Time Series Analysis of Precipitation in Lake Chad using the Prophet Forecasting Procedure

S. J. Okonkwo*; P. A. Ukoha; E. D. Adedoyin; A. R. Adewoye

The Lake Chad Ecosystem Modelling Group,
Department of Environmental Modelling and Biometrics,
Forestry Research Institute of Nigeria, Forest Hill, Jericho Ibadan, Nigeria.
E-mail: sj.okonkwo@gmail.com*

Abstract — **Lake Chad, African's largest endorheic lake is a source of fresh water and irrigation for over 40 million inhabitants. Historically, Lake Chad total water surface was once over 25, 000 Kilometre square and is currently on the decline. Aside seasonal supplies of water from its large tributaries or basins which span 8 countries, making it Africa's largest basin. The lake precipitation regime is itself an important source of water to Lake Chad: The objectives of this study were to understand the behavioural patterns of precipitation and their relationship with each other over the last two decades. Lake Chad was divided into five strata, namely: South Basins, Archipelagos, Great Barriers, North Basins, and Inundated Areas. Daily precipitation values (7,164 observations per stratum) were extracted from the National Aeronautics and Space Administration's Global Precipitation Measurement (GPM) Integrated Multi–SatellitE Retrievals archived in Google Earth Engine (GEE) platform. GPM datasets from January 2001 to August 2020 were extracted for each stratum using java scripts in GEE. Number of lags used for each stratum were selected using Akaike Information Criterion (AIC): 25, 32, 22, 32, and 35 respectively. Augmented Dickey–Fuller (ADF) test revealed each time series to be stationary at 0.05 confidence level. Prophet Forecasting was carried out on each time series with different change point prior scale. Accuracy was judged with Mean Square Error (MSE), Root Mean Square Error (RMSE) and Welch's t-test at 0.05 confidence level.**

Keywords - *ADF, AIC, Change Point, Precipitation, Prophet Forecasting, RMSE.*

I. INTRODUCTION

Lake Chad is situated within the borders of Cameroun, Chad, Niger, and Nigeria. It supports the socioeconomic activities of almost 40 million people (Pham–Duc et al., 2020) and the ecological activities of a wide variety of vegetation and aquatic life (Zeiba et al, 2017). The lake has been described as shrinking at an alarming rate (Ikusemoran et al., 2018) with only 90% decrease from 1990 (Policelli et al., 2018). Decadal climate change has been cited as a major culprit (Gao et al., 2011) with consequent developmental, food, health and security challenges (PLAN International, 2018).

This study answered the following questions: How has precipitation changed in Lake Chad over the last two decades? If there are changes, has it been the same across the Lake? Can precipitation in one part of the lake be useful in predicting precipitation in another? What should we expect about future precipitation around Lake Chad? This study focuses on precipitation alone: it does not consider other sources of water to Lake Chad. Neither does it consider other environmental or socio–economic factors with respect to the lake.

Zeiba et al. (2017) examined the seasonal migration of people in search of water and other resources around the Lake Chad basin and came to the conclusion that the shrinking lake, political instability from the Nigerian border, and the increasing population around Lake Chad is further compounding the delicate situation of the inhabitants in that region. Ikusemoran et al. (2018) came to the conclusion that what is left of Lake Chad is towards the South East of the region and recommended a close monitoring of the climatic conditions of the Lake Chad region. Pham–Duc et al. (2020) provided an understanding of the hydrological dynamics in the Lake Chad region and proposed a developmental template. Policelli et al. (2018) "addressed important questions about the size and trend of Lake Chad's total surface water area". The only noticeable advantage of the shrinking Lake is the new farming opportunities it provides but that is quickly cancelled out by unsustainable agricultural practices which some are calling "the impacts of poor agricultural resource management" such as "land degradation and loss of

agricultural biodiversity" not to mention the reduced access to water (Zeiba, et al., 2017).

When carrying out time series analysis, there are numerous models to choose from, but Akaike Information Criterion (AIC) is a consistent and robust means of model selection (Pho et al., 2019). According to Paparoditis and Politis (2018): "the limiting distribution of the augmented Dickey–Fuller (ADF) test under the null hypothesis of a unit root is valid under a very general set of assumptions that goes far beyond the linear AR(∞) process assumption typically imposed".

The real issues associated with granger causality analysis lies with stationarity, linearity, noise, sampling rates and temporal or spatial aggregations that occur during data acquisition (Barnett et al., 2018). Producing "reliable and high quality forecast" is a daunting task because time series forecasting is a specialized skill that involves years of practice (Taylor & Letham, 2017).

