Lizzy Compton, Terra Ficke
Dr. Long
MAT 375
2/16/2018

Evaluation of Average Monthly Surface Temperature in Kouma-Konda, Togo

Background of Kouma-Konda:

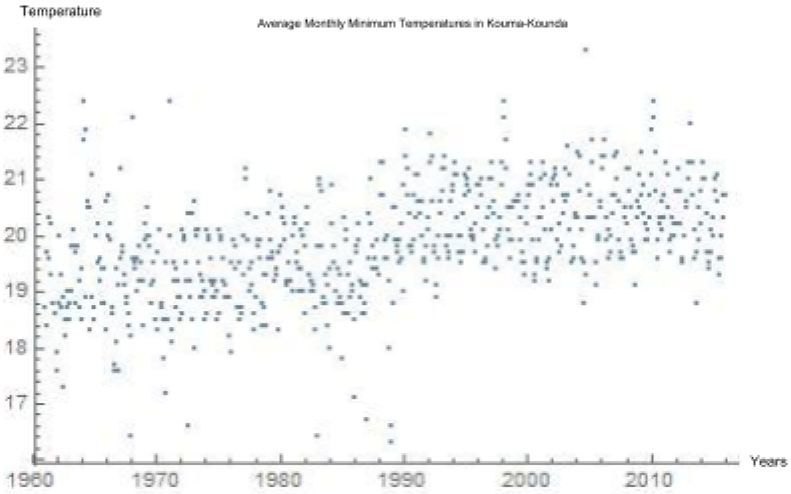
As you can see in the maps pictured to the side, Kouma-Konda is located in the South-Western part of Togo. It is a small village on the border. Because it is located in the middle of a beautiful green and lush region, it is a big spot for tourism, though Kouma-Konda has a population of just 550 persons. In current years, the forests have been subject to deforestation due to bushfires and also to make room for farmland. Rain is abundant from May to June and sometimes torrential from August to September. This makes sense since it located within a rainforest. December and January are the coolest months of the year. The altitude is 650 meters above sea level. It’s latitude and longitude in degrees is (6.950,+000.583).

Initial Report:

In the initial report for Kouma-Konda, there was a concerned about a potential outlier in 1964. With our new and improved data (now having monthly data as opposed to yearly) we can see that there are multiple data points that seem concerning, not just in 1964, but in other years as well. We have decided to throw out our data points that we think could possibly be a mistake and are throwing off our fitted models. A significant increase over time for the minimum temperatures was also found while no significant change in the maximum temperatures was seen. The calculations we did with our new data only further backed their claim (Dufec and May).

Monthly Minimum Temperatures:

The graph to the side shows the average monthly minimum temperatures (measured in degrees Celsius) plotted against the time (measured in years) for Kouma-Konda from January 1961 through December of 2015. Looking at the scatterplot, you can see that the overall trend of the data is increasing, however there is



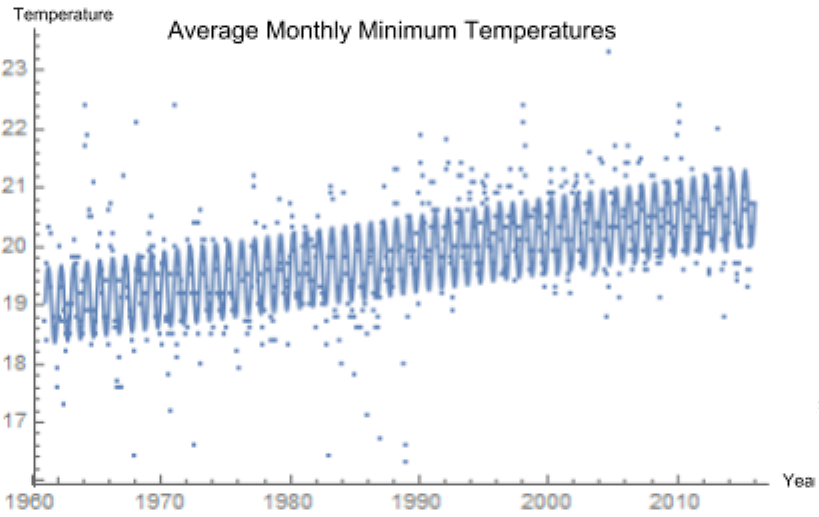
much variation throughout. To fully determine if there was a significant relationship between the minimum temperatures and time we ran our data through an ANOVA regression analysis in Minitab.

