MATH 426 Lab 2

Shane Cincotta, Fangzhou Li

10/18/19

Abstract:

The goal of this lab is to model the average temperature as a constant x_0 plus a linear trend $x_1 d$ plus a sinusoidal function with a one-year period representing seasonal changes. Our data is a collection of 55 years of daily average temperatures from a local weather station.

Detailed Strategy:

The temperature and time data were given to use in a CSV file called *McGuireAFB.csv*, before we could plot our data, we had to manipulate the *dates* column of our *csv* file to a usable type (initially it was type *datetime*). To achieve this, we casted the dates to strings, and then represented the dates as the number of days since the first day. Thus day 2 would be represented as 1 (1 day since the first day.)

```
data = csvread('McGuireAFB.csv');
dates = data(:,1);
formatteddates = datetime(string(dates),'InputFormat','yyyyMMdd');
DaysSinceDay0 = days(formatteddates-formatteddates(1));
```

Fig 1. Converting dates to usable format

After we had both the temperature and dates in a usable format, we had to fit the data to a model using least squares. As described above, our model will look like x_0 plus a linear trend $x_1 d$ plus a sinusoidal function with a one-year period representing seasonal changes. We began to build our normal matrix to satisfy those requirements, as well as solving for the coefficients.

```
%Begin to onstruct the normal matrix to fit a line and a curve fitting
 to the data.

%x_0+x_1*x+x_2*cos(2?d/365.25)+ x_3*sin(2?d/365.25) and x_0+x_1*x

%A1'*A1*x = A1'*temperatures is the function of normal matrix, and x
 is the
%coefficient matrix. Temperatures is is an n*1 matrix, A1 is the n-by-
m design matrix
%for the model and the model equatioin should be
%x_0+x_1*x+x_2*cos(2?d/365.25)+ x_3*sin(2?d/365.25).
A1= [0*x+1 x cos(2*pi*x/365) sin(2*pi*x/365)];
%A1'*A1; left side of normal equation except x.
normal = A1'*A1;
%A1'*temperatures right side of normal equation.
B1 = A1'*temperatures;
coefficient = normal\B1
```

Fig 2. Constructing normal matrix and solving for coefficients

After creating our normal matrix and solving for the coefficients, we plotted our new fit. After plotting, we noticed that the amplitude of our function does not match that of the original scatter plot, thus we had to scale our amplitude accordingly.

```
X = linspace(1, 20309, 20309);
plot(X,
 1.7*coefficient(4)*sin(2*pi*X/365.25)+1.7*coefficient(3)*cos(2*pi*X/365.25)+coeff
hold on;
%plot of sinusoid curve.


plot(x, coefficient(2)*x-6+coefficient(1), 'b');
```

Fig 3. Plotting and scaling amplitude

The resulting graph including the linear trend is shown below.

