MAT 375 Mini-Project 2: Modeling Temperature for Sotouboua, Togo

Jacob Englert and Adam May

Introduction

Sotouboua, Togo

Sotouboua is a town located in Togo, a Western African nation that gained its independence from France in April 1960. Geographically, Sotouboua is near the center of Togo both in terms of longitude and latitude. It is considered a mid-sized town at a population of approximately 21,000. Sotouboua is relatively flat, with an elevation of 650 feet above sea level, and is about average when it comes to rainfall among its Togolese counterparts, receiving an annual average of 1,318.5 mm of rain.

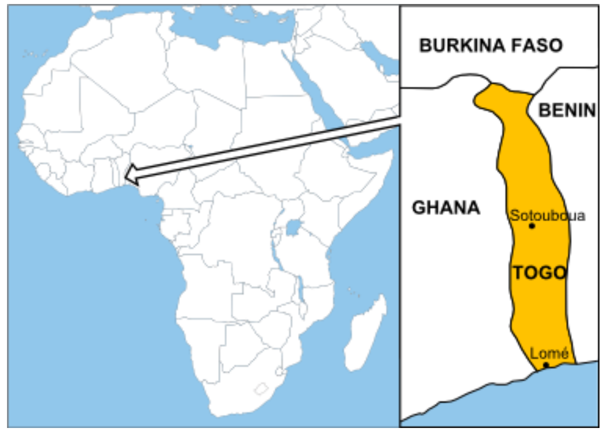


Figure 1: Map of Togo and Surrounding Nations

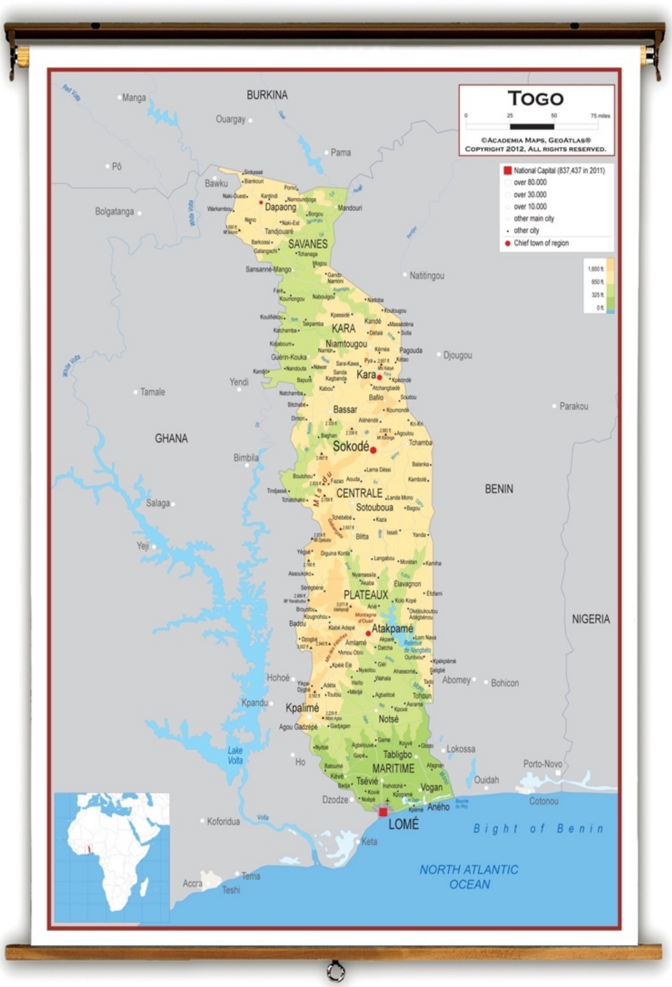


Figure 2: Contour Map of Togo

Problem

We have been tasked with the job of determining whether Togolese temperatures have experienced a significant increase over time, specifically in the town of Sotouboua. Large increases in temperature have the potential to not only wreck a nation's economy and ecosystems, but also to cause long-lasting climate problems, such as rising sea levels.

