[Rose-Hulman Undergraduate Mathematics Journal](https://scholar.rose-hulman.edu/rhumj)

Local Warming in Low Latitude Locations: A Time Series Analysis

Ryan Pittman Anderson University, ryanpittman1994@yahoo.com

Follow this and additional works at: [https://scholar.rose-hulman.edu/rhumj](https://scholar.rose-hulman.edu/rhumj?utm_source=scholar.rose-hulman.edu%2Frhumj%2Fvol19%2Fiss1%2F7&utm_medium=PDF&utm_campaign=PDFCoverPages)

Part of the [Climate Commons](http://network.bepress.com/hgg/discipline/188?utm_source=scholar.rose-hulman.edu%2Frhumj%2Fvol19%2Fiss1%2F7&utm_medium=PDF&utm_campaign=PDFCoverPages)

Recommended Citation

Pittman, Ryan (2018) "Local Warming in Low Latitude Locations: A Time Series Analysis," Rose-Hulman Undergraduate Mathematics Journal: Vol. 19 : Iss. 1 , Article 7. Available at: [https://scholar.rose-hulman.edu/rhumj/vol19/iss1/7](https://scholar.rose-hulman.edu/rhumj/vol19/iss1/7?utm_source=scholar.rose-hulman.edu%2Frhumj%2Fvol19%2Fiss1%2F7&utm_medium=PDF&utm_campaign=PDFCoverPages)

Local Warming in Low Latitude Locations: A Time Series Analysis

Cover Page Footnote

Dr. Diana Ivankovic, Professor of Biology, Anderson University, and her student Eslie Aguilar did some of the data collection for this paper; we are thankful for their contribution. Also, this research was partially funded by a student/faculty research grant from South Carolina Independent Colleges and Universities.

Rose-Hulman UNDERGRADUATE MATHEMATICS J_{OURNAL}

LOCAL WARMING IN LOW LATITUDE LOCATIONS: A TIME SERIES ANALYSIS

Ryan Pittman^a

Volume 19, No. 1, Fall 2018

Sponsored by

Rose-Hulman Institute of Technology Department of Mathematics Terre Haute, IN 47803 mathjournal@rose-hulman.edu scholar.rose-hulman.edu/rhumj

^aAnderson University, ryanpittman1994@yahoo.com

Rose-Hulman Undergraduate Mathematics Journal Volume 19, No. 1, Fall 2018

LOCAL WARMING IN LOW LATITUDE LOCATIONS: A TIME SERIES ANALYSIS

Abstract. Global warming is a well-known and well-studied phenomenon pertaining to a gradual increase of average global temperatures over time. Many global warming mathematical models make certain assumptions regarding the factors that impact global temperature. These assumptions include effects from increased global carbon dioxide levels in the atmosphere and the melting ice sheets, among others. This paper draws conclusions about temperature changes without the assumptions needed for the global warming mathematical models. Instead of using computer models to project temperatures on a global scale, 33 low-latitude locations in the southern United States were individually studied to see if each one has warmed over time. Each location's daily high temperature was obtained for each day since January 1, 1970, and the data is analyzed using a statistical model that contains a linear effect coefficient, an annual seasonal trend, and a 10:7-year solar cycle trend, both well-known and commonly- accepted periodic trends. The data is also analyzed using Fourier frequencies to check for other lesser-known periodic trends. The strongest trend is then added to the original model. Linear effect coefficients are calculated for each location using both the updated model and the original model to see how they compare to each other and global warming study results. Only using the 31 locations where these models fit, the original model yielded an average linear increase of 2:29 degrees Celsius per century, while the updated model showed an average linear increase of 2:33 degrees Celsius per century. Showing less than a 2% difference indicates that the original model is sufficient and that temperatures in low-latitude locations are increasing.

Acknowledgements: Dr. Diana Ivankovic, Professor of Biology, Anderson University, and her student Eslie Aguilar did some of the data collection for this paper; we are thankful for their contribution. Also, this research was partially funded by a student/faculty research grant from South Carolina Independent Colleges and Universities.

