

STATISTICAL MODELING OF THICK DUST HAYS HAZARDS IN SOKOTO METROPOLIS USING ARIMA MODEL

BY

S. U. Ibrahim, M. O. Oladejo and U. L. Okafor

Department of Mathematics, Faculty of Science, Nigerian Defence Academy, Kaduna, Nigeria.

Corresponding Author; shehusakkwato@gmail.com;

Abstract

ARIMA model have been used in this research to model hazards values of a climatic factor such as thick dust hays in Sokoto metropolis under Box and Jenkins (1976) methodology. The climatic factor data series were obtained from Nigeria meteorological unit (NIMET) Sokoto as an annual average secondary data series for period of fifteen years (2002-2016) which were assumed to be stationary both in mean and variance over time. Hypotheses were tested in this research against time plots to test for certain statistical facts findings, i.e. test for auto correlations of errors and test for stationary process. The results of the tested hypotheses favored the climatic factor data series assumptions that the series were stationary over time. In the model parameters estimation of the climatic factor satisfied both stationarity and invertibility conditions i.e. at least one of the autoregressive parameter estimate should be negative and at least one of the moving average parameter estimate should be positive respectively, and the estimated ARIMA model parameters are: AR (1); 1.752, AR (2); -0.992, MA (1); 1.708, MA (2); -0.689 and MA (3); -0.053. The ARIMA model estimated parameters yielded ARIMA model order as ARIMA (2, 0, 3). The order of the best fitted ARIMA model was found to be adequate at all lags and auto correlations of errors were not significant.

Keywords: Auto regressive, moving average and White noise process.

1. INTRODUCTION

Background of this study

Sokoto metropolis has been suffering some problems which occurred as a result of nature of a climatic factor in the region. Such climatic factor is: thick dust hays.

Climatic factor such as thick dust hays has serious effects on health sector, air flight operations, transportations, life and properties in the region of Sokoto metropolis.

Thick dust hays

Mineral dust has significant impacts in the region of Sokoto metropolis. Airborne mineral dust can have numerous repercussions on human health, such as allergies, respiratory diseases and eye infections. Airborne mineral dust reduces visibility. Wang et al (2008) confirmed that, despite the problems in road and air transportation, dust storms also have negative impacts on agriculture causing loss of crop and livestock. Stefanski and Sivakumar (2009) stated that, desert dust deposition also influences the biogeochemical cycles of both oceanic and terrestrial ecosystems.

Okin et al. (2004), Jickells et al. (2005), Mahowald et al. (2005), and Schulz et al. (2012) illustrated that, dust releases iron into river water. Nickovic et al. (2013) stated that, indeed, due to the many connections with the Earth's systems, mineral dust can also impact the carbon cycle and atmospheric CO₂. Jickells et al. (2005), Hamza et al. (2011) agreed that, Mineral dust also has a significant impact on Earth radioactive budget through both direct and indirect effects. The fact that Sokoto state share border with Niger republic where desert travel from in to Sokoto metropolis, all the above mentioned problems affected Sokoto metropolis.

Data for this research

This research obtained data series of such climatic factor from the available record of meteorological unit Sokoto for period of fifteen (15) years (2002 - 2016) of the climatic factor as an annual mean average secondary data for estimation and modeling future values of such climatic factor mentioned above.

Generalized forms of ARIMA model

Box and Jenkins (1976) introduced ARIMA model as a time series process model which take values at equally spaced intervals. Box and Jenkins defined time series process as an ordered sequence of observations at equal space intervals. Box-Jenkins described ARIMA model as a model that have many forms and represent different stochastic processes. And that the mathematical representation of such ARIMA process, initiate with autoregressive process, denoted by AR process, in which AR (p) process, is mathematically represented as:-

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \tag{1}$$

$$e_t \sim N(0, \sigma^2)$$

Box, Jenkins and Riesel (1994) and Box & Jenkins (1968) confirmed that, $e_t \sim N(0, \sigma^2)$ as a white noise process, which were assumed to be normally, identically and independently distributed random variables (the sequence of uncorrelated errors with constant zero mean and σ^2).