Data

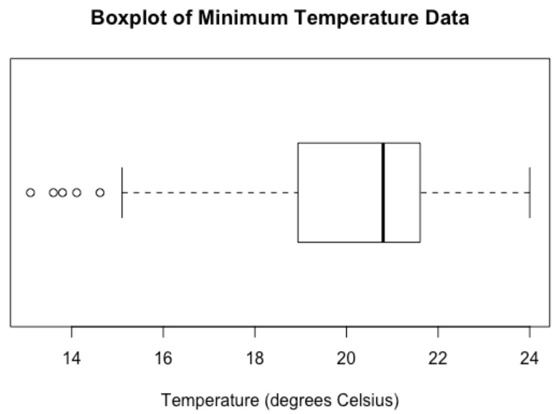
The data provided by the Togolese government consists of temperatures for ten different cities in Togo, but we will solely focus on the data concerning Sotouboua. Both minimum and maximum temperatures were recorded every month beginning with 1982 and extending to 2015. For the sake of this report, these recordings are assumed to have happened at the half-way point each month. Correct annual means were also calculated using these values. A few models using these annual averages were produced in a paper written by Alyssa Farmer and Patrick Nielsen. The mathematical modeling detailed in this report was performed in R utilizing the monthly data.

There are a few missing data points in the dataset we were given. In both the minimum and maximum temperature data, values are missing for January and February of 2013. For the purpose of constructing models we have omitted these observations.

While the maximum temperature data is free of outliers, the minimum temperature data contains several outliers. To detect these outliers, a boxplot of the data was created in R. The five observations that were marked as outliers were all abnormally low. These values have been excluded from the models discussed in this report. The values, the month and year the measurements were taken, and the boxplot can be seen here:

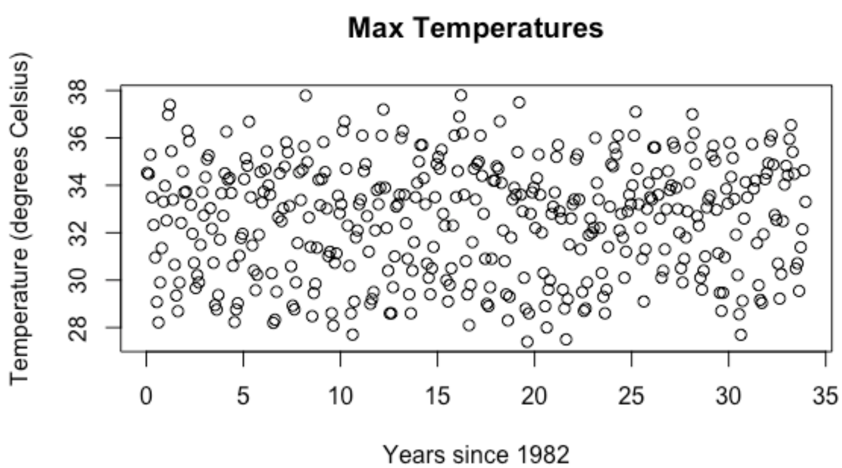
Minimum Data (mean = 20.171°C)

Outlier Month	Outlier Value
December 1986	14.61935°C
January 1989	14.10968°C
December 2003	13.1°C
January 2004	13.8°C
December 2006	13.6°C