1 Introduction

According to The Stanford Solar Center (2008), global warming is defined as "a gradual increase in planet-wide temperatures," and it enjoys broad acceptance within the scientific community [9]. Countless studies have been done to determine the causes of this global phenomenon and its magnitude. As stated by NASA (2017), evidence such as rising sea levels, shrinking ice sheets, and warming oceans are all signs of rapid climate change around the globe [6]. Based on tests performed by NASA, global temperatures are 0.99°C warmer than the mid-20th century mean. This is just one of many studies that conclude that global temperatures are on the rise.

In a test to determine whether or not global warming was in the midst of a hiatus, a mathematical model of global temperatures developed by Rahmstorf (2014) indicated that global temperatures have increased $0.175 \pm 0.047^{\circ}$ C per decade, or a 1.28°C to 2.22°C increase over the past century if the same trends continue [8]. Rahmstorf's study also concluded there has not been a statistically significant reduction in global warming, but there has been significant warming primarily beginning around 1970. Morano (2016), offers a contrasting position regarding global warming [5]. His research does conclude that there has been a pause in global warming over the past 18 years and 9 months (since *el Niño* in 1998). He is a climate change skeptic and believes that global warming fears must be stopped if this pause continues any longer.

In the context of these global warming studies, Robert J. Vanderbei of Princeton University (2012) has developed a different approach to study longterm temperature changes. While prior studies analyzed factors such as carbon dioxide levels in the ozone to support warming conclusions, Vanderbei focused solely on the temperature data itself in an attempt to determine if McGuire Air Force Base in New Jersey, located near Princeton University, has experienced any warming over a period of past 55 years (January 1, 1955 to August 13, 2010) [10]. Thus, rather than modeling this phenomenon on a global scale, Vanderbei chose to test the theory of global warming locally. Vanderbei obtained temperature data from the National Oceanic and Atmospheric Administration (NOAA), which collects and stores weather data from thousands of weather stations around the globe. Using this data, he created a statistical regression model that takes into account two known periodic weather trends in the data: the annual seasonal trend and the 10.7 year solar cycle.

The model he created is:

$$
T_d = a_0 + a_1 d + a_2 cos\left(\frac{2\pi d}{365.25}\right) + a_3 sin\left(\frac{2\pi d}{365.25}\right)
$$

+
$$
a_4 cos\left(\frac{a_6 2\pi d}{10.7 \cdot 365.25}\right) + a_5 sin\left(\frac{a_6 2\pi d}{10.7 \cdot 365.25}\right)
$$
 (1)

where T_d represents the daily high temperatures and d represents time in days. The key part in this model is the a_1 term. This term represents the linear effect coefficient in the model, and it is the portion of temperature change that does not naturally reverse over time as part of a single or multiyear periodic cycle. The other variables are used to account for the naturally occurring periodic trends in the data. The a_2 and a_3 term are used to identify the annual seasonal trend while a_4 and a_5 account for the solar cycle. The $a₆$ parameter is an unknown parameter associated with the solar cycle. The a_1 term is the linear effect coefficient that cannot be taken into account by these trends. In this particular situation, Vanderbei discovers that the optimal parameter value for a_1 is $9.95 \cdot 10^{-5}$ °F per day, or approximately 3.63 °F per century. Based on these results, Vanderbei concludes that local warming is occurring at McGuire Air Force Base in New Jersey. It is important to note that these results are obtained using only raw data and well-known periodic trends rather than assumptions about carbon dioxide and other factors made by most mathematical global warming models.