Another ARIMA model component is, MA (q) process, which is represented as:-

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \tag{2}$$

$$e_t \sim N(0, \sigma^2)$$

Then, equation (1) and (2) gives an ARMA (p, 0, q) process, which is express as:-

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \tag{3}$$

$$e_t \sim N(0, \sigma^2)$$

ϕ : Is an autoregressive process parameter, θ : is a moving average process parameter.

p : Is a general order of autoregressive component, and q : is general order of moving average component.

If d component is integrated in the ARMA (p, q) process, then such process becomes ARIMA (p, d, q), p, d, q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, moving average parts of the models respectively and ‘d’ controls the level of differencing the ARMA model (p, q) process. ARIMA (p, d, q) is also represented as standard equation in terms of lag polynomials as:-

$$\phi(L)(1-L)^d y_t = \theta(L)e_t \tag{4}$$

$$e_t \sim N(0, \sigma^2)$$

Box, Jenkins and Riesel stated that time series data may exhibit a seasonal variation in a data series, therefore, the following formula represents seasonal variation in a data series as:-

$$\phi_p(L^S)\phi(L)(1-L)^d(1-L^S)^D y_t = \theta_q(L^S)\theta(L)e_t \tag{5}$$

$$e_t \sim N(0, \sigma^2)$$

$\phi_p(L^S)$: Is seasonal autoregressive parameter, $\phi(L)$: is a non-seasonal autoregressive parameter, $(1-L)^d$: is non-seasonal difference, $(1-L^S)^D$: is a seasonal difference, $\theta_q(L^S)$: is

a seasonal moving average parameter, $\theta_q(L)$: is a non-seasonal moving average parameter and e_t : is an error term distributed throughout a data series.

Statements of the problem

Sokoto metropolis has been facing so many challenges on natural disaster from the excess values of the climatic factor in Sokoto metropolis where the following problems were itemized as:-

1. Agricultural problem of poor productions where the dust particles covers crops at their early stage of shooting out from soil after planted.
2. Problem of poor visibility condition as a result of thick dust hays which causes accidents and concealment of air flight operations.
3. Spread of dust particles from the thick dust hays in eye and sneezed in nose which causes epidemic outbreak.

Aim:-

The aim of this research is to estimate and model future values thick dust hays using ARIMA model and determine best model by considering their best estimated parameters under some statistical facts.

Objectives:-

The objectives of this research are as follows:-

1. To provides accurate weather estimates in respect of the climatic factor as; thick dust hays using ARIMA model in Sokoto metropolis.
2. To use the estimates of such climatic factor and model visibility condition in Sokoto metropolis.
3. To determine the best fitted model for Sokoto weather condition out of the estimated model parameters.

Scope

This research only covers a climatic factor such as thick dust hays data series from Meteorological unit of Federal Air port Sokoto, Sokoto state. This research will only be concerned with estimation and modeling of such climatic condition using ARIMA model in Sokoto metropolis.

Limitation

This research is restricted to only one climatic factor data series such as, thick dust hays that its excess values caused problems in Sokoto metropolis.

2. LITERATURE REVIEW

Introduction

This chapter reviewed some related literature that contributed immensely on weather modeling and forecast, looked into particular models employed in analysis of weather modeling.

Authors and their contributions

Yakubu (2014) carried out research on some climatic elements and employed ARIMA model to estimate monthly average temperature using a data series on temperature from Nigeria meteorological unit (NIMET) Sokoto and obtained seasonal ARIMA model for prediction of monthly average temperature in Sokoto metropolis. The author indicated that, the results of seasonal modeling of Sokoto monthly average temperature have been obtained using seasonal

autoregressive integrated moving average modeling approach. Based on this seasonal modeling analysis, the author concluded that , the best seasonal model among the models that are adequate to describe the seasonal dynamics for Sokoto city temperature is SARIMA (3,0,1)(4,1,0) 12, SARIMA (1,0,0)(0,1,1) 12 and SARIMA (4,0,2)(5,1,1) 12 models.