Model Building

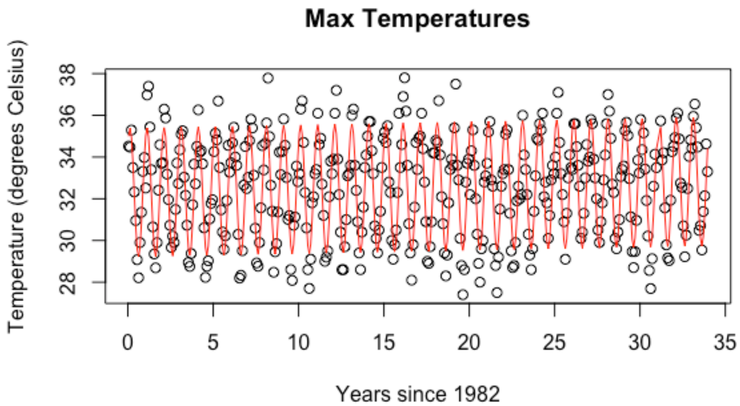
Maximum Temperatures



Above is a scatterplot of the maximum temperature data. Each of the 406 months with non-missing data is plotted. The plot appears to show relatively random scatter. In fact, in prior research focusing on the annual maximum temperatures (Alyssa and Patrick), no real linear relationship was found between times and temperature. However, with monthly data, we were able to construct models that account for the seasonal variation and include significant explanatory terms. These models are detailed in the next section.

Models

Linear Oscillating Model



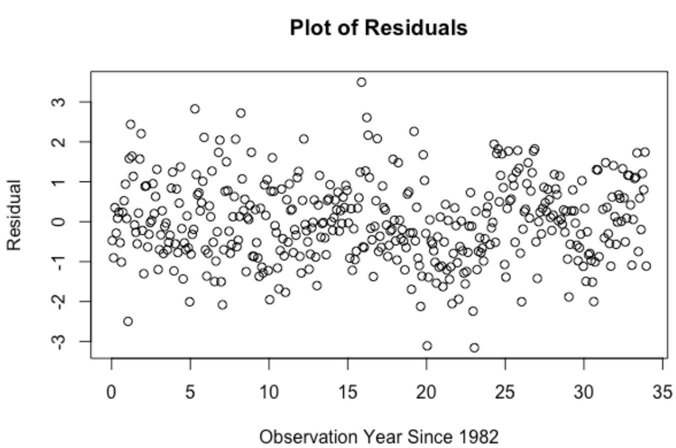
The first model constructed for the maximum temperature data contains terms for the year (scaled to start at 0), as well as a combination of sine and cosine terms using a period of one year to capture the seasonal oscillation in the data. All three of the terms are significant in the prediction of Sotouboua monthly maximum temperatures, as can be seen in the table of parameters below:

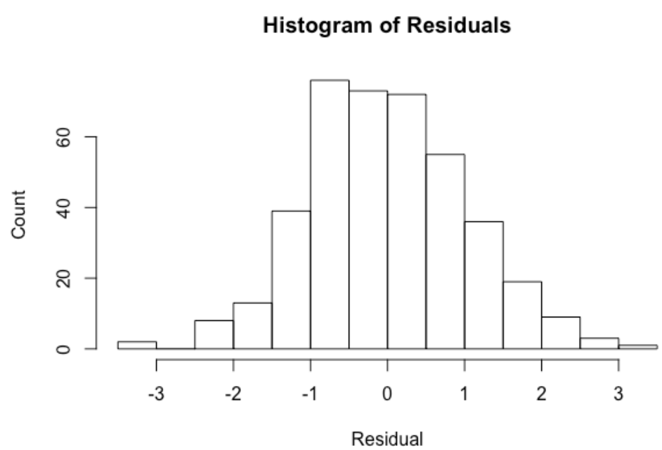
	Estimate	Standard Error	t-Statistic	P-Value	95% CI
Constant	32.296293	0.101253	318.966	< 2e-16***	(32.097241, 32.495344)
t (years)	0.015629	0.005178	3.018	0.00271***	(0.005449, 0.025808)
Sin(2πt)	2.135572	0.071608	29.823	< 2e-16***	(1.994798, 2.276345)
Cos(2πt)	2.220925	0.071736	30.960	< 2e-16***	(2.079901, 2.361949)

*p-value < 0.05 **p-value < 0.01 ***p-value < 0.001

This model performs very well, as it found that 82.10% of the variation in monthly maximum temperatures for Sotouboua was explained by the relationship to time.