This paper starts by applying Vanderbei's model to low latitude locations in the southern United States. There were 33 such locations analyzed to extract a possible linear trend from each location's dataset. With such a high number of similarly located cities being analyzed, if they all show similar warming trends, then this study acts as a different, less assumption-filled, method to support global warming. Additionally, this study seeks to verify that the seasonal and solar trends are the only statistically significant trends that need to be included in the model and that there are not additional periodic trends that must be considered before a valid local warming conclusion can be made. If there are other more relevant periodic trends within the data, then Vanderbei's conclusion may not be accurate.

This paper is organized as follows. Section 2 describes the data collection and a time series method for filling in missing data points. Section 3 contains the calculations of the linear trend variable using Vanderbei's model. Section 4 describes the use of Fourier frequencies and periodograms to check the data for additional periodic trends that are not considered in Vanderbei's model and modifies the results of Section 3 for these additional trends. Section 5 summarizes the results and gives some directions for further study.

2 Data Collection and Organization

The first step of the research was data collection. This data was obtained by submitting requests on the National Oceanic and Atmospheric Administration (NOAA) website for each individual location [7]. For inclusion in the data set, the location had to be in a low latitude location, between 24.555 °N (Key West) to 36.747 °N (Fresno) while most locations were near 30 °N. Additionally, each location must have an individual weather station that has been in operation since 1970. The low latitude requirement is imposed to add geographical consistency among the locations as well as to limit strong seasonality (i.e. the fact that low latitude locations in the United States do exhibit seasonal trends, but they are not as drastic as those in higher latitude locations). The requirements that the weather station had to be active and have nearly complete high temperature data since 1970 are imposed for data quality reasons. The year 1970 was chosen because it is a small enough window where it is feasible to have a complete data set from a sufficiently large number of locations, and it is a long enough period of time to offer reliable results. Also, Rahmstorf's study cited in the introduction concluded that significant warming began in 1970. Cities with at least one weather station that meets these qualifications include Atlanta, Charleston (SC), Dallas, Miami, San Antonio, and San Diego.

Before analyzing a dataset using a statistical model, the data points were counted to verify that there were no missing days in the temperature data. Because each dataset has the same beginning and ending date, each dataset should have 17,208 distinct data points. For data sets with fewer data points than the full number, the missing days were determined. Some locations were only missing one or two days within a set of over 17,000 days, and the maximum number of missing days for any selected location was ten. Despite finding over 30 locations that fit the criteria mentioned above, nearly twenty percent of the datasets contain missing points. Gelman and Hill (2006) claim that these missing points can create bias in the results, so finding an appropriate method to fill these missing points was critical [3]. There are numerous imputation techniques to handle missing values within data, but Gelman and Hill explain that the easiest way to fill in each missing point is through mean imputation. The only drawbacks to using this approach is that it can underestimate standard error and cause a distorted relationship between variables. Neither one of these problems are relevant to this study.

Taking the mean imputation method one step further, Cryer and Chan (2011) suggest using Autoregressive Moving Average Models (ARMA) to complete the data [2]. This way each missing point varies based on a weighted average of observations located around that missing value. Essentially, ARMA is a family of time series models that regresses the time series on itself.

$$
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t,
$$
\n(2)

where Y_t is the observed time series with an assumed mean of zero, ϕ_i are population parameters, ϵ_i is a white noise random variable with mean 0 and variance σ^2 that is independent of the time series Y_t , and p is a positive integer.

The next step is determining what order p within this family of models provides the best fit for the temperature time series. Crawley (2013) recommends using the ARMA function in R and manipulating the autoregressive element of function while leaving all else constant [1]. Once 25 different ARMA models were created and stored in R, the Akaike Information Criterion (AIC) was calculated for each model. Whichever model yields the lowest AIC is the best fit for the data. Next, the coefficients for the given model had to be calculated and then incorporated into the dataset to forecast what the missing data point would be based on the preceding data points in the time series. This process was repeated for every missing observation in each dataset. With missing data points filled by estimates, the data sets are ready to be analyzed.

