5-1-2019

ANALYSIS AND FORECASTING OF CLIMATE CONDITIONS IN SOUTH FLORIDA– A LAKE OKEECHOBEE CASE STUDY

Ohimai Imoukhuede

ANALYSIS AND FORECASTING OF CLIMATE CONDITIONS IN SOUTH FLORIDA– A LAKE OKEECHOBEE CASE STUDY

by

OHIMAI IMOUKHUEDE

Presented to the Faculty of the Honors College of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

HONORS BACHELOR OF SCIENCE IN GEOLOGY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2019

ACKNOWLEDGMENTS

I would like to thank God for strengthening me to complete this project, even though it started out roughly. I would also like to thank Dr. Hyeong Moo-Shin for accepting to help me with this project when I had to change advisors. Last, but certainly not least, I would like to thank my family, both nuclear and extended for all their support throughout my undergraduate experience.

May 03, 2019

ABSTRACT

ANALYSIS AND FORECASTING OF CLIMATE

CONDITIONS IN SOUTH FLORIDA–

A LAKE OKEECHOBEE

CASE STUDY

Ohimai Imoukhuede, B.S. Geology

The University of Texas at Arlington, 2019

Faculty Mentor: Hyeong Moo-Shin

The tropical nature of Florida's climate allows for a variety and unpredictability in conditions. The main focus of this study is to examine such variety. With conditions ranging from extreme flooding in the rainy season (June to October) to extreme drought in the dry season (November to May), it is important to study and document past climatic conditions, their effects, and the possibility of reoccurrence.

The objective of this research is to utilize remote sensing and GIS techniques in the investigation of changes in the surface area of Lake Okeechobee. This lake is located in southern Florida, which is known for its drought and flooding potential. After examining the changes in lake surface area, the identified trends will then be used in a predictive statistics model to forecast possible future changes in the lake's surface area. These

forecasted changes can then be used to tell a story about the potential for drought and/or flooding in South Florida's near future.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Lake Okeechobee is a subtropical inland lake located in South Florida. It is the second largest lake which resides entirely in the continental United States, spanning an area of approximately 700 sq. mi. The lake is shallow, with an average depth of 2.7 m. Its water levels are significantly controlled by a system of dikes, canals and pumping stations which distribute water for human consumption and agricultural uses (Brezonik and Engstrom, 1998). The lake underwent major alteration in the 1900's in order to control flooding and drain areas for agricultural use. In this study, Lake Okeechobee is used as a marker for changes in climate conditions in the South Florida region.

The wavering nature of South Florida's climate allows for the ever-present risks of either flooding or drought. This is partly due to the highs and lows of natural water supply to the area as well as the natural and artificial structures put in place to accommodate these supplies. The latter is more so true with regards to flooding. Typically, the effects of high groundwater tables in conjunction with high canal water levels significantly increase the risk of flooding (Czajkowski et al., 2018). This is the particularly the case in South Florida during the rainy season (June to October). Conversely, during the dry season there tends to be a significant decrease in precipitation. This decrease has been further amplified overtime, as studies show that while several locations in the United States showed significant overall increases in precipitation from 1970 – 2013, Florida saw a significant decrease over this time period (Ganguli and Ganguly, 2016). Still within this time period, South Florida experienced two major cases of drought, one in 2001 and the other spanning from July 2007 to April 2008 (James, 2016). As a result, it is highly imperative that water levels be monitored for the awareness and safety of the public.