Usman et al. (2013) contributed by conducting research on visibility condition in Sokoto metropolis by presenting some results of measurement and analysis of thick dust hays and three linear regression models for estimating the monthly mean visibility in Sokoto state. The authors stated that, fraction of sunshine hours; relative humidity and maximum temperature covering a period of 5 years (2002-2006) were obtained from the Meteorological Department Sultan Abubakar (III) International Airport Sokoto.

Yusuf (2002) presented a qualitative research paper which mainly highlighted the hazards of some climatic elements and the problem they posed on effective educational planning and management by sensitization in Sokoto metropolis.

Therefore, this research used Box and Jenkins methodology (1976), estimated and modeled thick dust hays data series in Sokoto metropolis for period under review (2002-2016).

3. METHODOLOGY

Most time series applications are not stationary. And to estimate a time series model parameters, this research imposed the condition that they are covariance stationary. This means that this research incorporated the concept of non-stationarity in time series which leads to ARIMA processes. ARIMA denotes Autoregressive integrated moving average. Or, an ARIMA process is an integrated ARMA process. Here is a class of non-stationary processes which becomes stationary after a finite number of differencing. For such process d denotes that number of times the process needs to be differenced to become stationary and d is called the order of integration of the processes. When such derived stationary processes can be estimated as ARMA, then we call the original process ARIMA because it is a non stationary process we need to differentiate it d times to become stationary, which if it is stationary already then it can be estimated as ARMA (p, q), otherwise ARIMA (p, d, q) process.

This research, adopted Box-Jenkins methodology for the ARIMA model parameters estimation and modeling excess values of the climatic factor. The methodology involved the following: (1) Identification and time plots including statistical tests (hypothesis testing). (2) Model parameters estimation (through packages). (3) Diagnostic checking and model selection based on certain statistical reasons. Packages such as: GRETEL, J-MULTI and NCSS 11.0 for obtaining suitable model for this research were used.

ARIMA model identification

The model identification stage enables this research to select a subclass of the family of ARIMA models appropriate to represent a time series. This involves stationary transformation, regular differencing, seasonal differencing, unit root and stationarity tests (hypothesis testing).

Statistical tests i.e. hypothesis testing for the variable forecasting model. KPSS Test

This test (KPSS) has been proposed by Kwiatkowski *et al.* (1992) where the hypothesis that the Data generating process (DGP) is stationary is tested against a unit roots, if there is no linear trend term (i.e. trend stationary). The Data generating process is given by:

$$H_0 : y_t \sim I(0)$$

$$H_1 : y_t \sim I(1)$$

That is the null hypothesis that the data generating process (DGP) is stationary and alternative hypothesis otherwise is tested against a unit root. Kwiatkowski, Philips, Schmidt

and Shin have derived a test for this pair of hypothesis. If there is no linear trend term, they start from a DGP.

$y_t = x_t + z_t$ where x_t is a random walk i.e. $x_t = x_{t-1} + e_t, e_t \sim N(0, \sigma^2)$ and Z_t is stationary process.

Test statistic is:

$$KPSS = \frac{1}{T^2} \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}_\varepsilon^2} \tag{6}$$

Decision rule: Reject null hypothesis if the test statistic is greater than the asymptotic critical values.

Portmanteau-test

The test was introduced by L. Jung Box statistics for correlation (1978) to test the presence of auto correlations, where Number of lags as u and h are predetermined.

The hypothesis stated that:

$$H_0 : P_{u,i} = \dots = P_{u,h} = 0$$

$H_1 : P_{u,i} \neq 0, i.e.$ For at least one $i = 1, \dots, h$ is tested i.e. at least one lag with non zero correlations.