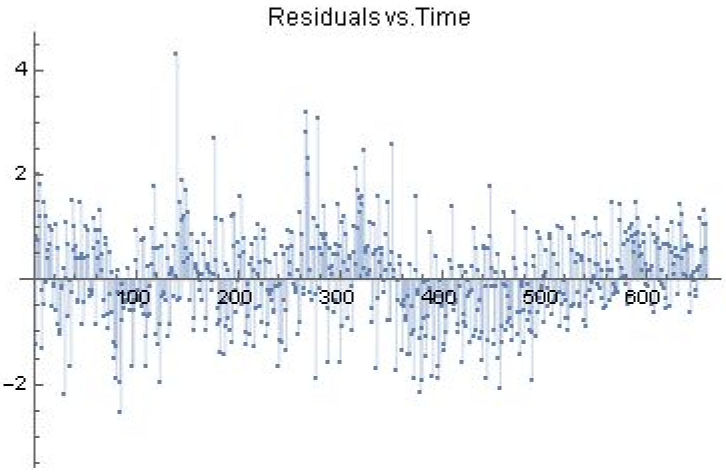
Analysis of Variance					
Source	DF	SS	MS	F-value	P-value
Regression	12	333.9	27.822	62.36	0.000
Year	1	157.1	157.104	352.13	0.000
Month	11	176.8	16.069	36.02	0.000
Error	647	288.7	.446		
Total	659	622.5			

From these calculations, it is clear that the year and the month are significant factors in determining the average monthly minimum temperatures. We now want to determine the specific relationship between the two factors and the temperatures. To do this we will use the LinearModelFit function in Mathematica to determine the best model for our data.

The linear model for the data is:

Temperature(t)= -41.3202 + 0.0307556x + 0.0682634Cos[2(pi)x] + 0.64454Sin[2(pi)x]





This model further shows that there is a significant increase in the average monthly minimum temperatures as the confidence interval for the linear term is both significant and positive.

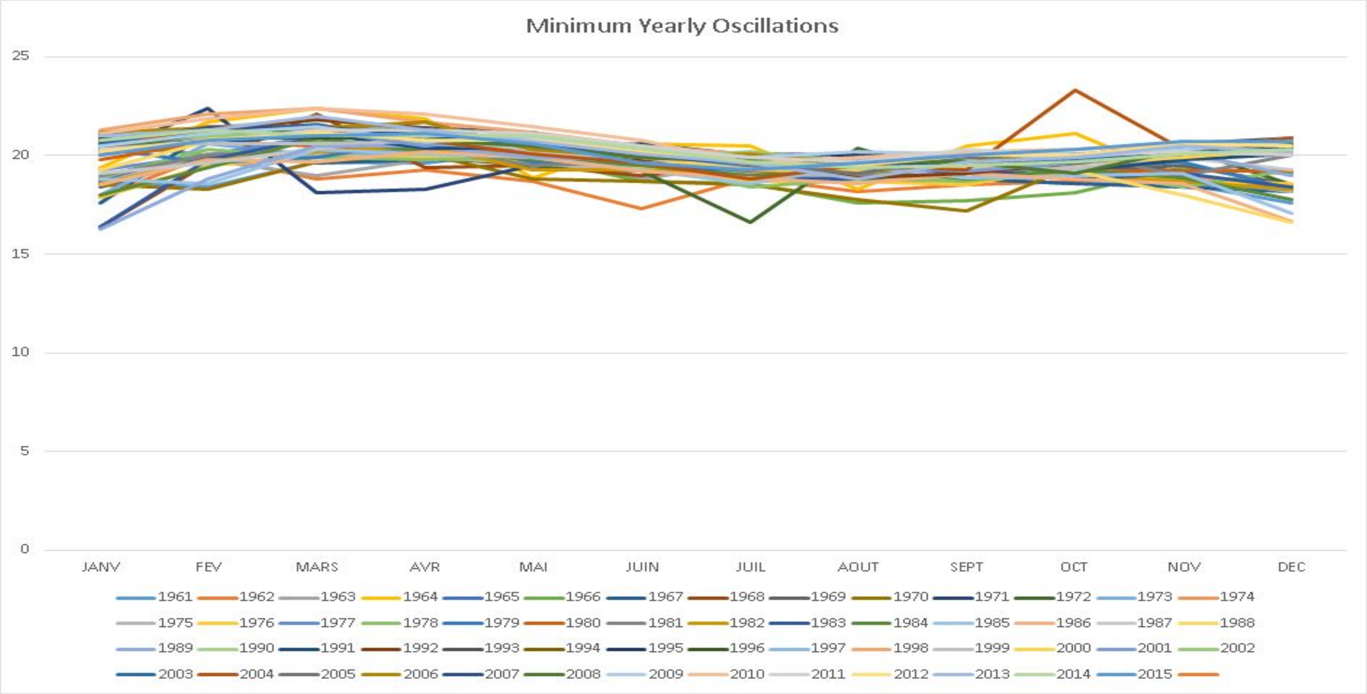
Parameter Confidence Intervals:
Intercept (a): {-48.1182, -34.5223}
Linear Term (b): {0.0273371, 0.0341741}
Sin Coefficient (α): {0.567782, 0.721299}
Cos Coefficient (β): {-0.00848698, 0.145014}

The linear term is significant and with 95% confidence we estimate that for each additional year, the average monthly minimum surface temperatures in Kouma-Konda increases by between .0273371 and