The Prophet Forecasting Procedure makes it easy to adjust the parameters of a decomposable time series model by an optimization algorithm that terminates at convergence. Taylor and Letham (2017) proposed "a modular regression model with interpretable parameters" that is easily adjusted by anyone with a practical knowledge and understanding of time series.

II. METHODOLOGY

The study area is divided into five classes of interest (Archipelago, Great Barrier, Inundated, North Basin and South Basin), here called strata (Figure 1). Four of the strata (South Basin, Archipelago, Great Barrier, and North Basin) were based on stratifications suggested by Lemoalle et al., 2012. An extra stratum (Inundated Area) was added to the Northern Part of the Lake. Five (5) geometric points were randomly placed within each stratum and precipitation values were extracted from the Global Precipitation Measurement (GPM) Integrated Multi–SatellitE Retrievals (IMERG) archive (Huffman et al., 2019) using Google Earth Engine (Gorelick et al., 2017). A buffer of 30 metre radius was created around each point and data extracted from these buffers were averaged for each stratum. Each stratum had a total of 7,164 observations from 1st of January 2001 to 12th of August 2020.

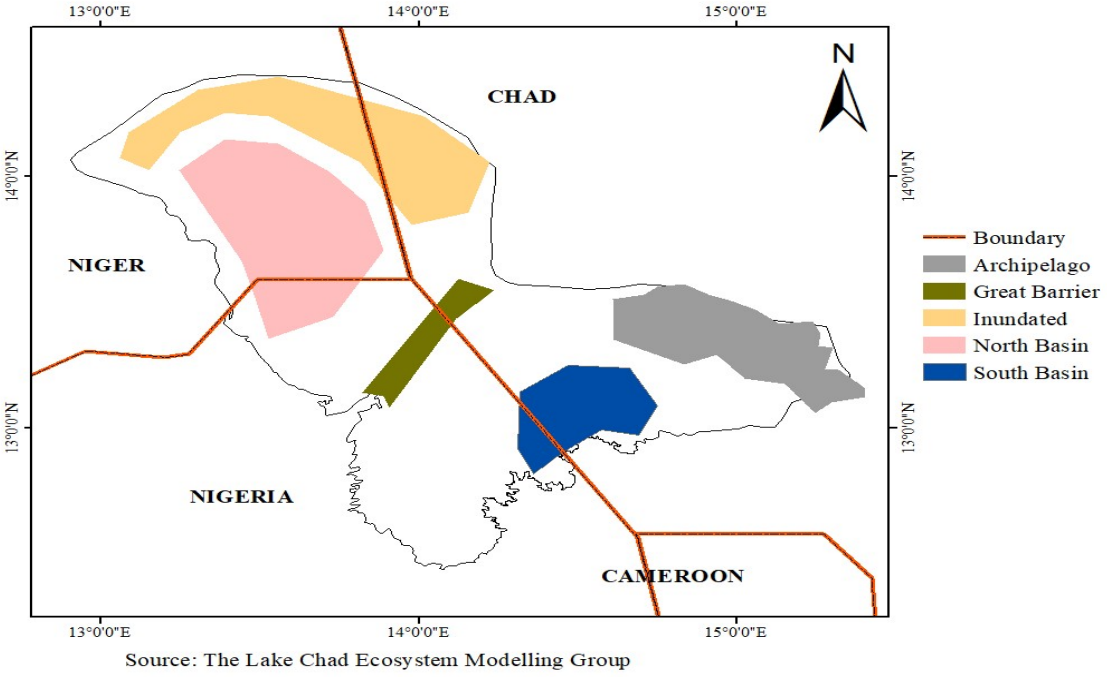

Figure 1: The Study Area (Lake Chad) showing the Different Strata

Consequently, let $\boldsymbol{y_i}$ be the precipitation series observed at the stratum of interest, $\boldsymbol{i}$ $(= 1,2, \dots ,5)$; that is,

$$\boldsymbol{y_i} = [y_i(1), y_i(2), \dots, y_i(t)], \quad (1)$$

where $y_i(t)$ represent the observed value of precipitation at stratum $\boldsymbol{i}$ at time $\boldsymbol{t}$ and each category of the class of interest (that is, Archipelago, Great Barrier, Inundated, North Basin and South Basin) is presented by $\boldsymbol{i}$.