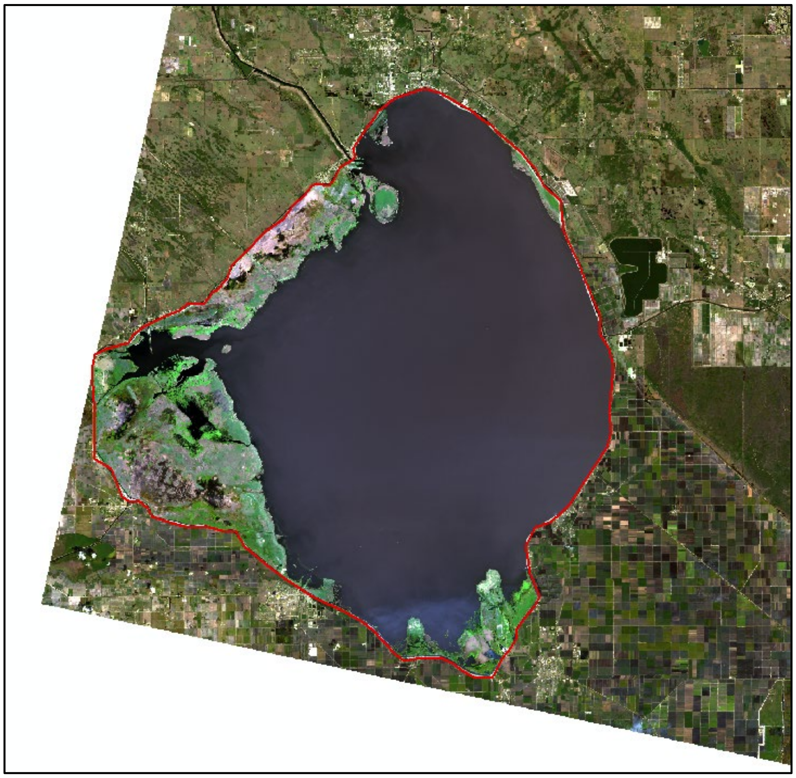

Figure 1.1: Imagery from 2018. The area outlined in red is the study area, Lake Okeechobee

1.2 Image Classification

Waterbody surface area extraction from satellite imagery is an effective way of monitoring relative changes in water levels. However, there is room for improvement with regards to the absolute accuracy of these measurements. The two most popular methods for obtaining these measurements are known as Supervised and Unsupervised image classification (Rozenstein and Karnieli, 2010). Supervised classification involves the use of machine learning, whereby the user provides samples of what each land cover should ideally look like, and the computer follows an algorithm to reproduce these land features along with their pixel counts. On the other hand, unsupervised classification requires little input from the user. It is a more automated process carried out by the computer with the user only specifying the maximum number of classes to group land features. Both of these classification techniques are used to group land cover into distinct classes based on their land use. However, studies have shown that unsupervised classification is superior to supervised classification (Rozenstein and Karnieli, 2010) based on the accuracy of the classification. Two other methods for image classification are also popular among environmental research scientists. These are Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI). NDVI is used to distinctly separate out vegetation from other land use classes in a satellite image, while NDWI is used to distinctly separate out water from other land use classes in a satellite image (McFeeters, 1996). These methods are based on normalized indices, which are used to interpret the degree of vegetation health (NDVI) and the degree of water occurrence in landcover (NDWI). NDWI has since been modified to account for differentiation between water and built-up features (Xu, 2006). This modification allows for the image to look sharper by increasing its spectral signature due to an increase in wavelength (near infrared to mid infrared). This new index is known as Modified Normalized Difference Water Index (MNDWI).

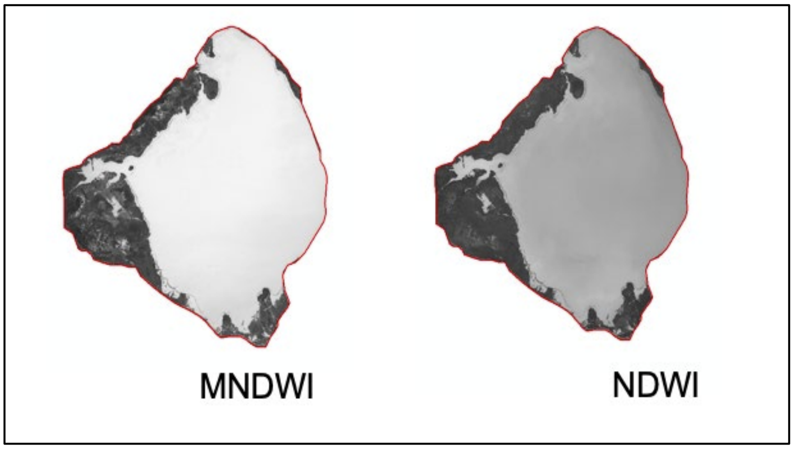

Figure 1.2: Contrast between MNDWI and NDWI Images

1.3 Significance of Research

The main significance of this research is to document, understand and analyze environmental changes in the drought and flood prone South Florida region, using water surface area change as a marker. As stated earlier, it is very important to monitor changes in water levels overtime for the safety and awareness of the public.