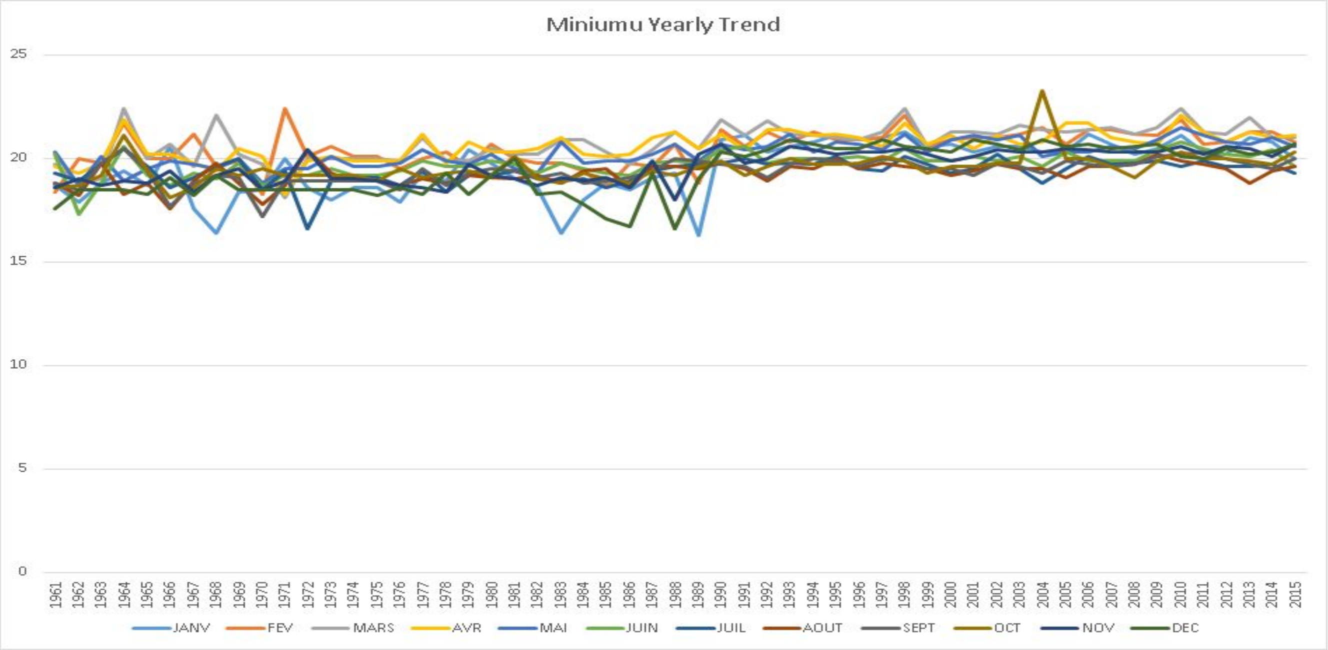
.0341741 degrees Celsius. This indicates that there is a positive linear relationship between the time and average monthly temperatures. The sin term parameter is also significant and with 95% confidence, we estimate the Sin coefficient is between .567782 and .721299. This coefficient indicates that there is a significant yearly oscillation within the data. Within each year our data supports that the temperature increases from October through April, and from April to October the monthly temperature decreases. After further researching about Togo as a whole, this claim seems very plausible and probable (Climate-Togo).

While the model and the parameter confidence intervals are indicating that there is a significant linear relationship here, the R-Squared term for this model is only .468721. That means that only 46.87% of the variation in the data is explained by the linear model. This can be classified as a weak to moderate relationship. Additionally, looking at the plot of the residuals, there appears to be a positive and negative pattern to them. Also there are a few residuals much higher than those around them, indicating that these may be outlier points. Because of this apparent lack of fit to the model, we looked into the quadratic model to see if it would be a better fit. However, when we calculated the LinearModelFit for the quadratic model, we found that only the Sin coefficient was significant, which means that this model only accounts for the oscillation in the data, not for the overall trend.

After completing these analyses, it is easy to see that the linear model is not fitting the data all that well. One of the reasons for this is because there appears to be quite a few outliers in the data. When we plot the temperature vs. month for each year (in Excel) we can clearly see the yearly oscillation pattern in the data and there are quite a few outliers that stick out.



Additionally when we plot the temperature vs. years for each month (again in Excel), we can see the overall increasing trend of the model, and we see even more outliers present.



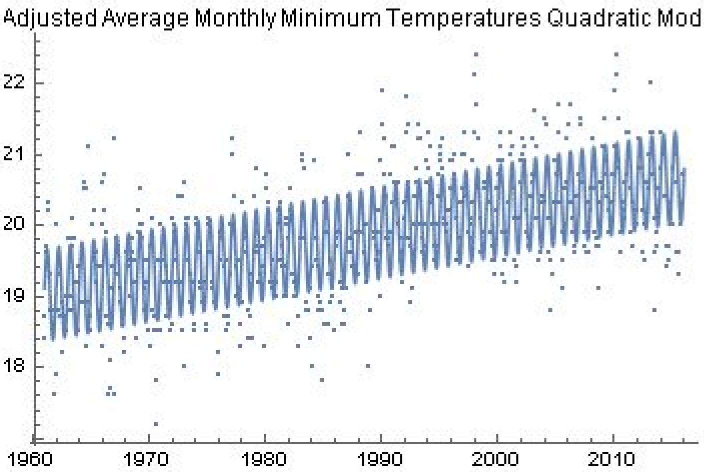
Because of there are so many data points that are obvious outliers, we have decided to remove these points and rerun the analysis to determine the effect that they have on the model.