Akaike Information Criterion (AIC)

Let Stratum 1 be our Stratum of interest, and let $y_{Stratum\,1}$ be a time series for precipitation in Stratum 1. Let AR(1), AR(2), . . ., AR(n), AR($n+1$), AR($n+2$), . . ., AR(m) be the possible number of autoregressive processes: $1, 2, \ldots, n, n+1, n+2, \ldots, m$ are the lag values for the respective AR processes, m is the maximum lag, and k_1, k_2, . . ., k_n, k_{n+1}, k_{n+2}, . . ., $k_m$be the number of estimated parameters in models AR(1), AR(2), . . ., AR(n), AR($n+1$), AR($n+2$), . . ., AR(m) respectively. The Akaike Information Crite--rion (AIC) (Akaike, 1973; Akaike, 1974; McElreath, 2016) of an arbitrary AR(n) process is given as:

$$AIC_n = 2k_n - 2\ln(\hat{L}_n) \qquad (2)$$

where $\hat{L}_n$ is the maximum value of the likelihood function of the AR(n) process. Given AR(1), AR(2), . . ., AR(n), AR($n+1$), AR($n+2$), . . ., AR(m) with the respective AIC values AIC_1, AIC_2, . . ., AIC_n, AIC_{n+1}, AIC_{n+2}, . . ., AIC_m. The preferred AR process is the one with minimum AIC value.

Augmented Dickey–Fuller Test

Let Stratum 1 be our Stratum of interest, and let $y_{Stratum\,1}$ be a time series for precipitation in Stratum 1. The autoregressive model of $y_{Stratum\,1}$ depicted with the autoregressive process AR(l) and written as:

$$y_{t_{Stratum\,1}} = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_l y_{t-l} + \varepsilon_t \qquad (3)$$

Where θ_0 is a constant, $\theta_1, \ldots, \theta_l$ are coefficients of the lags $y_{t-1}, \ldots, y_{t-l}$ respectively, t is the time point of the variable of interest, l is the number of lags, and ε_t is the error term. The Augmented Dickey Fuller Test (Elliott et al., 1996; MacKinnon, 2010) is applied to the model:

$$\Delta y_{t_{Stratum\,1}} = \theta_0 + \varphi t + \partial_1 y_{t-1} + \partial_2 y_{t-2} + \ldots + \partial_{l-1} y_{t-l+1} + \varepsilon_t \qquad (4)$$

Where θ_0 is a constant, φ is the time trend coefficient, and l is the number of lags of the AR process. The lag length is determined by examining the Akaike Information Criterion (AIC) and picking the AR model with the appropriate lag length.

Prophet Forecasting Procedure

Prophet model (Taylor & Letham, 2017) is an additive regression technique developed by Facebook to handle business–related modelling problems. The model is developed on the idea that a time series is decomposable into three component parts: the trend, seasonal effect and holiday effect (Harvey & Peters, 1990). In a combined state, the Prophet Forecasting Model assumes the form:

$$y_i(t) = \varphi(t) + \tau(t) + \omega(t) + \varepsilon_t \qquad (5)$$

where $y_i(t)$ is the value of precipitation for stratum $\boldsymbol{i}$ at time $\boldsymbol{t}$, $\varphi(t)$ is the trend, τ_t is the additive seasonality, ω_t is the holiday, and ε_t is the error term. Multiplicative seasonality can be achieved through log transformation (Taylor & Letham, 2017). In this work, the precipitation measurements are considered to be continuous throughout the period of observation. In other words, there are no occasions of halting measurement. Therefore, the holiday effect, $\omega(t)$, takes the value zero. Therefore, equation (5) is simplified to:

$$y_i(t) = \varphi(t) + \tau(t) + \varepsilon_t \qquad (6)$$

Since the maximum precipitation observation do not exhibit a saturating growth, the trend component of equation (6) is modelled by a piecewise linear growth model of mathematical form:

$$\varphi(t) = (r + a^T(t)\vartheta)t + (p + a^T(t)\delta) \qquad (7)$$

where r is the growth rate, $a^T(t)$ is the transpose of the vector $a(t) \in \{0, 1\}$, which is the vector of the rate at any time t, $\vartheta \in R$ is a vector of rate of adjustment, p is the offset parameter, and δ is the actual adjustment.