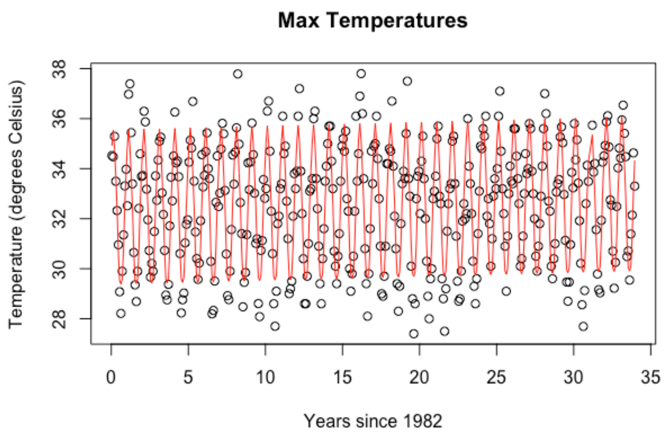
Residuals





The residual plots look very good for the maximum temperature data. We see relatively random scatter with constant variation in the plot of residuals. In the histogram of residuals, we see what appears to resemble a normal distribution, however a few values to extend past three standard deviations. After some digging, it was found that these observations resulting in poor fits were particularly low-max Januaries (2002 & 2005) on the negative side, and a particularly high-max November (1997) on the positive side. None of these observations would be considered outliers. In the plot of residuals, we see a bit of a "moustache" effect, which was also noticeable in the annual temperature analyses conducted by Patrick and Alyssa. The next model factors this fluctuation in.

Linear Oscillating Model (including 16-year oscillation)



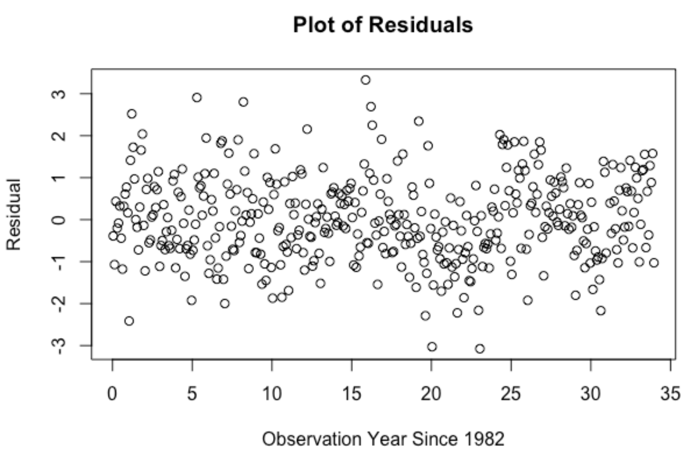
The first few terms in this model are identical to those in the first model. In an effort to capture the oscillation noted above, new sine and cosine terms were introduced, both using an estimate for the period of 16 years. However, the estimate for the sine term was not found to be significantly different from zero, and thus was dropped from the model. The cosine term was found to be significant at the 0.05 level. The new parameter values are shown below:

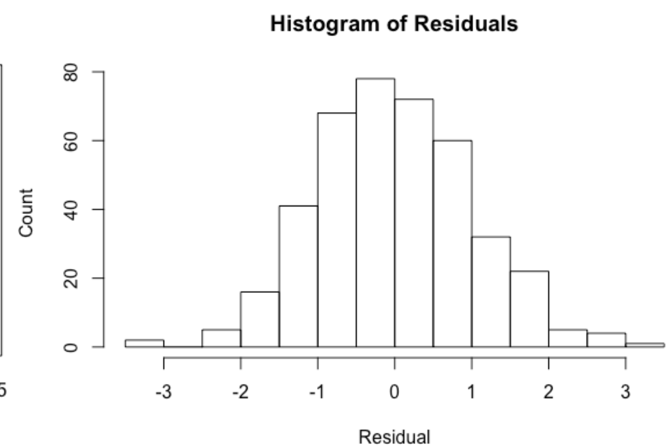
	Estimate	Standard Error	t-Statistic	P-Value	95% CI
Constant	32.29264	0.10071	320.664	< 2e-16***	(32.094661, 32.490615)
t (years)	0.01566	0.00515	3.041	0.00251**	(0.005536, 0.025784)
Sin(2πt)	2.13606	0.07121	29.995	< 2e-16***	(1.996061, 2.276056)
Cos(2πt)	2.22112	0.07134	31.134	< 2e-16***	(2.080871, 2.361363)
Cos(2πt*16)	0.17023	0.07272	2.341	0.01973*	(0.027264, 0.313200)