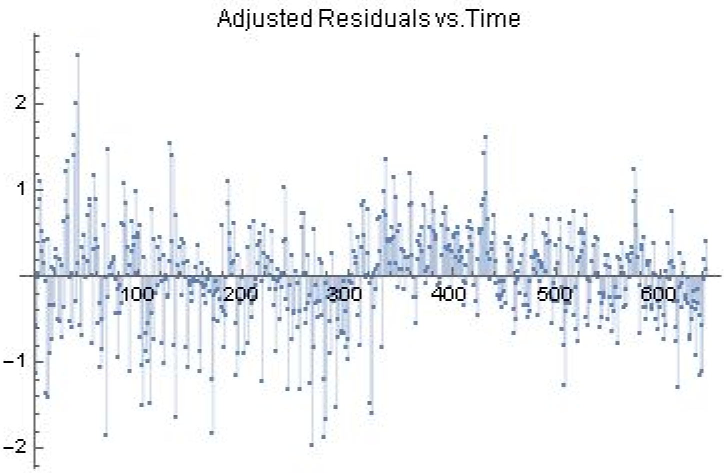
The outliers that we removed are:

{{Jun. 1962}, {Feb, Mar, Apr, 1964}, {Jan, Mar 1968}, {Feb, Mar, Apr 1971}, {Jul 1972}, {Jan 1983}, {Dec 1985}, {Dec 1986}, {Dec 1988}, {Jan 1989}, {Oct 2004}}

After making the adjustments in the data, our new linear model is:

Temperature(t) = -40.8326 + 0.0305195x + 0.097428Cos[2(pi)x] + 0.635158Sin[2(pi)x]





Parameter Confidence Intervals:

Intercept (a): {-46.4293, -35.2359}

Linear Term (b): {0.0277055, 0.0333335}

Sin Coefficient (α): {0.572202, 0.698114}

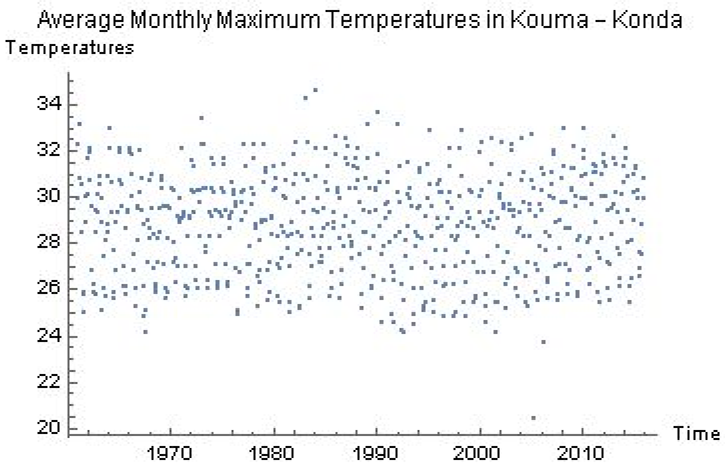
Cos Coefficient (β): {0.0343033, 0.160553}

All parameters are significant within the adjusted model. With 95% confidence, we estimate that for each additional year, the average monthly minimum surface temperatures in Kouma-Konda increases by between .0277055 and .0333335 degrees Celsius. Additionally, the R-Squared term for the model has improved to .573599 by adjusting the data. This indicates that 57.3599% of the variation in the data is accounted for by the model. This can be classified as a moderate linear relationship. While, we see improvements through these aspects, the residual plot actual for the adjusted data actually looks to have more pattern in it that the first residual plot indicating non-constant variance. This is a concern for our model but since when we looked into the quadratic model for the data we found that it was not significant, the linear model will stand as the most appropriate model for the average minimum monthly data.

Although this model using sine and cosine did give us a linear increase and had significant parameters, it a still a relatively weak model. Hopefully the next group can find a model that accounts for more than 57.3599% of the variation. This model made us happier than the quadratic and the original model that still included the outliers, but fitting this data can still be improved upon.

Maximum Monthly Temperatures:

The graph to the side shows the average monthly maximum temperatures (measured in degrees Celsius) plotted against the time (measured in years) for Kouma-Konda from January 1961 through December of 2015. Looking at the scatterplot, you can see that there is an overall flat trend, indicating that the temperatures are neither increasing nor decreasing. To fully determine if there was a significant relationship between the minimum temperatures and time we ran our data through an ANOVA regression analysis in Minitab.