The test statistic is given by:

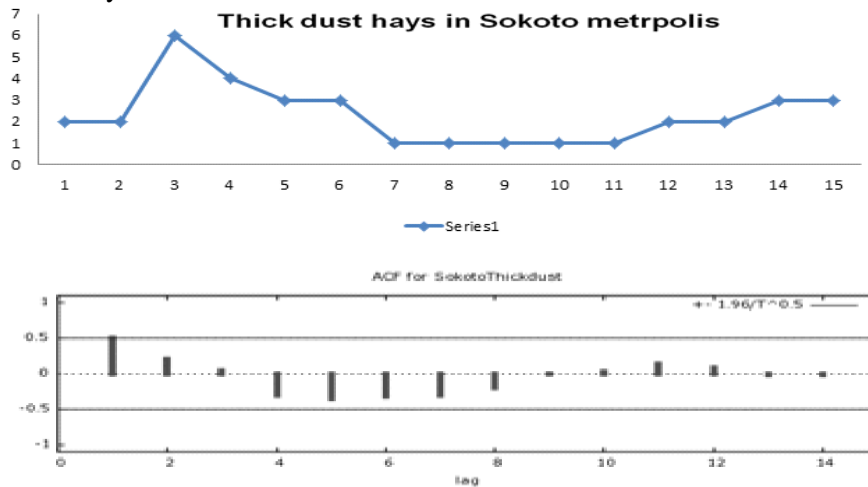
$$LB_h = T^2 \sum_{j=1}^h \left(\frac{1}{T-j} \right) \hat{\rho}_{u,j}^2 \tag{7}$$

Decision rule: reject H_0 if P-value is less than the significant level α

4. DATA ANALYSIS

Time series data analysis is main focus on this research and identification of various time series pertains through visual inspection by time plots and statistical tests, parameters estimations, diagnostic checking and selected appropriate model of the climatic factor original data series in order to determine the best fitted ARIMA.

The series plots in figure: 1 above was assumed to be stationary due to the fact that the series were not randomly distributed.



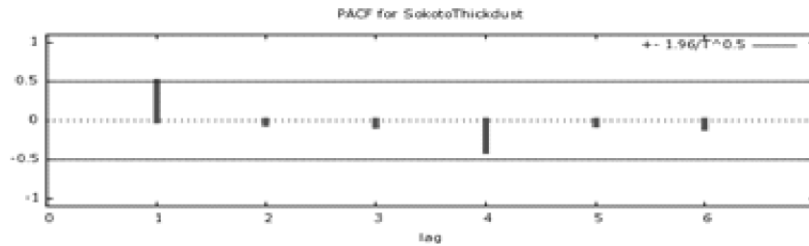


Figure 2. ACF and PACF of thick dust data series (2002-2016)

The figure 2 ACF and PACF plots above were observed under a range of -0.5 to 0.5 respectively in which the respective time series of both ACF and PACF plots of the climatic factor were found within the range of -0.5 to 0.5 as such the respective auto correlations of residuals were assumed to be insignificants in correlations over time. Such assumption would be backed up with statistical tests.

RESULTS AND DISCUSSION

Table: 1 summary results of the tested hypothesis in respect of the climatic factor such as thick dust hays in Sokoto metropolis.

FACTOR	TEST	T. STATISTIC (5%) level	P-VALUE	DECISION	CONCLUSION
(A) Thick dust hays.	KPSS test	-2.010	0.537	H ₀ is accepted	The series has no unit roots. i.e. the series are stationary.
	Portmanteau test	65.46	0.082	H ₀ is accepted	All lags correlations are zero.

Based on the above tested hypothesis and model identifications of the respective climatic factors data series, this research identified the variable for the model is: - Y-ARIMA (p, 0, q) and the variable can be mathematically expressed generally as:-

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \tag{8}$$

$$e_i \sim N(0, \sigma^2).$$

And $\mu + (1 - \phi_1 L - \phi_2 L^2) y_t = (1 - \theta_1 L - \theta_2 L^2) e_t \tag{9}$

$$e_i \sim N(0, \sigma^2).$$

The equation (8) and (9) are ARIMA models standard equations in line with Box, Jenkins and Riesel.

