship between our variables is computed and presented in Table 3 with the help of chi-square and Bartlett's test. There are indications of a strong correlation as the coefficient value is .731. By observing the significant probability value is sufficiently smaller compared to our 0.05 alpha level of confidence. Through the mentioned facts we deduced that a significant relationship exists among the variables.

Table 3. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy. [a]		.731
Bartlett's Test of Sphericity [b]	Approx. Chi-Square	64.654
	Df	21
	Sig.	.000

Kaiser-Meyer-Olkin Measure of Sampling Adequacy- This measure varies between 0 and 1, and values closer to 1 are better. A value greater than .7 is suggested. In this analysis, the value is 0.731 which implies that the sample is adequate.

Bartlett's Test of Sphericity- This tests the null hypothesis that the correlation matrix is an identity matrix. An identity matrix is a matrix in which all of the diagonal elements are 1 and all off-diagonal elements are 0. The p-value in the test (<0.001) indicates that the null hypothesis is to be rejected. Taken together, these tests provide a minimum standard that should be passed before a principal components analysis should be conducted [7,9].

Table 4. Communalities [a]

Months (variables)	Initial [b]	Extraction [c]
April	1.000	.698
May	1.000	.608
June	1.000	.513
July	1.000	.603
August	1.000	.517
September	1.000	.675
October	1.000	.679

Extraction Method: Principal Component Analysis

Interpretation:

Table 4 is Communalities extraction for the PCA. In this table, we have seen the proportion of each variable's variance which accounts for explaining by the PCs. These extractions are termed as the sum of squares of component loadings and are represented symbolically as h^2. By explanation, this initial communality of PCA results to be unity meanwhile underlying the capability of our scrutiny variable possibly certain to offers a significant interpretation for the undergoing analysis. The values in column c, variables with greater values may be well characterized in the shared constituent space, whereas variables with smaller values could not be well characterized.

Table 5. Total variance explained

Component [a]	Initial eigenvalues [b]			Extraction sums of squared loadings [f]			Rotation sums of squared loadings		
	Total [c]	% of variance [d]	Cumulative % [e]	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	1.810	25.862	25.862	1.810	25.862	25.862	1.622	23.177	23.177
2	1.303	18.612	44.474	1.303	18.612	44.474	1.460	20.856	44.033
3	1.180	16.853	61.327	1.180	16.853	61.327	1.211	17.294	61.327
4	.839	11.985	73.312						
5	.756	10.799	84.111						
6	.660	9.429	93.540						
7	.452	6.460	100.000						

Extraction Method: Principal Component Analysis

Interpretation:

Table 5 is the PCA result refers to the total variance explained. It indicates exactly how the quantity of extracted components in the course of our analysis. The involvement of many components is an outcome of the input variable in the system. The pilaster of initial eigenvalues is the respective variances for the PCs. As seen and achieved earlier to our employment of correlation matrix techniques, the sum of the total variance is also equivalent to the total number of variables in the analysis. In the total pilaster is the eigenvalues. By

this it was observed that through the first to the last components all justification for the most variance go downward descending meaning lager with lager contribution, each successive component will justification for a less and lesser variance [10,8].