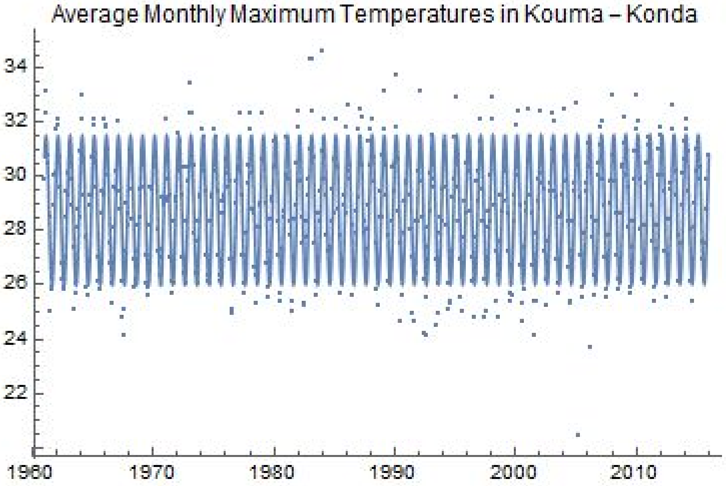
Analysis of Variance					
Source	DF	SS	MS	F-Value	P-Value
Regression	12	2613.88	217.824	258.17	0.000
Year	1	0.13	0.131	0.16	0.694
Month	11	2613.75	237.614	281.62	0.000
Error	647	545.89	0.844		
Total	659	3159.77			

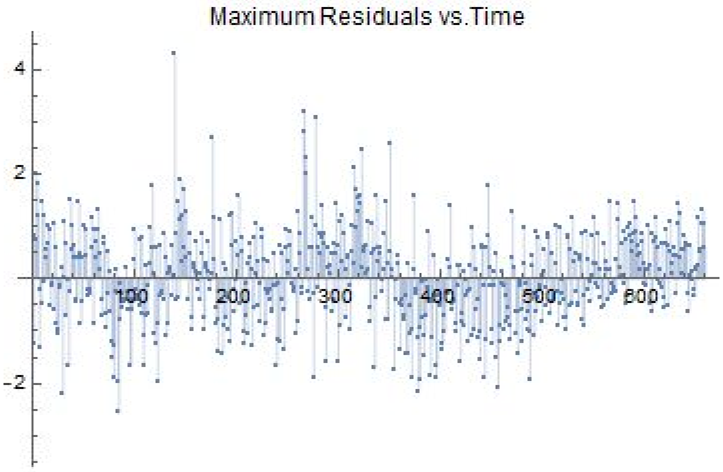
From these calculations we see that only the month is significant factor in determining the average monthly maximum temperatures. To determine the relationship between the months and the temperatures we will use the LinearModelFit function in Mathematica to determine the best model for our data which will give us the oscillation sin and cos coefficients. .

The linear model for the data is:

Temperature(t) = 27.0101 + .000883214x + 1.98763Cos[2(pi)x] + 1.892Sin[2(pi)x]

Parameter Confidence Intervals:
Intercept (a): {17.2985, 36.7217}
Linear Term (b): {-0.0040005, 0.00576692}
Sin Coefficient (α): {1.78234, 2.00165}
Cos Coefficient (β): {1.87798, 2.09727}





While the linear term is not significant, both the Sin and Cos coefficients are. With 95% confidence, we estimate the Sin coefficient is between 1.78234, 2.00165 and with 95% confidence, we estimate the Cos coefficient is between 1.87798, 2.09727. These coefficients indicate that there is a significant yearly oscillation within the data. Each year April tends to be the warmest month much like the minimum temperatures, but unlike those, the maximums seem to have a smaller time increment of decreasing temperatures. It looks like they average their low for the year in about September instead of October. Although this may seem like that would mean we would have a positive relationship between time and temperature in our model, it does not. When the temperature decreases from April to September, it decreases to low enough temperatures that it cancels out the shorter time it takes to do so.

While the model and the parameter confidence intervals are indicating that there is a not significant linear relationship here, the R-Squared term for this model is .786378. That means that 78.6378% of the variation in the data is explained by the linear model and the oscillations from the sin and cos terms. This can be classified as a moderately strong relationship. However, looking at the plot of the residuals, there appears to be a positive and negative wave like pattern to them. There are a few residuals much higher than those around them, indicating that these may be outlier points. Even though, the linear term wasn't significant, we looked into the quadratic model to see if it would be a better fit.

When we calculated the LinearModelFit for the quadratic model, we found that all of the parameters were significant.