Palm Beach, Broward and Miami-Dade, the three most populous counties in Florida, are located on the southeast coast of the state. Therefore, the effects of flooding and drought would have significant economic impacts, as the residents of these areas are important end-users of publically supplied water withdrawn from Florida aquifers today (Czajkowski et al., 2018). Hence, if a significant drought occurs in these areas, a large population of people would be affected.

Though the existing structure of high groundwater table and high canal levels aids in meeting government requirements for water supply in the dry season, the other side (flooding) also suffers from the same structure in the rainy season. Moreover, despite the structure put in place, significant drought still occurs in these areas in some years. Interpreting trends in water level and changes definitely aids in better understanding the possible effects of global warming. This would lead to better preparation and caution so as to minimize anthropogenic impacts on earth's natural system.

CHAPTER 2

METHODOLOGY

2.1 Input

Landsat Imagery from the United States Geological Survey (USGS) spatial database is used as input in this study. This imagery covers a portion of South Florida that contains the study area. Depending on the year in focus, various types of Landsat imagery are used. From 2007 to 2011, Landsat 4-5 imagery is used, as this is the best available Landsat imagery data during that time. This imagery contains seven bands corresponding to light in the electromagnetic spectrum. For 2012, Landsat 1-5 imagery is used. This satellite imagery contains four bands, with the Blue band and all other bands after the Middle Infrared band excluded. Finally, from 2013 to 2018, Landsat 8 imagery is used for analysis. This imagery contains 11 bands, out of which seven will be used for analysis.

2.2 Intermediate Parameters

Intermediate parameters are those used during the analysis that are neither absolute input nor absolute output. In other words, they are the outputs of initial inputs, which are then used as inputs for final outputs. These parameters are created in the ArcGIS environment by built-in geoprocessing tools and scripts which work to change the form of already existing data.

In this study, the following geoprocessing tools are used to create intermediate parameters:

- *Composite Bands*: This geoprocessing tool takes in Landsat raster band data and combines them to create a natural look image. The natural look image created is the first intermediate parameter.
- *Clip*: After the natural look image is created, the clip tool serves as a “cookie-cutter” which aids in cropping out the specific study area, which for this study (as stated earlier), is the area enclosed by the visible boundary surrounding Lake Okeechobee.
- *Raster Calculator*: This is a Spatial Analyst geoprocessing tool which performs map algebra to get a desired output. For this study, map algebra is used to perform a Modified Normalized Difference Water Index (MNDWI) calculation on the study area, producing an image that combines Green and Near Infrared light to better identify and distinguish water features from vegetation and bare earth features.
- *Iso Cluster Unsupervised Classification*: This is a built-in ArcGIS script which is accessible from the Image Classification Window. This script contains a machine learning algorithm that accepts user input of number of classes (more than the number of feature classes), and minimum and maximum classes to be used. This algorithm then enables the computer to distinguish (with minimal supervision from the user) between feature classes. In this experiment, four feature classes are used: vegetation, water, wetlands and bare earth. This output is then further examined by the user to ensure accuracy and precision, after which it is ready for post-processing, which gives the desired final output.

2.3 ModelBuilder Workflow in ArcGIS

In order to ensure accuracy and efficency in repetition, a workflow is created using ModelBuilder in ArcGIS. ModeBuilder is a simple, visual programming language in ArcGIS that allows for automation of geoprocessing workflows. Due to a high amount of repetition of similar tasks over a 12-year period of study, it becomes imperative to automate these workflows. Figure 2.3 shows the workflow that was used specifically for this study.

2.4 Stage 1 Output

After the unsupervised image classification is done, the next step is post-processing reclassification to remove any noise and null values from the image. This classification is done using the *Reclassify* Spatial Analyst geoprocessing tool. This tool takes in unique numerical values corresponding to each feature class as examined by the user. The output from this step is an image of the study area with realistic colors distinguishing features for visualization purposes. More importantly, this output contains an attribute table with pixel count values that are converted to surface area in square miles based on the projected coordinate system used (Sabliov et al., 2002). These surface area values are then imported into Microsoft Excel to work toward the Stage 2 (final) output.