For example, El Paso was one of the cities missing data points. Approximately ten dates, within the dataset of 17,208 points, were missing. As mentioned above, model10 to model250 were created using R where model250 uses an autoregressive term of 25, a difference term of 0, and a moving average term of 0. Once these models were created, each AIC was calculated. The results are as follows:

Based on these results, Model180 produces the smallest AIC, meaning that an ARMA model with $AR = 18$ and $MA = 0$ is the best model to use with this dataset. Once each ϕ_i coefficient in model180 was calculated, the model can accurately be used to fill in the missing data-points in El Paso.

3 Temperature Trend Estimation with Known Periodic Trends

In this section, Vanderbei's model is regressed against each location's data with the goal being to obtain a measure of goodness-of-fit and the linear effect coefficient a_1 for each location. Kutner et al. (2008) say that because the data exhibits known periodic trends, transformations must be applied to the data [4]. Transforming the data based on known trends flattens out the data to reveal only minute cycles and the all-important linear effect coefficient. Once these transformations are applied to each location, an estimate is obtained from the a_1 term in Vanderbei's model. Using the lm command in R, each location's temperature data was regressed against the variable included in Vanderbei's model. The results of this test showed how well the model fit the raw data from each location, yielding an adjusted R-Squared value that shows how closely the data falls around the line from Vanderbei's equation. A higher adjusted R-squred value suggests that the model more accurately fits the collected data. If the model shows good fit, then the parameter estimates are more reliable. Thus, the estimate for the a_1 term, is a good indication of how much the temperature is increasing, on average, per day. The lm command in R also produces a p value for each estimate. A small p value indicates that the value is significant, while a large p value suggests that the term is insignificant, and is most likely zero. These estimated coefficients along with the calculated p value and adjusted R-squared for each location are shown in Appendix A: Vanderbei Model Results.

Out of the 33 locations analyzed, only Los Angeles and San Diego showed evidence of poor fit using Vanderbei's model. Despite the lack-of-fit in these two locations, the average daily temperature increase in these 33 locations is 0.000105°F per day or 3.83°F per century. This is equivalent to 2.128°C per century, which is within the global study's range of 1.28° C to 2.22° C that is mentioned in the introduction. Taking out the two ill-fitting locations, the 100 year average increase is 4.12°F (2.29°C). Given that this value is only only 0.07 °C away from the interval given in aforementioned results. Every location except for Los Angeles and San Diego had a reliable p value of less than $\alpha = 0.05$. Most of them are computer 0, meaning that the linear effect coefficient $H_0: a_1 = 0$ can be rejected and $H_a: a_1 \neq 0$ is accepted. Additionally, each location's adjusted R-squared was calculated, and every location except the two poorly fitted data points noted earlier had an adjusted R-squared greater than 0.6 with most being above 0.7. The average adjusted R-squared is 0.6673 with the poorly fitted points and 0.6889 without the two poorly fitted points. Thus, Vanderbei's model fits the data well for most of the selected cities. Moreover, since these results fall within Rahmstorf's interval for average increase in global temperatures, these results are supported by global warming research results, without relying heavily on assumptions.