In general, looking at the cumulative percentage pilaster, the third rackets described 61.327%. By considering this statistical evidence altogether the first three components explain almost about 61.327% of the total variance.
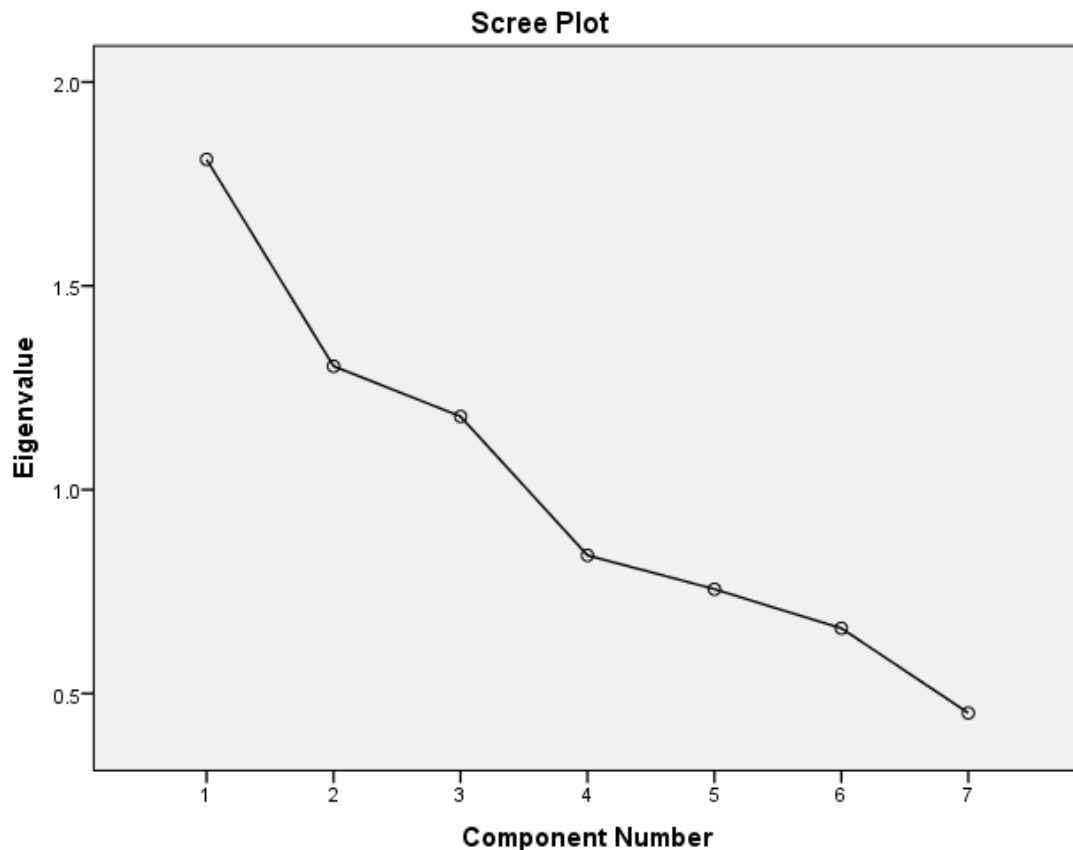

Fig. 1. Scree plot

Interpretation:

Fig. 1 is called the scree plot. It displays the eigenvalue alongside the component number. One can see clearly that our values in the first leading two pilasters of Table 5 are the fourth component in our scree plot is nearly very plane. A consequence of that, we infer at each succeeding component, is a reckoning for lesser and slighter quantities of the total variance explained. Comprehensively, the purpose remained in retaining only such individuals' PCs of which eigenvalues are larger than unity (1). The remaining constituents by eigenvalue of lesser than unity (1) is justification for a smaller amount of variance than prepared by the novel variable and consequently are of little practice. Therefore, individuals can realize that the fact behind the PCA is to reallocate the variance in the matrix of correlation to reallocate the variance to leading extracted components.

Table 6. Component matrix[a][b]

Months (variables)	Component [c]		
	1	**2**	**3**
April	.086	.122	.822
May	.264	.733	.027
June	.533	.470	.085
July	.738	.219	-.102
August	.595	-.320	-.246
September	.681	-.440	-.132
October	.294	-.430	.639

Extraction Method: Principal Component Analysis

Table 7. Rotated component matrix[a]