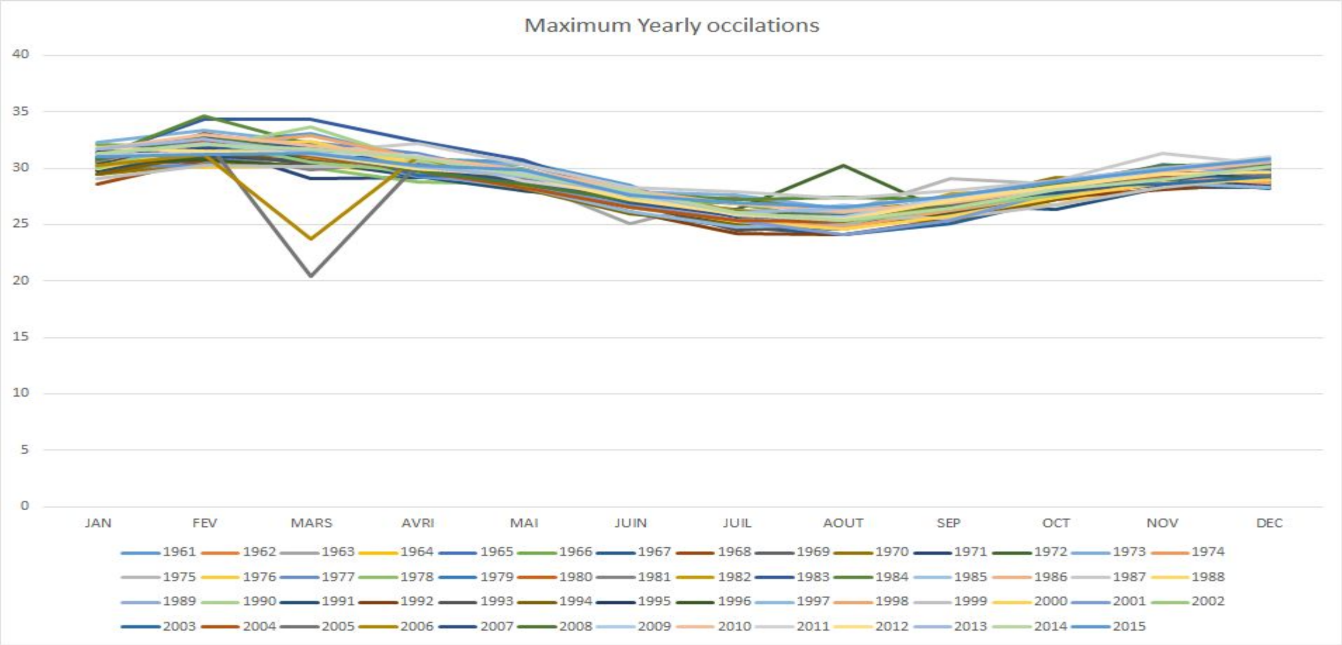
The quadratic model:

$$\text{Temperature}(t) = 1935.31 - 1.91857x + 0.000482639x^2 + 1.98758\text{Cos}[2(\text{pi})x] + 1.892\text{Sin}[2(\text{pi})x]$$

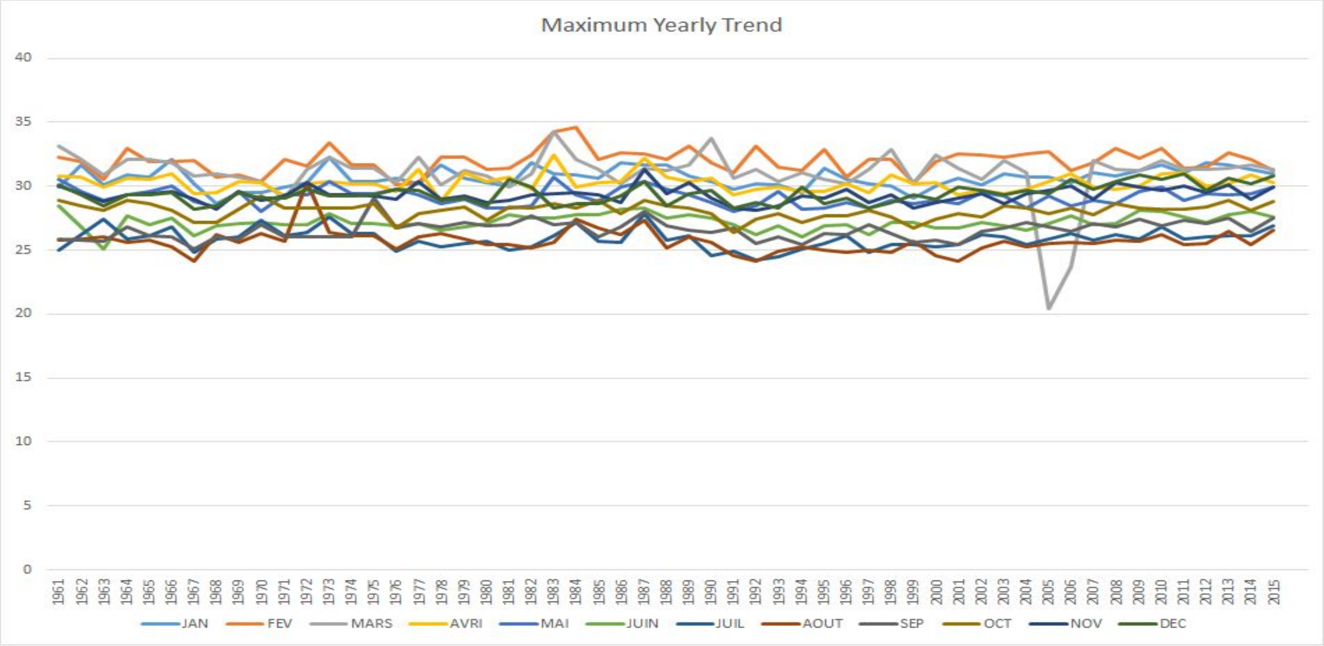
Parameter Confidence Intervals:
Intercept (a): {582.523, 3288.09}
Linear Term (b): {-3.27924, -0.557904}
Quadratic Term (c): {0.000140507, 0.000824771}
Sin Coefficient (α): {1.78289, 2.0011}
Cos Coefficient (β): {1.87849, 2.09667}