ARIMA model parameters estimations

Model parameters estimation is achieved through software package designed for Box-Jenkins ARIMA model programme. As such, numbers of iterative and repetitive Procedures are involved. This procedure is concerned with different combinations of model orders in terms of autoregressive and moving average orders in different times. Such different order of combinations of autoregressive moving average should be run in the software during the analysis repeatedly, until the required variable for the model parameters are achieved. As such, this research obtained the following estimated model parameters in the tables below as:-

Table: 2. estimated ARIMA (2, 0, 3) parameters based on annual average thick dust hays with trend effect in Sokoto metropolis

MODEL	P/ ESTIMATE	S. ERROR	T. VALUES	P/LEVEL	R.M.S.E	P-SEU.R ²	R. AUTOC.
ARIMA (203)	AR(1) 1.751707	0.09188229	19.0647	0.000000	1.113667	54.624037	NOT SIGNIF. AT ALL LAGS
	AR(2) -0.9920747	0.1039673	-9.5422	0.000000			
	MA(1) 1.708099	0.2575017	6.6333	0.000000			
	MA(2) -0.6887443	0.5001813	-1.3770	0.168516			
	MA(3) -1.05337318	0.26198	-0.2037	0.838565			

The above respective ARIMA model estimated parameters are on the basis of ARIMA report on auto correlations of residuals where the errors autocorrelations not significant if correlations values not greater than 0.516398 as shown below and revealing model in adequacies respectively:

Table: 3. Autocorrelations of Residuals on ARIMA (2, 0, 3) with Trend

Autocorrelations of Residuals

Lag	Correlations	Lag	Correlations	Lag	Correlations	Lag	Correlations
1	-0.025230	4	-0.104415	7	-0.069620	10	-0.094736
2	-0.301635	5	0.001911	8	-0.097986	11	0.002135
3	0.172635	6	0.160786	9	0.021352	12	-0.021166

Significant if |Correlation| > 0.516398

Table: 4. Portmanteau Test on ARIMA (2, 0, 3)

Lag	DF	Portmanteau test values	Prob. Level	Decision (0.05)
6	1	3.41	0.064611	Adequate Model
7	2	3.57	0.167848	Adequate Model
8	3	3.92	0.270328	Adequate Model
9	4	3.94	0.414390	Adequate Model
10	5	4.40	0.493885	Adequate Model
11	6	4.40	0.623179	Adequate Model
12	7	4.43	0.728577	Adequate Model

Selection of appropriate ARIMA model

Diagnostic checking: Portmanteau-test of Box and Jenkins (1970), checks if the estimated residuals were in closed form with white noise process. It also checks for the residuals values to be closed to zero. Diagnostic checking checks the model fitness by revealing the in adequacies of the models and by considering the residuals autocorrelations not significant. Therefore, based on the above mentioned statistical reasons table 10 was obtained from tables of estimates. This research diagnosed and selected ARIMA (2, 0, 3) for annual moving average thick dust hays with trend effect and the general mathematical representations of the estimated model in terms of its order is as follows:-

$$y_t = \mathcal{I}_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} \tag{10}$$

And the mathematical representations of the diagnosed and selected fitted model based on its estimated parameters are:-

$$^* Y_t = 2.958 + 1.752 y_{t-1} - 0.992 y_{t-2} + 1.708 e_{t-1} + 0.689 e_{t-2} + 0.053 e_{t-3} \tag{11}$$

CONCLUSION

From ongoing discussions in this research, it is clear that climatic factor such as thick dust hays posed hazards to planning and management, agricultural productivities, air flight

operations, life and properties in Sokoto metropolis. As such successful implementations of the above mentioned resides on both materials and favorable climatic conditions, where favorable climatic conditions depends on reliable weather / climatic estimates which also depend on accurate weather modeling, that:

ARIMA model has been employed on the original time series data of the climatic factor such as, thick dust hays, in Sokoto metropolis and came up with fitted ARIMA model as ARIMA model (2, 0, 3). This research concluded that the model is best fitted model modeling visibility of Sokoto metropolis.