Months (variables)	Component		
	1	**2**	**3**
April	-.224	.194	**.781**
May	-.179	**.756**	-.066
June	.160	**.691**	.096
July	.509	**.587**	.008
August	**.715**	.053	-.051
September	**.815**	.007	.100
October	.283	-.158	**.757**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization
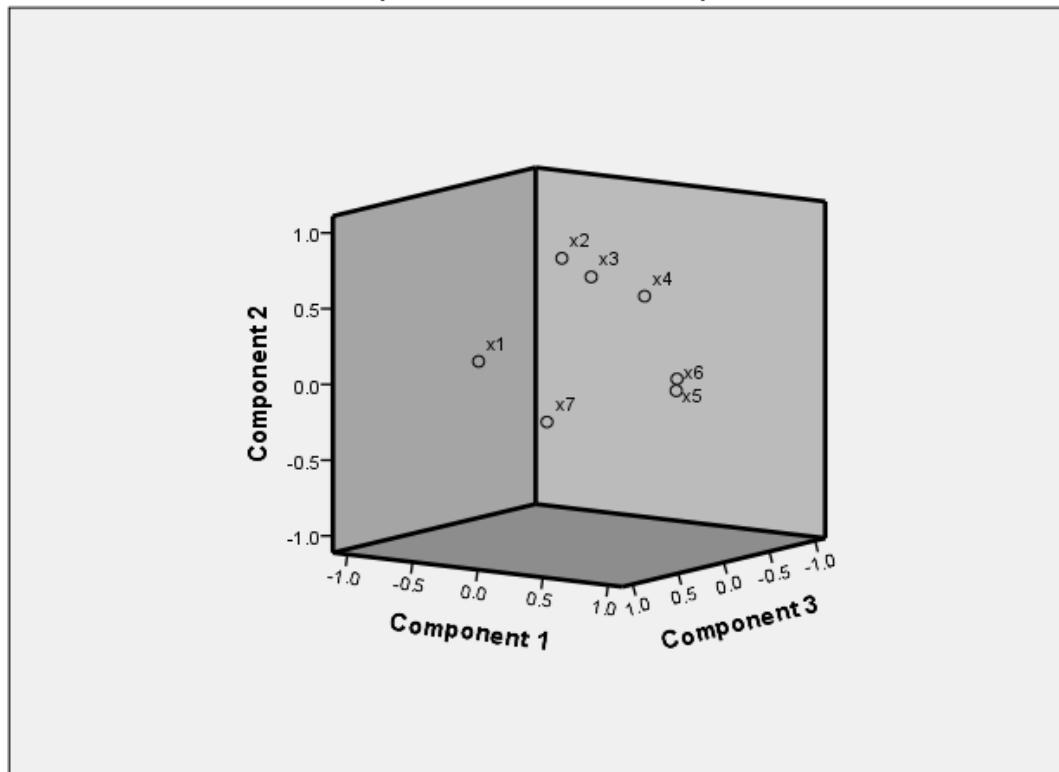
Fig. 2. Component plot in space

Fig. 2 shows the variables clustering together from where they would be categorized into new components.
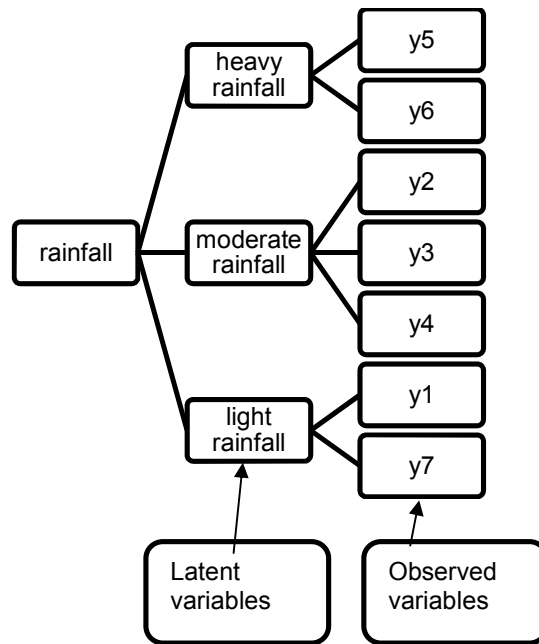
Regrouping the variables

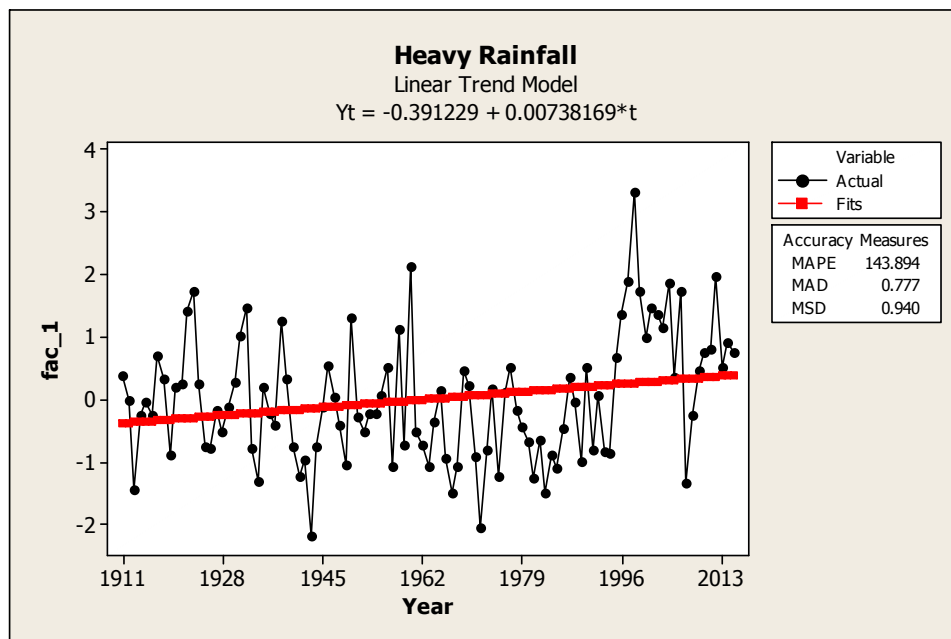**Fig. 3. Regrouping the variables into their components**