The multiperiod seasonal component of equation (6) is modelled using the Fourier series expressed in the form:

$$\tau(t) = M(t)\partial = \sum_{i=1}^{N}\left(\alpha_i \cos\left(\frac{2\pi i t}{P}\right) + \beta_i \sin\left(\frac{2\pi i t}{P}\right)\right) \qquad (8)$$

where

$$M(t) = \left[\cos\left(\frac{2\pi(1)t}{365.25}\right), \ldots, \sin\left(\frac{2\pi(N)t}{365.25}\right)\right] \qquad (9)$$

P is the regular, or annual, periodicity ($P = 365.25$) while N is the number of parameters. The seasonality model is fitted by estimating the $2N$ parameters $\partial = [\alpha_1, \beta_1, \ldots, \alpha_N, \beta_N]^T$, which is assumed to have a Gaussian distribution ($\partial \sim Normal\ (0, \sigma^2)$).

Similarity measure

The measure of similarity is built around three intuitions: (1) the more features two objects share, the more similar they are; (2) the more differences there are between two objects, the less similar they are; and (3) maximum similarity occurs when two objects are similar. It is quantified as either the cost of transforming one object into another or as the inverse of the distance between two or more objects (Cleasby et al., 2019). To assess the similarity between two time series variables, the pairwise distance of the two variables' trajectory is measured.

Several methods are available for time series trajectory quantification. One of the most popular and

robust of these methods is the Dynamic Time Warping (DTW) algorithm (Cleasby et al., 2019). DTW algorithm is an elastic measure (can be used with trajectories of different lengths or phases) for computing the distance and alignment between two time series (Seto et al., 2015). For the purpose of this work, the algorithm of the DTW is presented as follows.

Let $\boldsymbol{y_i}$ be as defined in equation (1) above, where $\boldsymbol{i}$ $(= 1,2,\dots,5)$, and each $\boldsymbol{i}$ is of length $\boldsymbol{t}$. The DTW algorithm for any two of the series in $\boldsymbol{i}$, given $\boldsymbol{y_i} = [y_i(1), y_i(2), \dots, y_i(t)], t \in \mathbb{N}$ and $\boldsymbol{y_j} = [y_j(1), y_j(2), \dots, y_j(t)], t \in \mathbb{N}$, involves derivation of the distance, or local cost, matrix, $\mathbb{C} \in \mathbb{R}^{T \times T}$ (where $T = t$), which is the pairwise distances between the alignment of the two sequences, $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$:

$$\mathbb{C}_l \in \mathbb{R}^{T \times T}: c_{i,j} = \|y_i - y_j\|, i \in [1:T], j \in [1:T] \quad (10)$$

The alignment path built by the DTW algorithm is a sequence of points $p = (p_1, p_2, \dots, p_K)$ with $p_l = (p_i, p_j) \in [1:T] \times [1:T]\ for\ l \in [1:K]$ which must satisfy the boundary condition $(p_1 = (1,1)$ and $p_K = (T,T))$, monotonicity condition $(t_1 \leq t_2 \leq \cdots \leq t_K)$ and step–size condition $(p_{l+1} - p_l \in \{(1,1),(1,0),(0,1)\})$. The cost function associated with the warping path computed with respect to the local cost is:

$$c_p(y_i, y_j) = \sum_{l=1}^{L} c(y_{tl}, y_{tl}) \quad (11)$$

while the optimal warping path, $c_{p*}(y_i, y_j)$, also referred to as the warping path with the minimal associated cost, is given as:

$$c_{p*}(y_i, y_j) = min\{c_p(y_i, y_j), p \in P^{T \times T}\} \quad (12)$$

where $P^{T \times T}$ is the set of all possible warping paths which builds the accumulated cost matrix at time cost $O(T^2)$. The value of $c_{p*}(y_i, y_j)$ ranges from zero to infinity with larger values connoting lower similarity.

III. RESULTS

Data analysis was carried out using Python Programming Language version 3.7 (Van Rossum & Drake, 2009). Packages used include: NumPy (Oliphant, 2006; van der Walt et al., 2011), Matplotlib (Hunter, 2007), Pandas (McKinney & others, 2010), SciPy (Virtanen et al., 2020), Seaborn (Waskom et al., 2017), Statsmodels (Seabold et al., 2010) and Prophet (Taylor & Letham, 2017).

Table 1: Pearson's Correlation Matrix of Time series

	South Basin	Archipelago	Great Barrier	North Basin	Inundated
South Basin	1.0000				
Archipelago	0.7788	1.0000			
Great Barrier	0.7950	0.7166	1.0000		
North Basin	0.6159	0.5996	0.7963	1.0000	
Inundated	0.4830	0.5050	0.6070	0.7831	1.0000

Preliminary analysis showed that there has been a steady decline in the amount of surface water in Lake Chad (Figure 2). Also, for the most part, there was strong positive correlation between pairs of precipitation observations from the different strata (Table 1).