*p-value < 0.05 **p-value < 0.01 ***p-value < 0.001

This model performs slightly better than the first, as 82.35% of the variation in monthly maximum temperatures is now explained by the relationship to time.

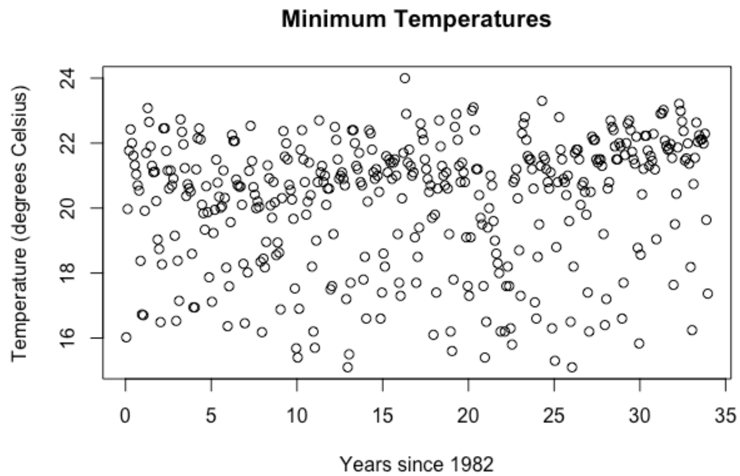
Residuals





The residuals for the adjusted linear oscillating model are very similar to those of the original. The histogram of residuals appears to resemble a normal distribution slightly better than the first. Also, the plot of residuals looks to be almost identical, but if observed closely one can see that the “moustache” effect seems to have been mitigated, which is what the new model aimed to fix.

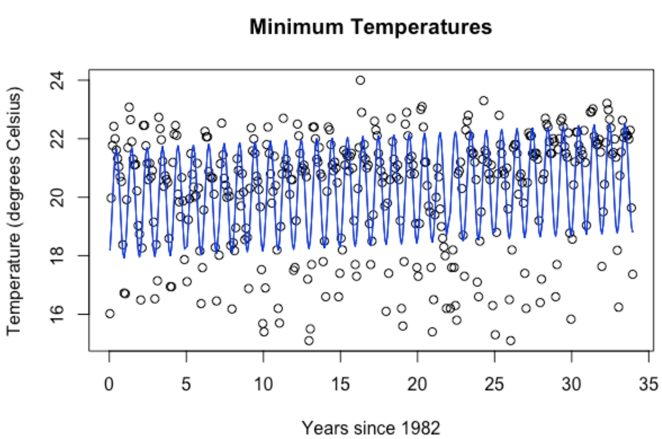
Minimum Temperatures



Above is a scatter plot of monthly minimum temperatures in Sotouboua measured in degrees Celsius starting in 1982. At first glance there is no obvious pattern, however temperatures do seem to cluster around the 22°C mark. Alyssa and Patrick found in their research a significant positive linear trend in the minimum temperatures. The models we have constructed seek to address monthly fluctuations as well as this trend.