2.5 Stage 2 Output

At this stage the work within the ArcGIS space ends, while work begins in the Microsoft Excel space. In Excel, the surface area values are merged into a table with values representing Dry and Wet seasons in Florida's tropical climate, spanning over a period of 12 years from 2007 till 2018. In this table, the environment is prepared to carry out a Classic Multiplicative Time Series forecast Model. The attributes in each column include:

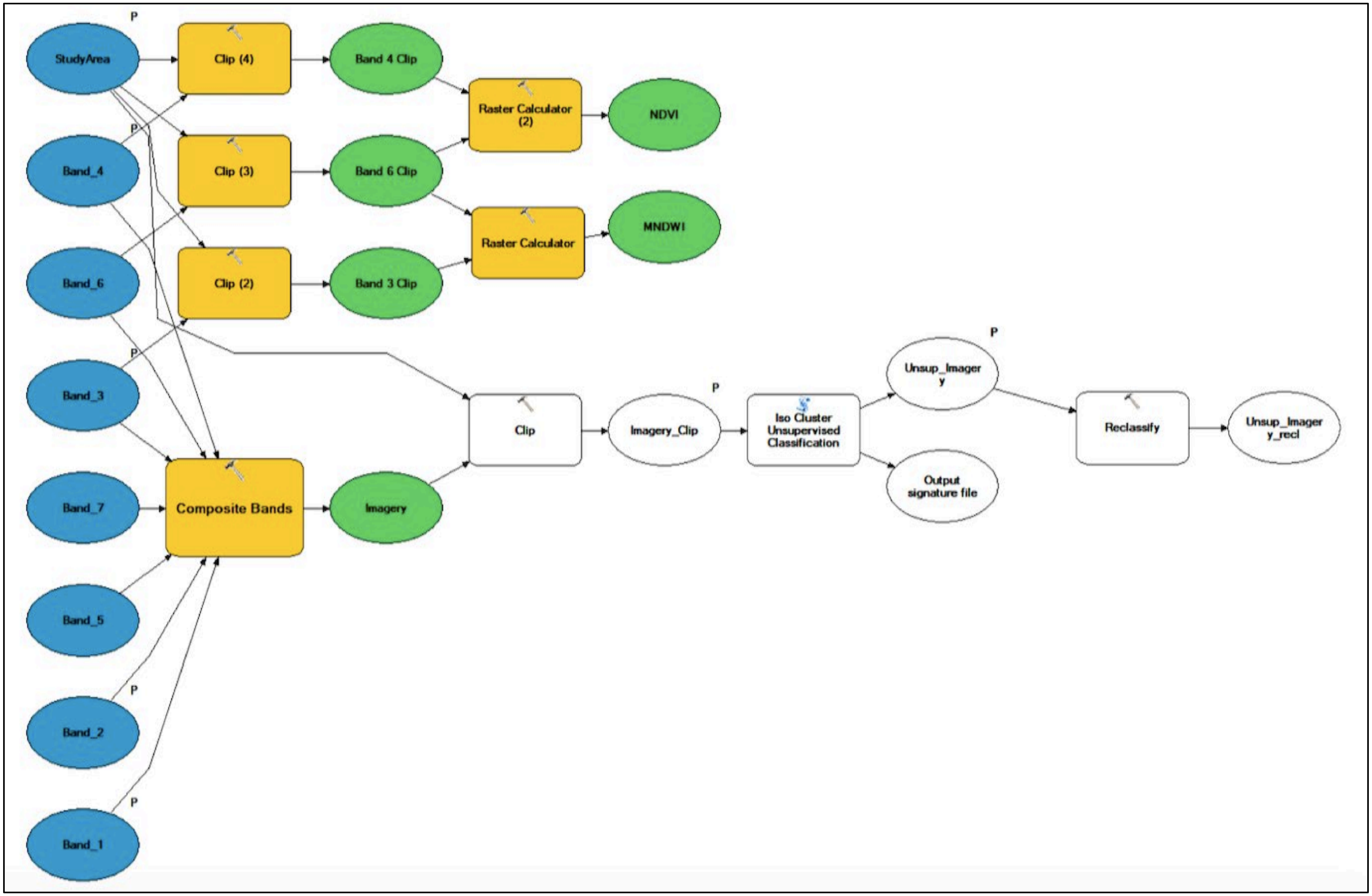

Figure 2.1: ModelBuilder illustration of workflow for each landsat image

- *Years*: Years in study ranging from 2007 – 2018.
- *Periods*: This refers to the number of seasons that data was collected for. In this case, data was collected for two periods for each year: Dry Season and Wet Season.
- *Surface Area:* Surface area values derived from pixel counts in ArcGIS.
- *Moving Average:* Smooths out fluctuations in the data to show a pattern or trend more clearly.
- *Centered Moving Average*: A more accurate representation of the moving average.
- *Seasonality and Irregularity*: A representation of how far away each data point with a combined degree of seasonality and irregularity strays from the centered moving average. For example, a value of 1.10 suggests that the data point is 10% above the centered moving average line.
- *Seasonality:* A representation of how far away each periodic component (dry and wet) is from the centered moving average.
- *Deseasonalize:* The process of removing the seasonality component from data. These values then undergo a polynomial regression to provide "predicted" values (not future values) of the general trend excluding seasonality and irregularity.
- *Trend:* The general state of changes in surface area, excluding the seasonality and irregularity component.
- *Forecast*: A representation of how well the selected model was able to match with the original data, excluding the irregularity component.
- *Future*: Using the forecasted values (containing trend and seasonality components) to predict future surface area trends.