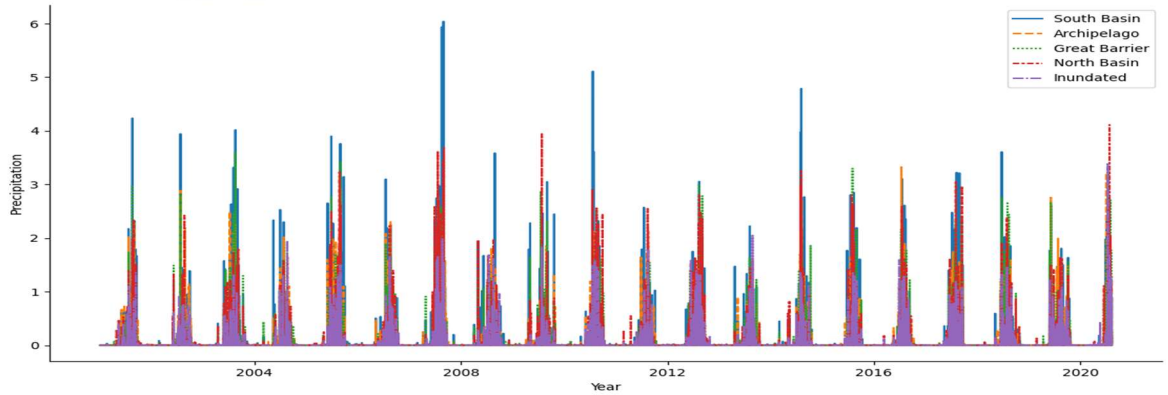

Figure 3: Time Series Graph of Precipitation from all Strata

Figure 3 shows a seasonality in the amount of daily precipitation level over the years. It also shows a rhythm in the seasonality of precipitation from the different strata. Each time series was found to be stationary after undergoing Augmented Dickey–Fuller (ADF) test at 0.05 confidence level (Table 2).