Models

Linear Oscillating Model



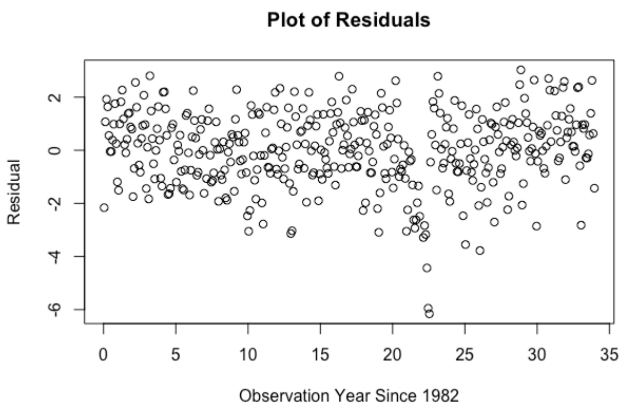
We first modeled Sotouboua's minimum temperature with four terms: a constant term, a linear term, and sine and cosine terms with periods of one year in an effort to account for seasonal variation in temperature (see left). In this model, each of the four terms were significant in the prediction of Sotouboua monthly minimum temperatures, and thus contribute important information to the model. The table of parameters can be seen below:

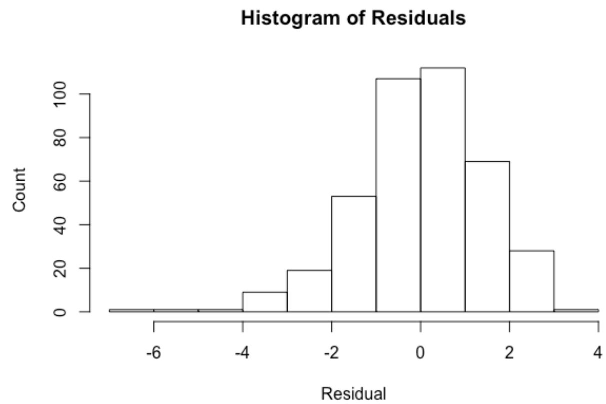
	Estimate	Standard Error	t-Statistic	P-Value	95% CI
Constant	19.770954	0.142842	138.412	< 2e-16***	(19.49013288, 20.05177476)
t (years)	0.026610	0.007299	3.646	0.000302***	(0.01226014, 0.04095941)
Sin(2πt)	0.559488	0.100557	5.564	4.86e-08***	(0.36179700, 0.75717955)
Cos(2πt)	-1.795457	0.101856	-17.627	< 2e-16***	(-1.99570032, -1.59521279)

*p-value < 0.05 **p-value < 0.01 ***p-value < 0.001

This is model found a moderate relationship in that 47.23% of the variation in monthly minimum temperatures for Sotouboua was explained by the relationship to time.

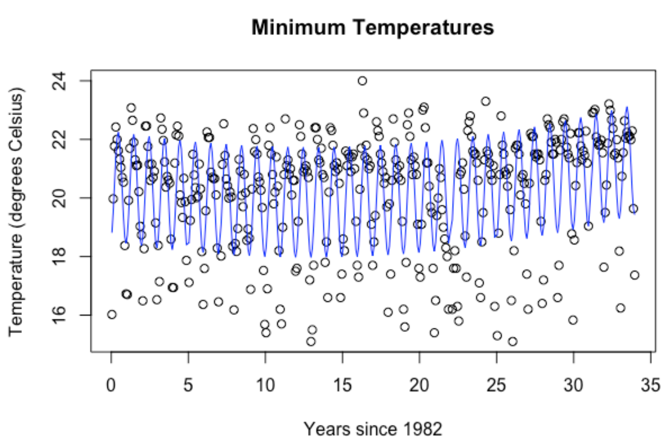
Residuals





The histogram of residuals looks slightly skewed left. This is due to the predictive model sitting higher on the plot, as this is where most of the data is. There are no real extremes above this clustering, but there are quite a few observations that sit below it. In the plot of the residuals, there are several large negative values in the 2000s despite our omitting of outliers. This is due to a few particularly low-minimum months (May, June, and July of ’04, & January of ’07, ’08, and ’15).