All of the parameters are significant in this model and the R-Squared term is .78852, indicating that 78.852% of the variation in the data is accounted for by the model. Looking at the residual plots, there still appears to be a slight wave pattern to the residuals, indicating non-constant variance throughout the model.

Again, after completing these analyses, we wanted to see if taking out the outlier points would affect the models. When we plot the temperature vs. months for each year (in Excel) we see a similar oscillation pattern that we saw in the minimum data as well as a few outlier points.



Additionally when we plot the temperature vs. years for each month (again in Excel), we can see the overall flat trend of the model and we see another clear outlier.



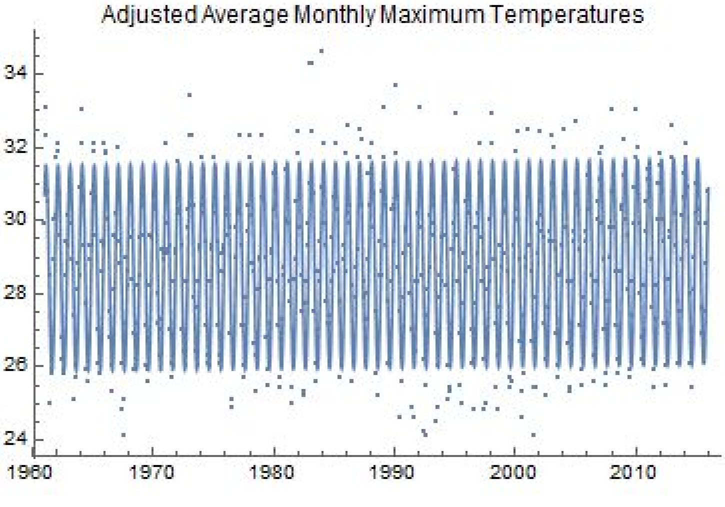
Just like in the minimum data, we have decided to remove these points and rerun the analysis to determine the effect that they have on the model.

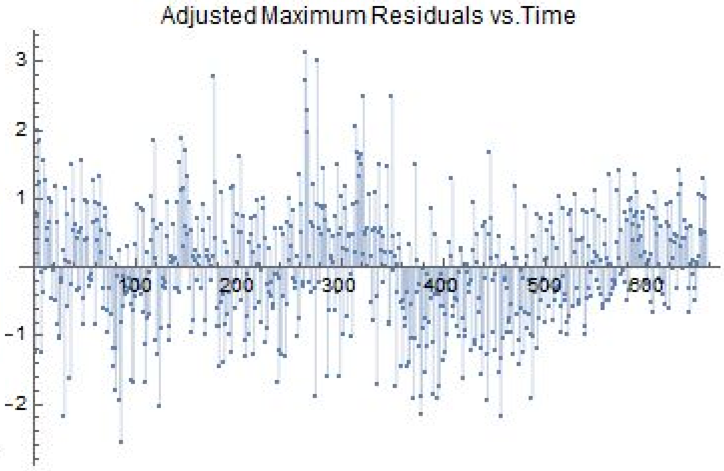
The outliers that we removed are:

{{Aug 1972}, {Mar 2005}, {2006 Mar}}

After making the adjustments in the data, our new linear model is:

Temperature(t) = 22.3539 + 0.00323549x + 2.01154Cos[2(pi)x] + 1.95598Sin[2(pi)x]





Parameter Confidence Intervals:

Intercept (a): {14.0922, 30.6155}

Linear Term (b): {-0.000919163, 0.00739015}

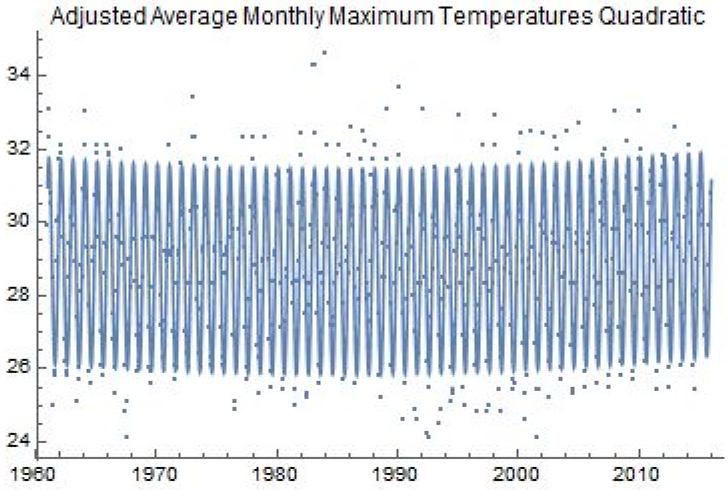
Sin Coefficient (α): {1.8626,2.04937}

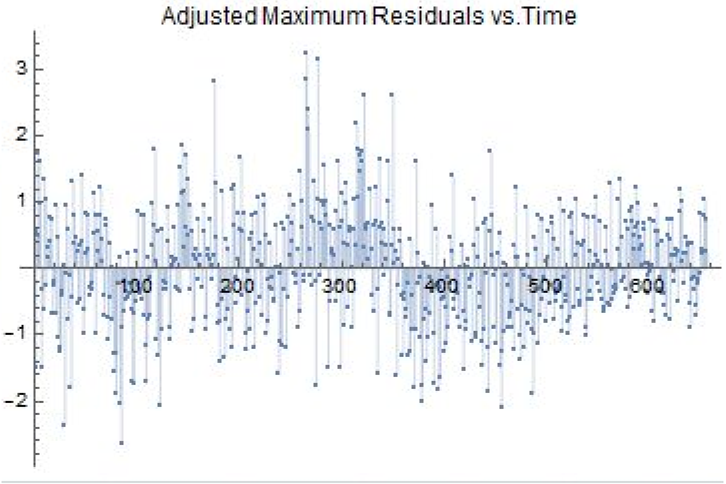
Cos Coefficient (β): {1.91842, 2.10466}

Once again, we see that the linear term is not significant in the model. The R-Squared term has increased though to .842012 indicating that more of the variation is now accounted for by the model. However, there still seems to be a wave pattern in the residual plot indicating non-constant variance. Because of this, we will again look into the quadratic model to see how it has changed now that we have removed the outliers.

Adjusted quadratic model:

Temperature(t) = 2020.25 -2.00635x + 0.000505303x^2 + 2.01153Cos[2(pi)x] + 1.95612Sin[2(pi)x]





Parameter Confidence Intervals:

Intercept (a): {582.523, 3288.09}

Linear Term (b): {-3.27924, -0.557904}

Quadratic Term (c): {0.000140507, 0.000824771}

Sin Coefficient (α): {1.78289, 2.0011}

Cos Coefficient (β): {1.87849, 2.09667}

All of the parameters are significant in this model and the R-Squared term has again increased to now at .84481 since we have taken out the outliers, indicating that 84.481% of the variation in the data is accounted for by the model. This is considered a strong relationship. Looking at the residual plots though, there still appears to be a slight wave pattern to the residuals indicating non-constant variance. This is something that we will need to ask the Togolese meteorologists about. We'd like to see if their data collection has any effect on this, for this has been a constant problem throughout all our models.

The adjusted quadratic model for our maximum temperatures is by far the best option. It has a high percentage of accounted for variation and complete significant parameters. Although the linear term shows a negativity, we can see that it's not important enough to consider the maximum temperatures to be decreasing over time. Also, the quadratic term is positive. And although that may seem interesting, the parameter is so low that it doesn't mean much. In this model there is higher cosine and sin values, so we know that there is definite, meaningful oscillation. Yet, over time for our maximum values there does not seem to be neither a significant increase or decrease in temperatures.

Additional Comments and Questions:

There are still many questions we have for the Togolese meteorologists. Such as questions about where they took their temperatures, was is at the same location every time? Also, what's up with the outliers, are some of them mistakes on their part or are some legitimate? With the constant pattern we see in our residuals, there has to be something we are missing. Is there maybe some pattern that spans every few years? What could be causing this pattern? Could it be astronomical factors? Possibly something to do with the rainfall? Kouma-Konda *is* located within a rainforest. Or perhaps it is purely coincidental. Nevertheless, finding out more from the Togolese could help us understand the data better so that we could get a feel for the country's temperature fluctuation as a whole and more easily fit proper models. Moving forward, we need to figure out how to account for the residual patterns in order to account for more variation. Also, for our minimum data, we need to find a better model since none of the ones we fit accounted for as much as we might've hoped. Perhaps a non-linear model would be more appropriate? It will be very interesting to see what strides the next group can make for this data. It is already interesting that there is an increase in minimum temperatures in almost every town.

Citations:

"Climate - Togo." *Climate in Togo: temperature, precipitation, when to go, what to pack*, www.climatestotravel.com/climate/togo

Dufec, S. and May, A. (2018). *Mini Project 1*.

"Missions au Togo - Kouma Konda." *Missions Togo Kouma Konda - santé, education, environnement... | Urgence Afrique*, urgenceafrique.org/fr/mission-togo-kouma-konda.

"Posts about Kouma Konda village on Pogo in Togo." *Pogo in Togo*, pogointogo.wordpress.com/tag/kouma-konda-village/.

Weather Data for Kouma-Konda in Togo, geographic.org/global_weather/togo/kouma_konda_653780_99999.html.