4 Fourier Frequencies and Periodograms

This section further analyzes the historical time series data to determine if other periodic trends within each data set need to be included in the data transformation. The goal is to find trends within the data and then incorporate them into the model to filter out these significant trends and clarify the linear piece. To do this, Fourier frequencies were calculated to find the most significant periodic trends within each dataset. An example of the output from Fresno's dataset is shown in Figure 1 below.

```
[1] 5.929788e+03 3.926289e-02 2.135916e-01 3.539790e-02 1.827143e-01
 [6] 1.026968e-01 2.492499e-03 9.936100e-02 1.657764e-01 9.305174e-03
[11] 5.207586e-03 2.830768e-02 3.804759e-02 4.360504e-02 1.486879e-02
[16] 1.179141e-01 4.322475e-03 4.986951e-03 1.917139e-02 4.751598e-02
[21] 8.262730e-03 1.604308e-02 3.291613e-03 2.636179e-02 3.028745e-02
[26] 1.427547e-02 8.106531e-02 2.271389e-02 4.309256e-02 1.664013e-03
[31] 7.549440e-03 2.528892e-02 6.615890e-02 4.337759e-03 4.929212e-02
[36] 1.492907e-01 9.082137e-03 3.813635e-02 1.972014e-02 4.130392e-02
[41] 3.937865e-02 5.983131e-02 9.560332e-02 9.693716e-02 1.665400e-01
[46] 2.285920e-01 1.058332e+00 1.109449e+02 1.541770e+00 5.888157e-01
[51] 4.793564e-02 5.615661e-02 2.203298e-01 5.478005e-02 5.505321e-02
[56] 3.431195e-02 3.621914e-03 1.240024e-02 2.447070e-02 5.204609e-02
[61] 4.916943e-02 1.050346e-01 7.238188e-02 4.589599e-03 1.367608e-02
F66] 8.450989e-03 8.316856e-02 5.756013e-02 5.640707e-03 7.967042e-03
F71] 1.456154e-02 5.061607e-02 2.093302e-02 2.305560e-03 4.598421e-03
F761 6.502961e-03 1.603540e-02 7.091931e-03 1.444577e-02 4.622930e-03
[81] 2.164716e-02 3.637457e-03 2.638316e-02 3.451892e-02 5.033562e-03
E867 3.716168e-03 5.294699e-03 1.964083e-02 9.841463e-03 2.192647e-02
[91] 5.989924e-03 5.955355e-02 8.521438e-02 1.109466e-01 1.985378e+00
[96] 3.010449e-01 4.214225e-03 3.310296e-02 5.626670e-03 2.568391e-02
```
Figure 1: Fresno Fourier Frequencies.

As expected, the highest output value takes place around the $n = 47, n =$ 48 spot in the matrix. This is an indication that there is a strong trend that

Any trend revealed after the $n = 48$ mark is ignored because it indicates potential trends happening within a year. This study is primarily concerned with multiyear trends, meaning that the next significant trend occurs at $n = 6$. This translates into a periodic trend of roughly 8 years within this particular dataset. Another important note is that the solar cycle, which is represented by $n = 5$, does not stand out among Fourier Frequencies for most of the data sets in this study. However, removing the solar cycle from the Princeton model worsens the model's fit of the data, meaning that even though it does not appear distinctly in the Fourier Frequencies, it is beneficial to include it in the model. Vanderbei notes this observation in his conclusions as well, citing the solar cycle is real but has a small impact on the results.

As previously stated, the linear effect coefficient in Fresno using the Princeton model revealed an increase of $1.25 \cdot 10^{-4}$ °F per day with a p value of near 0 and an adjusted R-squared of 0.7983. Once the newly discovered eight year trend was literally added onto the end of Vanderbei's model, the model for the Fresno data set becomes:

$$
T_d = a_0 + a_1 d + a_2 cos\left(\frac{2\pi d}{365.25}\right) + a_3 sin\left(\frac{2\pi d}{365.25}\right)
$$

+
$$
a_4 cos\left(\frac{a_6 2\pi d}{10.7 \cdot 365.25}\right) + a_5 sin\left(\frac{a_6 2\pi d}{10.7 \cdot 365.25}\right)
$$
(3)
+
$$
a_7 cos\left(\frac{2\pi d}{8 \cdot 365.25}\right) + a_8 sin\left(\frac{2\pi d}{8 \cdot 365.25}\right)
$$

Repeating the regression calculations for this new model, the linear effect coefficient a_1 increases slightly to 1.30 $\cdot 10^{-4}$ °F per day, retaining the same near zero p value and slightly increasing the adjusted R-squared to 0.7984. The lack of difference between the two models suggests that Vanderbei's model accurately incorporates all significant periodic trends, as the addition of new trends into the model decreases the model's degrees of freedom. To more confidently conclude that Vanderbei's model accurately reflects what is occurring in the data, the same process was repeated in all of the other locations. The results are shown in Appendix B: Adjusted Model Results.