Generally, irregularity is excluded from the model due to its difficult predictability. Hence, the model is a slightly simplified version of the real world and would therefore give a simplified forecast and future prediction. It is important to not predict too far into the future, as the absence of irregularity in the model would allow for a high level of uncertainty. As such, this study only predicts the average dry and wet surface area conditions for 2019.

This predictive model is used for the forecasting of water from Lake Okeechobee, including smaller bodies of water within the lake's boundaries. The values of the actual surface area data (shown in the results section) are averaged from a simple random sample over the 12 years.

The results section will illustrate the surface area plotted with necessary parameters from time-series forecasting.

CHAPTER 3

RESULTS

3.1. Water Surface Area

The values of surface area over the 12-year period show a decrease from 2007 – 2008. From 2008 until 2014, there is a high level of periodicity, indicating fairly consistent variations in wet and dry season surface areas during that time. However, there was a spike in water surface area in 2015, increasing from 480 sq. mi to 520 sq. mi. This large water surface area has remained relatively large since 2015, with a steady increase noticeable from the start of 2017. Here is a graphical illustration of these water surface area changes:

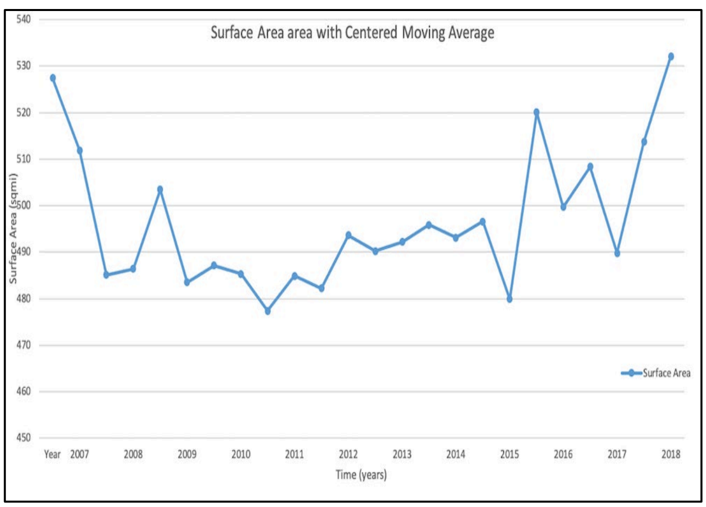

Figure 3.1: Change in water surface area with time

The next set of figures in this section illustrate the process of Time-Series Forecasting for Water Surface Area, starting with the calculation of the Centered Moving Average and ending with the predictive forecast.

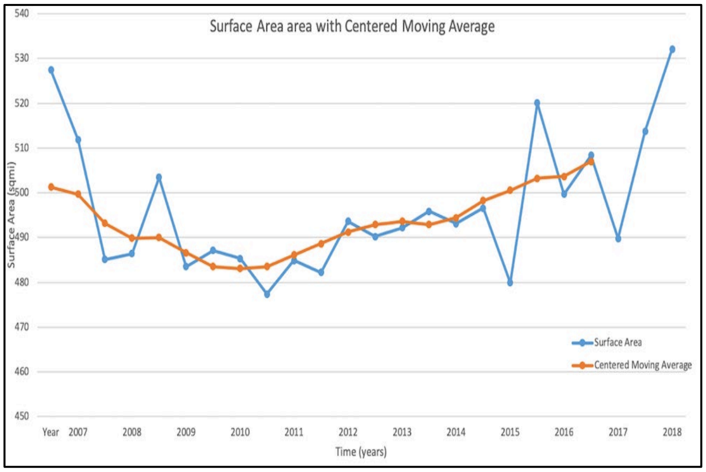

Figure 3.2: Change in water surface area and moving average with time

A sixth order polynomial regression analysis is conducted on deseasonalized data in order for the trend to be derived. This regression analysis is done with 95% confidence. Shown below are the key outputs for that analysis:

Multiple R	0.846428734
R Square	0.716441601
Adjusted R Square	0.616362167
Standard Error	9.198972386
Observations	24

Table 3.1: Regression statistics for analysis

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	561.108331	23.62836963	23.74723012	1.77768E-14
t	-47.4317579	23.33448168	-2.032689586	0.058007128

Table 3.2: Validation for acceptance or rejection of independent variable

The most important output in the two tables above is the P-value. Because the confidence level is 95% (0.95), the p-value required for this analysis to be acceptable is one that is less than or approximately equal to 0.05. The focus here is on t (time), which is the independent variable. The value of 0.058 is close enough to 0.05. Hence the regression analysis to "predict" the general trend for surface area changes is fairly acceptable. This is further buttressed by the high R Square value, which signifies how well the independent variable (time) would predict the dependent variable (surface area). Therefore, it can be stated that time is able to predict 71.6% of the changes in surface area. This is not always the case, however, as one would expect that time could not single-handedly predict changes in surface area with the best accuracy. Other factors such as temperature and atmospheric pressure would play a significant part. However, given the purpose of this study (to evaluate climate conditions over time), time is used as a single independent variable.

Shown next is the general trend derived from regression analysis:

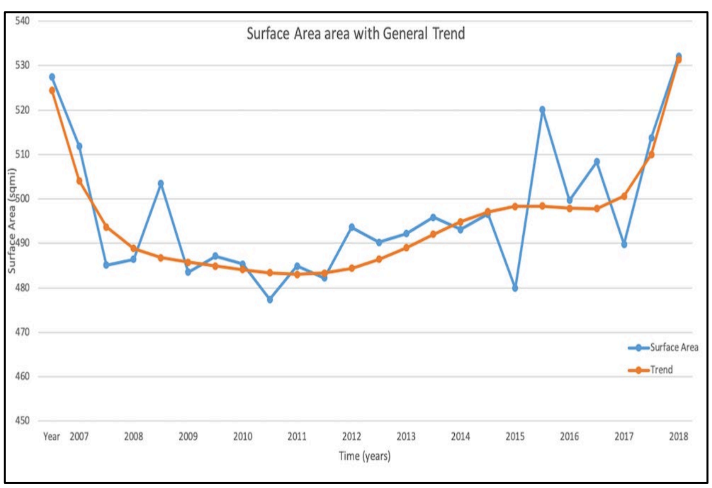

Figure 3.3: Change in surface area and the general trend of the data with time

The next figure shows a forecast created by inserting a seasonality component to the trend. This is done by multiplying the averaged-out seasonality (periodicity) in dry and wet seasons by the trend component (multiplication is done because the model used is the classical multiplicative model).

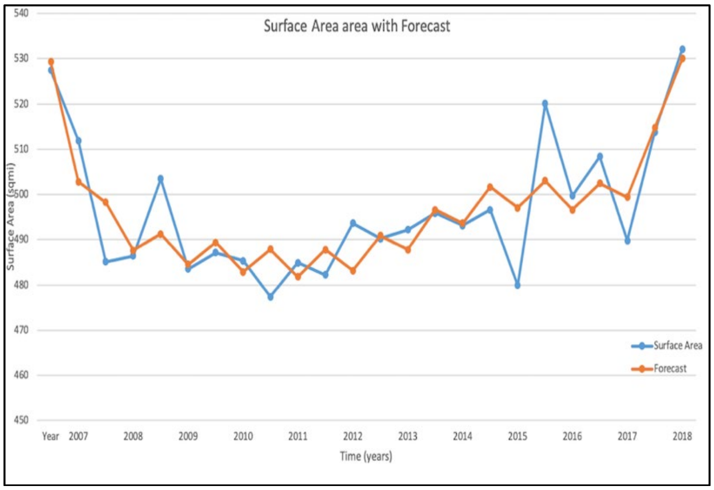

Figure 3.4: Change in surface area and forecast of the data with time (no prediction yet)

Finally, the forecast is used to predict surface area values for 2019 by utilizing the seasonality and trend values as shown in Table 3.1.