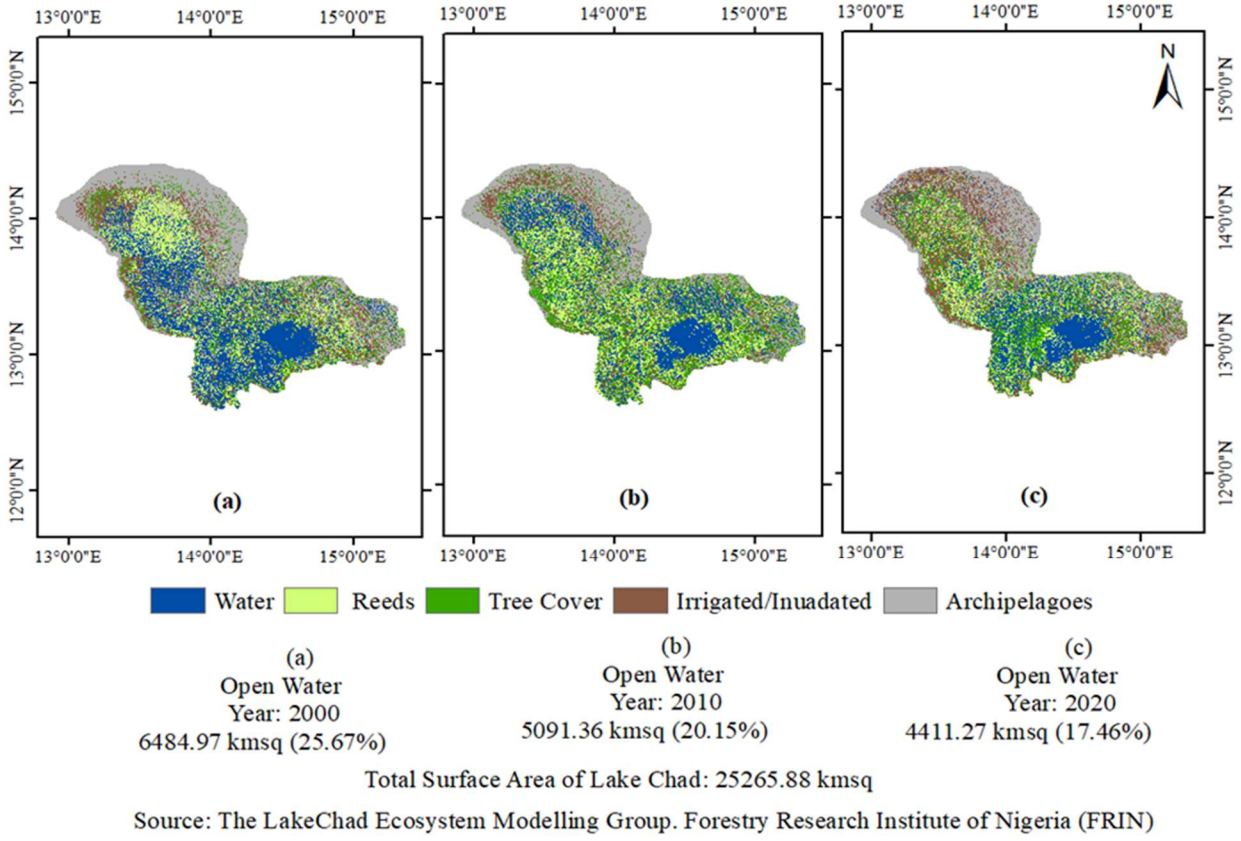

Figure 2: Classified Images of Lake Chad Showing the Decline in Open Water

Figure 2 contains classified images of the study area over the years from 2000 to 2020. The images contain five classes: Water, Reeds, Tree Cover, Inundated areas, and Archipelagoes. There has been a steady decline in the amount of Open Water over the years. From 25.67% in Year 2000 to 17.46% in Year 2020.

Table 2: Augmented Dickey–Fuller Stationarity Test

Stratum	Number of Lags	ADF Test Statistic	P–value	
South Basin	25	-8.8176	1.9150e-14	**
Archipelago	32	-8.0691	1.5669e-12	**
Great Barrier	22	-9.2393	1.5981e-15	**
North Basin	32	-8.6937	3.9781e-14	**
Inundated	35	-7.8339	6.1889e-12	**

Number of lags used in the study were based on Akaike Information Criterion (AIC). Welch's t–Test was carried out at 0.05 confidence level between pairs of variables to see if there is a statistical significant difference in

precipitation from the different strata. Granger causality was also tested between pairs of time series at 0.05 confidence interval. The time series each with 7,164 time points were split into two parts for the forecasting. 6,939 points were used for training while 225 were used for validating the model. Forecasting was carried out on each time series with different change point prior scales (0.01, 0.05, 0.10, and 0.50). The model converged for each time series after different number of iterations. The Prophet optimization algorithm terminates once convergence has been detected. Convergence is detected when the relative gradient magnitude is below the tolerance level. Accuracy was judged with Mean Square Error (MSE), Root Mean Square Error (RMSE) and Welch's t-test at 0.05 confidence level.

There is no statistically significant difference in the amount of precipitation in the Archipelagos, Great Barrier, and the North Basin (Table 3). However, there is a significant difference in the amount of precipitation from the South Basin and the Inundated areas against each other and against other strata.

Table 3: DTW Distance and Welch's t–Test of pairs of Strata from Lake Chad

Strata Pair		Welch's t–Test		Dynamic Time Warping (DTW) Distance
Stratum 1	Stratum 2	Statistic	P – value	
South Basin	Archipelago	5.9850	2.2242e-09 **	15.7908
South Basin	Great Barrier	4.9103	9.2015e-07 **	14.2463
South Basin	North Basin	4.7683	1.8776e-06 **	19.5438
South Basin	Inundated	10.7062	1.3256e-26 **	14.6710
Archipelago	Great Barrier	-1.0313	0.3024	10.8630
Archipelago	North Basin	-1.1391	0.2547	10.3869
Archipelago	Inundated	5.7400	9.6821e-09 **	11.6490
Great Barrier	North Basin	-0.1152	0.9083	12.5077
Great Barrier	Inundated	6.4132	1.4790e-10 **	10.3426
North Basin	Inundated	6.4368	1.2680e-10 **	12.7397

** There is a significant difference in the amount of precipitation from the two strata at 0.05 confidence level.

Prophet Forecasting

There was no significant difference between the predicted and the actual time series data for precipitation from the different strata at 95% confidence interval. Tables 4, 5, 6 and 7 show the results at different change point prior scales: 0.01, 0.05, 0.10, and 0.50. The test statistics for the Welch's t–Test for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.01 (Table 4) are -0.6953, -1.6406, -1.0152, -0.8597, and -0.9612 respectively.