Quadratic Oscillating Model



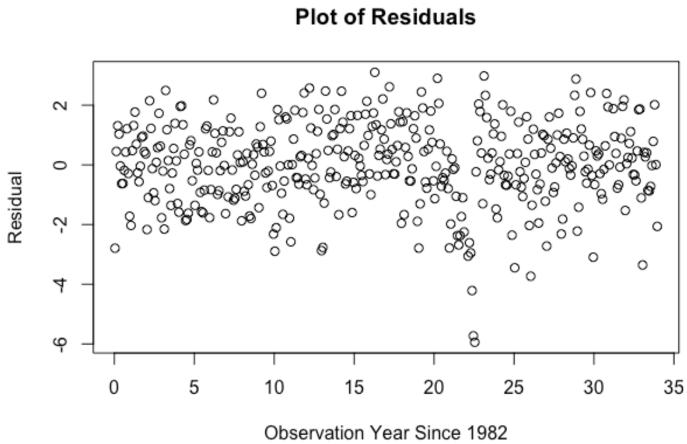
We then chose to model monthly minimum temperatures of Sotouboua with an additional quadratic term to try to explain any concavity in our minimum temperature data. The residuals from the prior model did not show a lot of curvature, but the data, especially around the clustering, does appear to show some. This was also evident in Patrick and Alyssa’s work. The quadratic term effectively models the decrease in minimum temperature beginning in 1982 and the slight increase in recent history.

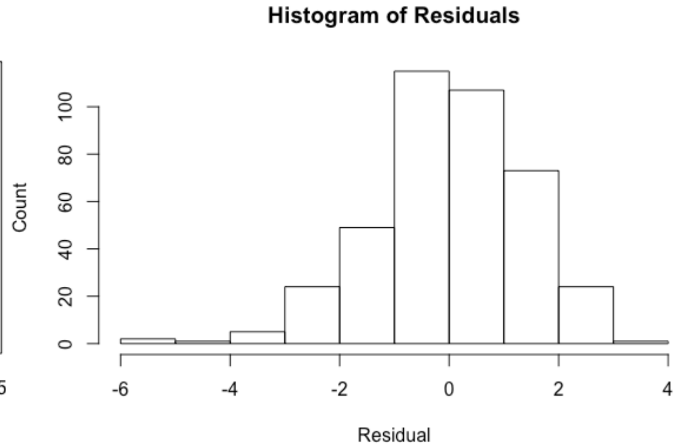
	Estimate	Standard Error	t-Statistic	P-Value	95% CI
Constant	20.4046635	0.2098243	97.246	< 2e-16***	(19.992154746, 20.817172203)
t (years)	-0.0853292	0.0284994	-2.994	0.00293 **	(-0.141358320, -0.029300097)
t^2	0.0032957	0.0008122	4.058	5.97e-05***	(0.001698999, 0.004892332)
Sin(2πt)	0.5610955	0.0986547	5.687	2.51e-08***	(0.367143194, 0.755047896)
Cos(2πt)	-1.7949187	0.0999277	-17.962	< 2e-16***	(-1.991373883, -1.598463582)

*p-value < 0.05 **p-value < 0.01 ***p-value < 0.001

This model finds a slightly stronger relationship between time and monthly minimum temperature than its predecessor as it found that 49.34% of the variation in the minimums was explained by the relationship to time. A discussion of the residuals is found on the next page.

Residuals





The plot of residuals and histogram look strikingly similar to the linear model, and this is because the coefficient on the quadratic term (0.0033) is small. However, it does appear that the issue of concavity in the residuals that came with the original model is missing from the new model, which was the goal of adding the new term. The histogram still shows left skewness, and this is for the same reasons mentioned above. Without these observations the graph would resemble a normal distribution quite well.