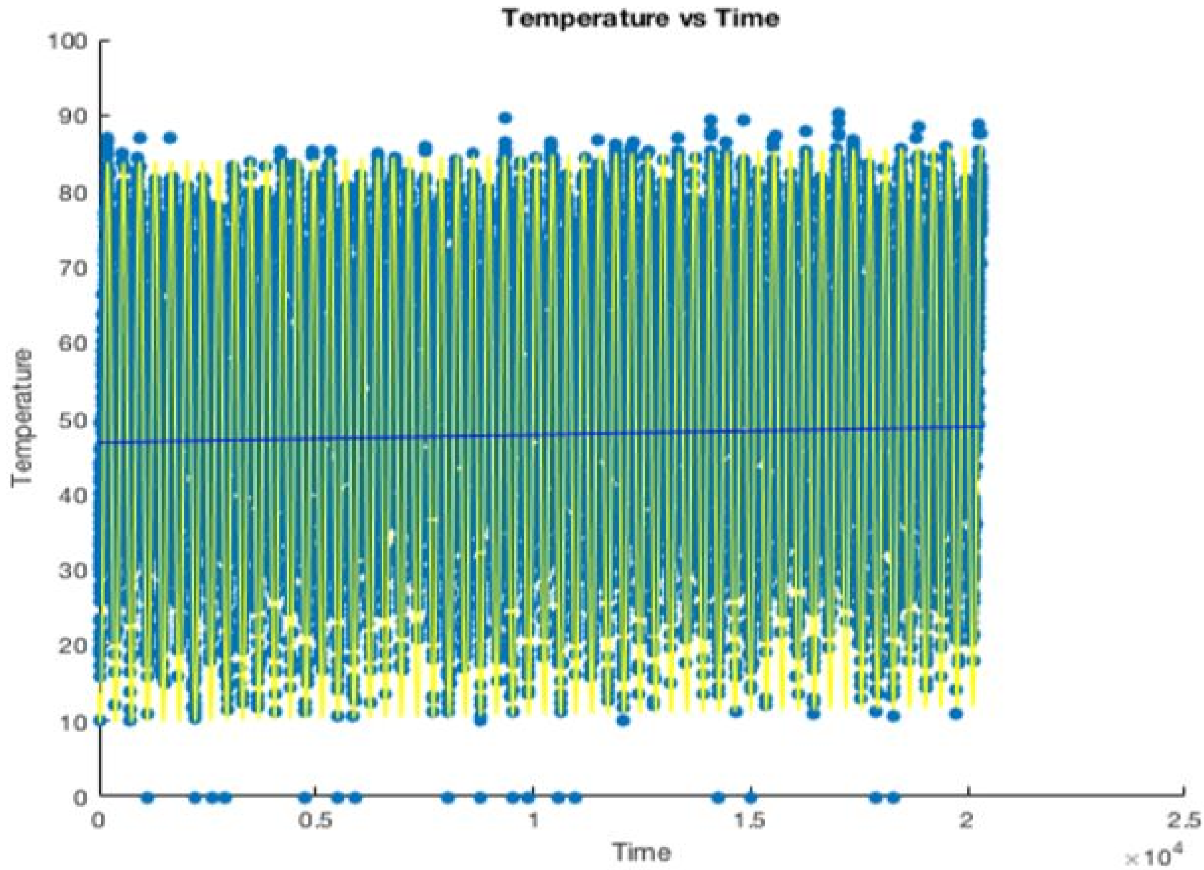

Fig 4. Best fit line and linear trend

Results:

Is the weather getting warmer or not?

Based on our graph, when one looks at the sinusoidal portion, the temperature does not seem to be changing with each season. However, when one looks at the linear trend, it is clear that the temperature is increasing, albeit not a lot. According to our linear trend, the temperature has increased a couple of degrees since the data was initially collected in 1955.

Propose a method for determining the solar cycle period or other nonlinear parameters. What are the difficulties? Discuss

If we wanted to determine other unknown nonlinear parameters we would first sample the data at a high rate for an extended period of time. The reason for this is since our unknown function is nonlinear, its behaviour can seem to change within short observational periods. I.e, the function can look drastically different from $0 < t < 3$ and $3 < t < 6$. After we have sampled our data, we would then need to find a function which behaves in the same way. For example, if we plotted our data and noticed that it oscillated between 2 numbers, we would be led to assume it can be modeled with a trigonometric function. After we found a function which behaves in the same way our data, we have to solve the coefficients in the same way that we did for this lab.

Finding nonlinear parameters can be difficult for a number of reasons. One reason is that nonlinear functions can be difficult to predict in time. What I mean by this is that the function can look drastically different from $0 < t < 3$ and $3 < t < 6$. Thus to be sure we are getting a correct view of our function over time, we need to sample our data over long intervals. This can take up lots of time and memory depending on the size of the dataset. Another issue that arises from finding nonlinear parameters is finding a function that represents the data. While this may seem obvious, it is important to mention because finding a nonlinear function to fit to data is harder than finding a linear function because it is changing exponentially. Thus if you have a longer observational window, it becomes more and more complex to fit a function to your data as the requirements of the functions become more complex. Finally, even after completing the above, one still needs to find the coefficients to those nonlinear functions, adding even more time.

Conclusion:

In conclusion, in this lab we were able to successfully parse more than 60 years of climate temperature data and model the average temperature as a constant x_0 plus a linear trend x_1d plus a sinusoidal function with a one-year period representing seasonal changes. We created this model using least squares. From our model we were able to see that the average temperature has been slowly and steadily increasing for the past 60+ years.