Table 4: Accuracy measurement at Change Point Prior Scale of 0.01

Stratum	Number of Iterations	Mean Square Error (MSE)	Root Mean Square Error (RMSE)	Welch's t – Test	
				Statistic	P–Value
South Basin	49	0.1403	0.3745	-0.6953	0.4874
Archipelago	44	0.1482	0.3849	-1.6406	0.1022
Great barrier	97	0.1644	0.4056	-1.0152	0.3110
North basin	96	0.1558	0.3947	-0.8597	0.3908
Inundated	114	0.0759	0.2756	-0.9612	0.3374

The Mean Square Error (MSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.01 are 0.1403, 0.1482, 0.1644, 0.1558, and 0.0759 respectively. The Root Mean Square error (RMSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.01 are 0.3745, 0.3849, 0.4056, 0.3947, and 0.2756 respectively.

The test statistics for the Welch's t–Test for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.05 (Table 5) are -0.54989, -1.2660, -0.8462, -0.9343, and -0.8443 respectively. The Mean Square Error (MSE) between the

predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.05 are 0.1401, 0.1472, 0.1642, 0.1560, and 0.0758 respectively.

Table 5: Accuracy measurement at Change Point Prior Scale of 0.05

Stratum	Number of Iterations	Mean Square Error (MSE)	Root Mean Square Error (RMSE)	Welch's t – Test	
				Statistic	P–Value
South Basin	255	0.1401	0.3743	-0.54989	0.5828
Archipelago	119	0.1472	0.3837	-1.2660	0.2067
Great barrier	107	0.1642	0.4052	-0.8462	0.3982
North basin	231	0.1560	0.3949	-0.9343	0.3511
Inundated	149	0.0758	0.2754	-0.8443	0.3993

The Root Mean Square error (RMSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.05 are 0.3743, 0.3837, 0.4052, 0.3949, and 0.2754 respectively.

The test statistics for the Welch's t–Test for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.10 (Table 6) are -0.4997, -1.1656, -0.7657, -0.8541, and -0.7197 respectively. The Mean Square Error (MSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.10 are 0.1401, 0.1470, 0.1641, 0.1558, and 0.0758 respectively.

The Root Mean Square error (RMSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.10 are 0.3743, 0.3835, 0.4050, 0.3948, and 0.2752 respectively. The test statistics for the Welch's t–Test for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.50 (Table 7) are -0.4093, -1.0327, -0.7078, -0.8303, and -0.7056 respectively.

Table 6: Accuracy measurement at Change Point Prior Scale of 0.10

Stratum	Number of Iterations	Mean Square Error (MSE)	Root Mean Square Error (RMSE)	Welch's t – Test	
				Statistic	P–Value
South Basin	335	0.1401	0.3743	-0.4997	0.6177
Archipelago	133	0.1470	0.3835	-1.1656	0.2449
Great barrier	181	0.1641	0.4050	-0.7657	0.4446
North basin	236	0.1558	0.3948	-0.8541	0.3939
Inundated	344	0.0758	0.2752	-0.7197	0.4724

The Mean Square Error (MSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.10 are 0.1401, 0.1470, 0.1641, 0.1558, and 0.0758 respectively.

Table 7: Accuracy measurement at Change Point Prior Scale of 0.50

Stratum	Number of Iterations	Mean Square Error (MSE)	Root Mean Square Error (RMSE)	Welch's t – Test	
				Statistic	P–Value
South Basin	973	0.1400	0.3742	-0.4093	0.6826
Archipelago	289	0.1468	0.3831	-1.0327	0.3027
Great barrier	468	0.1640	0.4050	-0.7078	0.4797
North basin	681	0.1558	0.3948	-0.8303	0.4072
Inundated	506	0.0757	0.2752	-0.7056	0.4811

The Root Mean Square error (RMSE) between the predicted and the actual values for South Basin, Archipelago, Great Barrier, North Basin, and Inundated at a change point prior scale of 0.50 are 0.3742, 0.3831, 0.4050, 0.3948, and 0.2752 respectively.

IV. Discussion and Conclusion

With the exception of the South Basin, and the Inundated areas, there is a no statistically significant difference in the amount of precipitation from most parts of the Lake (Archipelagos, Great Barrier, and North Basin). From Figure 2, we see that the majority of Open Water in the

Lake in Year 2020 is in the South Basin. The Prophet Forecasting Procedure was able to forecast Precipitation with good accuracy at different change point prior scale.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Republished in Kotz, S., and Johnson, N. L., (Editors). (1992). *Breakthroughs in Statistics*, 1, 610–624.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723. Doi:10.1109/TAC.1974.1100705

Barnett, L., Barret, A. B., and Seth, A. K. (2018). Misunderstandings regarding the application of Granger causality in neuroscience. In *Proceedings of the National Academy of Sciences of the United States of America*, 1 - 2. Doi: 10.1073/pnas.1714497115

Elliott, G., Rothenberg, T. J., Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64, 813–836.