Discussion of Results and Conclusion

Final models

The final model we have chosen to fit the maximum temperature data with is the linear oscillating model with the 16-year "beat". This model sees 82.35% of the variation in monthly maximum temperatures explained by the relationship to time. While this R^2 value is not much larger than the simpler model, it does seem to have mitigated the issue of a pattern in the residuals. The oscillation period for this model is one year. The highest temperatures in the waves occur mid-February, and the lowest occur mid-July (see Appendix A). The coefficient on the year term is positive, which indicates an increase in maximum temperatures since 1982. We do not yet know why there would be a 16-year beat in the data, however we believe this is important to incorporate into the model for the time being.

Final Models:

$$Maximum = 32.2926 + 0.0157t + 2.1361\sin(2\pi t) + 2.2211\cos(2\pi t) + 0.1702\cos(2\pi t * 16)$$

$$Minimum = 20.4047 - 0.0853\text{t} + 0.0033t^2 + 0.5611\sin(2\pi t) - 1.7949\cos(2\pi t)$$

The final model we have chosen to fit the minimum temperature data with is the quadratic oscillating model. There is certainly some curvature in the residuals of the linear model, and this model addresses this problem. This model finds a moderate relationship in that 49.34% of the variation in monthly minimum temperatures is explained by the relationship to time. We would expect the highest and lowest minimums to happen around the same time the highest and lowest maximums do, with the mean temperature shifting from 32.5°C to 20.2°C. The model introduced a quadratic term that carries with it a positive coefficient, meaning the model is concave-up. This means that in recent history the minimum temperatures have been rising at an increasing rate.

Ideas for Future Research and Inquiries for Togolese Meteorologists

Future modeling with this data should explore the residuals in the minimum temperature data more. These errors, as well as the data in general, are very left-skewed. We addressed the most significant outliers by removing them from our models, but simply removing five observations does not rid the data of its inherent skewness. The underlying population does not come from a normal distribution, so some transformation techniques may be useful.

For future reference, we believe it would also be beneficial to examine the data for outliers on a monthly and/or seasonal basis. After we removed the overall outliers from our data, we still had some lingering large residuals, and this may have been because there are observations that are not outliers when compared to the rest of the data, but would qualify as outliers when compared only to the same month (i.e. low summer temperatures will be closer to the mean of the entire dataset when all months are included, but may be found to be abnormal when only compared to other summer temperatures).

Also, there a few questions that arose during our analyses that we would like to ask the Togolese meteorologists. We were left with a few extreme residuals in both models for the maximum and minimum temperatures. These points were not outliers by any means, they just simply did not fit in with the models well. We would be interested in learning more about the following months: January ’02 and ’05 (low maximums), November ’97 (high maximum), May, June, and July ’04 (low minimums), January ’07, ’08, and ’15 (low minimums), as well as the outlier months mentioned in the report. Were extreme weather conditions occurring during these months? Or is it possible there was error in measurement? In addition to learning more about these months, we are curious to see if the Togolese have any explanation of their own for what appears to be a 16 year “beat” in maximum temperatures.

Appendix A

To see when the yearly maximums are at their highest and lowest points, we optimize the following expression:

$$\sin(2\pi t) + \cos(2\pi t)$$
$$2\pi\cos(2\pi t) - 2\pi\sin(2\pi t) = 0$$
$$\cos(2\pi t) - \sin(2\pi t) = 0$$
$$\cos(2\pi t) = \sin(2\pi t)$$
$$\frac{\sin(2\pi t)}{\cos(2\pi t)} = 1$$
$$\tan(2\pi t) = 1$$
$$2\pi t = {}^{\pi}/_{4} + 2\pi k \ or \ {}^{5\pi}/_{4} + 2\pi k$$
$$t = {}^{1}/_{8} + k \ or \ {}^{5}/_{8} + k$$

The highest points in the oscillation are at the 1/8 point in every year, while the lowest points in the oscillation are at the 5/8 point in every year. These points correspond to mid-February and mid-July.