Fig. 4. Linear trend plot for heavy rainfall

Interpretation:

The expected heavy rainfall intensity according to the model in Fig. 4 is given by the trend line $y_t = -0.391 + 0.007*t$. The heavy rainfall intensity above or below the trend line is caused by natural random variations.
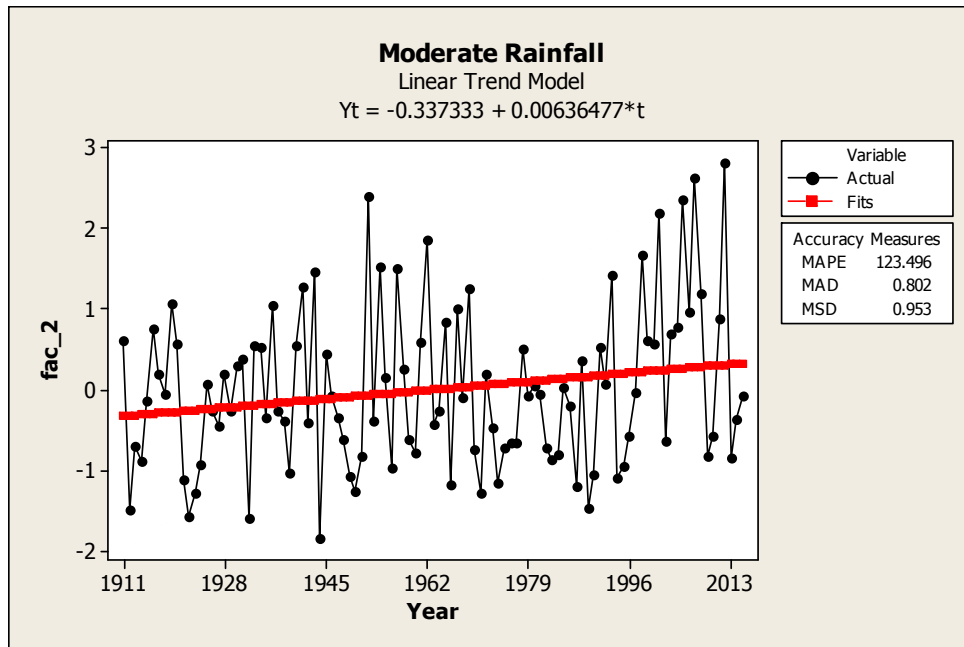
Fig. 5. Linear trend plot for moderate rainfall

Interpretation:

The expected moderate rainfall intensity according to the model in Fig. 5 is given by the trend line $y_t = -0.337 + 0.006*t$. The moderate rainfall intensity above or below the trend line is caused by natural random variations.
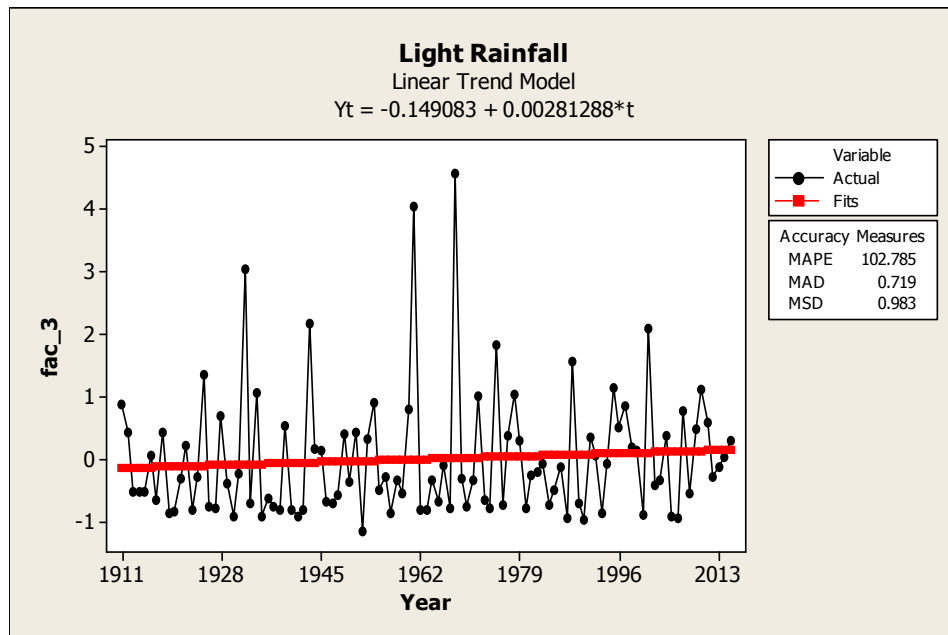