Gao, H., Bohn, T. J., Podest, E, Mcdonald, K., and Lettenmaier, D. (2011). On the causes of the shrinking of Lake Chad. *Environmental Research Letters*, 6, 034021. DOI: 10.1088/1748-9326/6/3/034021

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 37 (3): 424–438. Doi:10.2307/1912791

Harvey, A., and Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, 9, 89 – 108.

Huffman, G.J., E.F. Stocker, D.T. Bolvin, E.J. Nelkin, Jackson Tan (2019). GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Data Access Date], 10.5067/GPM/IMERG/3B-HH/06

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. Doi:10.1109/MCSE.2007.55

Ikusemoran, M., Alhaji, M. and Abdussalam, B. (2018). Geospatial Assessments of the Shrinking Lake Chad. *Adamawa State University Journal of Scientific Research*, 6, 114 – 130.

Lemoalle J. (2004) Lake Chad: A Changing Environment. In: Nihoul J.C.J., Zavialov P.O., Micklin P.P. (eds) Dying and Dead Seas Climatic Versus Anthropic Causes. *NATO Science Series: IV: Earth and Environmental Sciences*, 36. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0967-6_13

Lemoalle J., Bader, J–C., Leblanc, M., and Sedick, A. (2012). Recent changes in Lake Chad: Observations, simulations and management options (1973–2011). *Global and Planetary Change*, 80 – 81, 247 – 254. Doi:10.1016/j.gloplacha.2011.07.004

MacKinnon, J.G. (2010). Critical Values for Cointegration Tests. Queen"s University, Dept of Economics, Working Papers. Retrieved from https://ideas.repec.org/p/qed/wpaper/1227.html

McElreath, R. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. *Chapman & Hall/CRC Texts in Statistical Science*, 487.

McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*. 445, 51–56.

Oliphant, T. E. (2006). *A guide to NumPy* (Vol. 1). Trelgol Publishing USA.

Paparoditis, E. and Politis, D. N. (2018). The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. *Econometric reviews*, 37, 955-973.

Pham–Duc, B., Sylvestre, F., Papa, F., Frappart, F., Bouchez, C., and Cretaux, J–F. (2020). The Lake Chad hydrology under current climate change. Scientific Reports, 10, 5498. Doi: 10.1038/s41598-020-62417-w

Pho, K.–H., Ly, S., Ly, S., and Lukusa, T. M. (2019). Comparison among Akaike Information Criterion, Bayesian Information Criterion and Vuong's test in Model Selection: A Case Study of Violated Speed Regulation in Taiwan. *Journal of Advanced Engineering and Computation*, 3, 293 – 303. DOI: http://dx.doi.org/10.25073/jaec.201931.220

PLAN International (2018). *Adolescent Girls in Crisis: Voices from the Lake Chad Basin*. United Kingdom. Available at: https://plan-uk.org/file/lake-chad-reportpdf/download?token=4KhpByvu. (Accessed: 22 August 2020).

Policelli, F., Hubbard, A., Jung, H. C., Zaitchik, B., and Ichoku, C. (2018). Lake Chad Total SurfaceWater Area as Derived from Land Surface Temperature and Radar Remote Sensing Data. *Remote Sensing*, 10, 252. Doi: 103390/rs10020252

Seabold, Skipper, and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python.

In *Proceedings of the 9th Python in Science Conference*. .

Taylor S. J., and Letham, B. (2017). Forecasting at scale. *PeerJ Preprints* 5:e3190v2. Doi: 10.7287/peerj.preprints.3190v2

Van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13, 22–30. Doi:10.1109/MCSE.2011.37

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R.–J. Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.–A. Harris, C. R., Archibald, A. M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, Paul and SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261 – 272. Doi: 10.1038/s41592-019-0686-2

Waskom M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., de Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K., Miles, A., Ram, Y., Yarkoni, T., Williams, M. L., Evans, C., Fitzgerald, C., Brian, Fonnesbeck, C., Lee, A., and Qalieh, A. (2017). *mwaskom/seaborn: v0.8.1* (September 2017). Zenodo. Doi: 10.5281/zenodo.883859

Zeiba, F. W., Yenogh, G. T., and Tom, A. (2017). Seasonal Migration and Settlement around Lake Chad: Strategies for Control of Resources in an Increasingly Drying Lake. *Resources* 2017, 6, 41. Doi: 103390/resources6030041