Attached next is the forecasting chart, including the 2019 forecast:

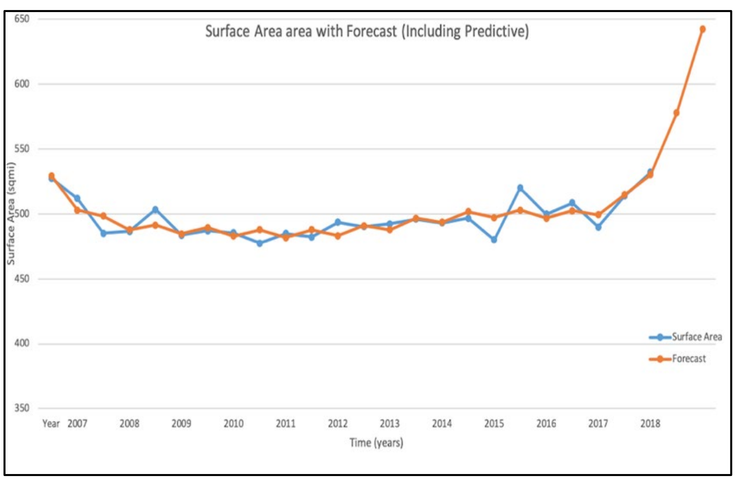

Figure 3.5: Change in surface area and forecast of the data with time (includes prediction)

Shown next is the table for Water Surface Area Forecasting. This table includes all the attributes necessary for a Time Series Forecast using the Classical Multiplicative Model, as well as the predicted forecasts highlighted in green:

t	Year	Period	Surface Area	Moving Average	Centered Moving Average	Seasonality & Irregularity	Seasonality	Deseasonalize	Trend	Forecast
1	2007	Dry	527.46				1.009	522.64	524.47	529.32
2		Wet	511.90	519.68			0.997	513.19	504.10	502.83
3	2008	Dry	485.13	498.51			1.009	480.69	493.70	498.25
4		Wet	486.44	485.78	501.33	0.9703	0.997	487.67	488.87	487.64
5	2009	Dry	503.48	494.96	499.73	1.0075	1.009	498.87	486.78	491.28
6		Wet	483.55	493.51	493.19	0.9804	0.997	484.77	485.74	484.52
7	2010	Dry	487.11	485.33	489.89	0.9943	1.009	482.65	484.93	489.41
8		Wet	485.30	486.20	490.00	0.9904	0.997	486.53	484.10	482.88
9	2011	Dry	477.41	481.35	486.60	0.9811	1.009	473.04	483.37	487.84
10		Wet	484.93	481.17	483.51	1.0029	0.997	486.15	483.01	481.79
11	2012	Dry	482.26	483.59	483.08	0.9983	1.009	477.85	483.29	487.76
12		Wet	493.63	487.94	483.51	1.0209	0.997	494.88	484.41	483.19
13	2013	Dry	490.26	491.94	486.16	1.0084	1.009	485.77	486.38	490.87
14		Wet	492.21	491.23	488.68	1.0072	0.997	493.46	489.04	487.80
15	2014	Dry	495.85	494.03	491.29	1.0093	1.009	491.32	492.02	496.56
16		Wet	493.13	494.49	492.92	1.0004	0.997	494.37	494.87	493.62
17	2015	Dry	496.58	494.85	493.65	1.0059	1.009	492.04	497.09	501.68
18		Wet	479.94	488.26	492.91	0.9737	0.997	481.15	498.30	497.04
19	2016	Dry	520.10	500.02	494.40	1.0520	1.009	515.34	498.41	503.02
20		Wet	499.73	509.91	498.26	1.0029	0.997	500.99	497.87	496.61
21	2017	Dry	508.34	504.03	500.56	1.0155	1.009	503.69	497.86	502.45
22		Wet	489.83	499.09	503.26	0.9733	0.997	491.07	500.67	499.40
23	2018	Dry	513.75	501.79	503.71	1.0199	1.009	509.05	510.01	514.71
24		Wet	532.14	522.95	506.96	1.0497	0.997	533.49	531.37	530.03
25	2019	Dry					1.009		572.52	577.81
26		Wet					0.997		643.88	642.26