Fig. 6. Linear trend plot for light rainfall

Interpretation:

The expected light rainfall intensity according to the model in Fig. 6 is given by the trend line $y_t = -0.149 + 0.003*t$. The light rainfall intensity above or below the trend line is caused by natural random variations.

PCA findings:

The three (3) leading PCs, gives account for about 61% of the entire disparity, is described. The revision displays and described PC1 is associated to the heavy intensity rainfall, whereas PC2 is connected to the moderate-intensity rainfall and finally PC_3 is linked to the light intensity rainfall of the region under study. By the scores of our PCs, uniform rainfall zones are established over the region of enquiring to which the yearly performance of rainfall is discussed. These means that PC1 explains the month of August and September, which usually experience heavy annual amounts of rainfall, while PC2 explains that the month of May, June, and July usually experience moderate annual amounts of rainfall and PC_3 explains that the month of April and October usually experience light annual amounts of rainfall Statistically all the three models for the various mode of rainfall intensity are significant, which serves as the annual pattern of rainfall in the study area.

4 Conclusion

This study has examined the rainfall patterns of Kano State using total monthly data from the period 1911 to 2015. The impacts resulting from these changes in socio-economic sectors, transport infrastructures were identified. The increase in the annual rainfall yield in recent periods is predominantly as a result of the increase in the June, July, August, and September rainfall as indicated by the statistically significant wetter conditions of those months. Observations show that flood occurrences which are an indication of climate change corresponded with months of increase in rainfall amount in the area. This study has also highlighted the vulnerable nature of socio-economic sectors, agriculture, transport infrastructures to extreme weather events especially increasing in amount and intensity of rainfall in the area.

Competing Interests

Authors have declared that no competing interests exist.

References

[1]   Abaje IB, Ati OF, Iguisi EO. Changing climatic scenarios and strategies for drought adaptation and mitigation in the Sudano-Sahelian Ecological Zone of Nigeria. In M. A. Iliya & I. M. Dankani (Eds), Climate Change and Sustainable Development in Nigeria. Ibadan: Crown F. Publishers. 2012a;99-121.

[2]   Abaje IB, Ati OF, Iguisi EO. Recent trends and fluctuations of annual rainfall in the Sudano-Sahelian Ecological Zone of Nigeria: Risks and opportunities. Journal of Sustainable Society. 2012b;1(2):44-51.

[3]   Abaje IB, Ati OF, Iguisi EO, Jidauna GG. Droughts in the Sudano-Sahelian Ecological Zone of Nigeria: Implications for agriculture and water resources development. Global Journal of Human Social Science (B): Geography, Geo-Sciences & Environmental. 2013;13(2):1-10.

[4]   Wang J, Yu Y, Gong Q, Yuan S. Debris flow disaster risk analysis and modeling via numerical simulation and land use assessment. Arabian Journal of Geosciences. 2020;13(19):1-14.

[5]   Das S. Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin, India. Remote Sensing Applications: Society and Environment. 2019;14:60-74.

[6]   African Institute for Applied Economics. Kano State. BECANS Business Environmental Report. 2007;1(21).
Available:http://www.aiaenigeria.org/becans/Reports/Kano.pdf

[7]   Ati OF, Muhammed SQ, Ati MH. Variations and trends in annual rainfall amounts and the onset of the rainy season for Kano for 87 Years (1916-2002). Journal of Applied Sciences Research. 2008;4(12):1959-1962.

[8]   Rencher AC. Multivariate statistical inference and applications. Jon Wiley and Sons, INC. 1998;559.

[9]   Richman MB. Rotation of principal components. Journal of Climatology. 1986;6:293-335. Stevens JP. Applied Multivariate Statistics for the Social Sciences (4th Ed.). Mahwaw, NJ: Lawrence Erlbaum Associates; 2002.

[10]   Karamizadeh S, Abdullah SM, Manaf AA, Zamani M, Hooman A. An overview of principal component analysis. Journal of Signal and Information Processing. 2013;4:173-175.