REFERENCES

- Box, G. E. P., and Jenkins, G. M. (1976). Time Series analysis: forecasting and Control, revised edition. Holden- Day, www.Sci epub.com/reference/56270.
- Box, G.E.P., Jenkins, G. M., and Riesel, G. C. (1994). Modeling the Coherence in short run. Nominal exchange rates: *A multivariate generalized ARCH approach. Review of Economics and Statistics.* <https://books.google.com.ng/booksisbn/1400839254>.
- Box, G., and Jenkins, G. (1970). Time series analysis: Forecasting and control, Holden-day, Sanfransisco. [www.scrip.org/\(s/I75map453edsnp55\)/reference/references_papers.aspx?](http://www.scrip.org/(s/I75map453edsnp55)/reference/references_papers.aspx?)
- Hamza, W., Alhassini, H., Enan, M., and Beer, D. (2011). Aquatic Eco system Health and Management. *14(3): 260 - 268: Doi:10.1080/14634988.2011.60127*
- Jickells, T. D., An, Z. S., Andersen, K. K; Baker, A. R., Bergametti, G., Brooks, N., CaoTegen, I. and Torres, R. (2005). *Global Iron Connections between desert dust, Ocean Biogeochemistry.* Doi: *10.1126/Science 1105959. Science 308 (5718): 67*
- Jung, L. Box, G. (1978). A diagnostic tool used to test the lack of fit of a time series model. <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4481.htm>.
- Kwiatkoswski D, Phillips P. C. B, Schmidt P. and Shin (1992). *Testing the null of stationarity against the alternative of a unit root. How sure are we that the economic time series have a unit root? Journal of Economic 54:159-178.*
- Mahowald, N. M., Baker, A. R., Bergametti, G., Brooks, N., Duce, R. A., Jickells, T. D., Kubilay, N., Propero, J. M., and Tegen, I. (2005). *Global Biochemical Cycles. Vol. 19, GB 4025, doc: 10.1029/2004 GB002402, 2005.*
- Musa, Y. (2014). Modeling an average monthly temperature of Sokoto metropolis using short term memory models. *International Journal of academic research in business and social science 4 (7).*
- Nichovic, S., Scarza, R. A., Mahowald, N., Ghan, S., Zender, Si, Kok, J. F., Liu, X., Zhang, Y, and Albani, S. (2013). *Atmospheric chemical Physics. 15, 537-561, 2015 www.atmos-chem-phys.net/15/537/2015/doi:10.5194/acp-15-537-2015*
- Okin, G. S., Mahowould, N., Chadwick, O. A., and Artako, P. (2004). *Impact of desert dust on Biochemist of Phosphorus in Errestrial Ecosystem Global Biogeochem. Chem.cycles, 18GB 2005, doc 10.1029/2003GB002145.*
- Stefanski, R. and Sivakumar, M. V. K. (2009). Impact of Sand Dust Storms on Agricultural and Potential Agricultural Applications of a ADSWS. *IOP Conference series earth and environmental science 7 (1). Doi: 10.1088/1755-1307/7/1/01206.*
- Schulz, M. H. Zerbino, D. R. Vingron, M., and Birney, E. (2012). The Atmospheric Transport and Disposition of Mineral dust to the Ocean: *Implication for Research, Environmental Science and Technology.* <https://dio.org/1029/2012GL051000>
- Usman, A., Olaore, K. O., and Ismaila, G. S. (2013). Analysis of visibility condition in Sokoto metropolis. *International Journal of Basic and Applied Science Vol. 01, NO. (04).*
- Wang, L., Evans, J., Andrews, H. K., Beckstead, R. B; Thumel, C. S; Bashirullah, A. (2008). *A genetic Screen identifies new regulators of steroid triggered Cell death in Drosophila Genetic. 180(1): 269-281 http://flybase.org/reports/FB.f0207097.html.*
- Yusuf, A. K. (2002). The hazards of some climatic elements on educational planning in Sokoto metropolis. *The Farfaru Journal of Multi-Disciplinary Studies*