Table 3.3: Table containing fields and values used for Time-Series forecasting

CHAPTER 4

DISCUSSION

4.1 Summary

The decrease in surface area of Lake Okeechobee from 2007 – 2008 would indicate a significant lack of rainfall during that period. Evidently, this is a good marker for the already documented severe drought conditions that were present in South Florida from July 2007 until April 2008 (James, 2016). Interestingly, that drought began in July, which is close to the middle of Florida's wet season (May – October). However, the reverse is projected to be the case for 2019. From the results section, it can be noticed that the average wet season forecast for Lake Okeechobee surface area is approximately 642 sq km. (with a standard error of 23 sq km). This is just over 85% of the total surface area of Lake Okeechobee (including wetlands, surrounding vegetation and bare earth). This is higher than the average for any other wet season from 2007 till 2018. It is also noticed that the dry season (November – April) also records a relatively large surface area.

4.2 Conclusion

What does this mean for South Florida? Seemingly bad news. An already heavily flooded South coast looks to be even more flooded. This means that the stakeholders need to find a way to significantly improve flood management in the area. A suggestion would be to reverse the processes that allowed for an increased risk of flooding in the first place. Canal levels and groundwater levels can be dropped to accommodate the flooding. However, water supply to the population also needs to be taken into account. It would require a team of highly skilled Hydrogeologists, Environmental Scientists and Engineers to tackle these alarming issues.

REFERENCES

Czajkowski, J., Engel, V., Martinez, C., Mirchi, A., Watkins, D., Sukop, M. C., & Hughes, J. D. (2018). Economic impacts of urban flooding in South Florida: Potential consequences of managing groundwater to prevent salt water intrusion. *Science of The Total Environment*, *621*, 465–478. http://doi.org/10.1016/j.scitotenv.2017.10.251

Ganguli, P., & Ganguly, A. R. (2016). Space-time trends in U.S. meteorological droughts. *Journal of Hydrology: Regional Studies*, *8*, 235–259. http://doi.org/10.1016/j.ejrh.2016.09.004

James, R. T. (2016). Recalibration of the Lake Okeechobee Water Quality Model (LOWQM) to extreme hydro-meteorological events. *Ecological Modelling*, *325*, 71–83. http://doi.org/10.1016/j.ecolmodel.2016.01.007

Rozenstein, O., & Karnieli, A. (2011). Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. *Applied Geography*, *31*(2), 533–544. http://doi.org/10.1016/j.apgeog.2010.11.00

Sabliov, C. M., Boldor, D., Keener, K. M., & Farkas, B. E. (2002). Image Processing Method To Determine Surface Area And Volume Of Axi-Symmetric Agricultural Products. *International Journal of Food Properties*, *5*(3), 641–653. http://doi.org/10.1081/jfp-120015498

Sarp, G., & Ozcelik, M. (2017). Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey. *Journal of Taibah University for Science*, *11*(3), 381–391. http://doi.org/10.1016/j.jtusci.2016.04.005

Schottler, S. P., & Engstrom, D. R. (2006). A chronological assessment of Lake Okeechobee (Florida) sediments using multiple dating markers. *Journal of Paleolimnology*, *36*(1), 19–36. http://doi.org/10.1007/s10933-006-0007-5

Smith, S. M., Gawlik, D. E., Rutchey, K., Crozier, G. E., & Gray, S. (2003). Assessing drought-related ecological risk in the Florida Everglades. *Journal of Environmental Management*, *68*(4), 355–366. http://doi.org/10.1016/s0301-4797(03)00102-6

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. http://doi.org/10.1080/01431160600589179

BIOGRAPHICAL INFORMATION

Ohimai Imoukhuede was born and raised in Lagos, Nigeria. He lived in Nigeria until he completed his secondary education there. Ohimai then moved over to the United States of America for university education, where he pursued an Honors Bachelor of Science in Geology (Engineering Option) with a minor in Business Administration. Ohimai is an independent learner and is interested in pursuing a career in Data Analytics, starting use cases in Earth and Environmental Sciences, and then moving on to Business Analytics.