Each location's model was individually modified using whichever trend the Fourier frequencies indicated to be most significant. Some common trends within the data were 4 and 15 year trends. Unfortunately, adding the next best trend also did not improve the results in Los Angeles or San Diego, as

the adjusted R-squared remained quite low. After removing these two bad locations, 31 cities are left. As seen in Appendix B, none of the results in these cities changed much after the addition of the new trends. The average change in a_1 was $2.02 \cdot 10^{-6}$ °F per day (individual location's p values given in Appendix B), and the maximum change was $2.24 \cdot 10^{-5}$ °F per day in Tampa, with a p value of $3.50 \cdot 10^{-12}$. The average linear effect coefficient changed from an increase of 4.12 °F per century in the original model to an increase of 4.19 °F (2.33 °C) per century, which is not a statistically significant change as it is less than 2% different than the results from the original model. Again, these results validate the original model and the conclusion that says that the average temperature in low latitude locations is on the rise.

5 Conclusion

This paper concludes that these low latitude locations are heating up over time. This result agrees with the results of previous global studies studies. However, this study's results were not found by making the assumptions of the other studies. In fact, the results were found solely using known periodic trends and the actual high temperature data, making the results more robust. Also, a Fourier frequency analysis shows that Vanderbei's model includes all significant trends. Lastly, it can be seen that the results found in low latitude locations are consistent with the results from New Jersey.

In future studies, modeling locations of different international latitudes could help discover if the Princeton model is also applicable to such locations, provided reliable data is available for locations outside the United States. If not, the Fourier frequencies could be used to create a new model to clarify the linear effect coefficient within the data's many trends. Expanding this study to locations in all parts of the world can distinguish locations of the Earth that are warming at a faster rate than others, and potentially lead to stronger conclusions about global warming. In addition to studying different locations, identifying the source of the discovered 4 and 15 year trends could yield a periodic trend currently unknown to climate scientists.

References

- [1] Crawley, Michael J. The R Book. 2nd ed. Chichester: Wiley, 2013. Print.
- [2] Cryer, Jonathan D., and Kung-sik Chan. Time Series Analysis with Applications in R. New York: Springer, 2011. Print.
- [3] Gelman, Andrew, and Jennifer Hill. Applied Regression and Multilevel/Hierarchical Models. Cambridge University Press, 2006. Print.
- [4] Kutner, Michael H., Chris Nachtsheim, and John Neter. Applied Linear Regression Models. Boston: McGraw-Hill, 2008. Print.
- [5] Morano, Marc."No Global Warming at All for 18 Years 9 Months ? a New Record ? The Pause Lengthens Again ? Just in Time for UN Summit in Paris." Climate Depot. N.p., 13 Jan. 2016. Web. 13 Mar. 2017. climatedepot.com
- [6] "Climate Change Evidence: How Do We Know?" NASA. NASA, 23 Feb. 2017. Web. 13 Mar. 2017. NASA.gov
- [7] National Centers for Environmental Information (NCEI). "Climate Data Online." Climate Data Online (CDO) - The National Climatic Data Center's (NCDC). National Oceanic and Atmospheric Administration, n.d. Web. 03 Apr. 2017. ncdc.noaa.gov
- [8] Rahmstorf, Stefan. "Recent Global Warming Trends: Significant or Paused or What?" RealClimate. N.p., 4 Dec. 2014. Web. 13 Mar. 2017. realclimate.org
- [9] "Global Warming Research Issues." Stanford Solar Center. N.p., 2008. Web. 14 Mar. 2017. solar-center.stanford.edu
- [10] Vanderbei, Robert J."Local Warming." SIAM Review 54.3 (2012): 597- 606. 17 Mar. 2011. Web. 13 Mar. 2017.

A

Vanderbei Model Results

B

Adjusted Model Results