

Rahevin Potter-Clark

CS 434 Data Mining

Final Project: Project Proposal

## Global and Sea Surface Temperatures Proposal

### Abstract:

This project proposal is to determine and find relationships between the Global Surface Temperature Initiative and the Sea Surface Temperatures between the dates of January 1988 to December 2007. The proposal is to find the average temperatures per month and year between these two sets of data and graph them to determine similarities between. I will also calculate the rate of change per month and per year to also try to discover more similarities between the two data sets.

### Introduction:

For this project we will be using two groups of data sets; The first data set will contain International Surface Temperature Initiative, which contains a monthly timescale mean, max, and min temperature from observations from the 1800s to the present. Which can be found at [https://catalog.data.gov/dataset/international-surface-temperature-initiative-isti-global-land-surface-temperature-databank-sta#topic=coastalflooding\\_navigation](https://catalog.data.gov/dataset/international-surface-temperature-initiative-isti-global-land-surface-temperature-databank-sta#topic=coastalflooding_navigation) we will refer to this data set as **GSTemp** for “Global Surface Temperature”. Our next data set is Climate Data Record of Sea Surface Temperature, which contains data from January 1988 – December 2007 with a 3-hourly observation recording. Which can be found at <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00887> we will refer to this data set as **SSTemp** for “Sea Surface Temperature”. We will be using this data to determine the monthly average, monthly and yearly rate of change, and if the data sets have any similarities. This data will provide important research for those interested in the differences between the temperatures of the Earth’s land and sear surfaces.

### Data Mining Task:

Tasks that would be included in the firsts steps would be displaying each data on a single graph. We would have to extract the recorded temperatures and data taken from the rest of the data and place them inside the graph. Then next step would reduce **GSTemp** to only contains temperatures from January 1988 – December 2007, and because this data set already has our averages we only need to find our rate of change between each month and year and for good measure we should calculate the total rate of change from Jan 1988 to December 2007. Next step is to convert our **SSTemp** into a monthly base set from the current 3-hour intervals. To do this we must first create an average for each day, then the next important step is to determine how we want to calculate our months given some have 28, 30 or 31, and the occasionally 29 every four years. Between the date range given there are a total of 5 leap years at 1988, 1992, 1996, 2000, and 2004. After the months and years are calculated we will calculate the total rate of change for each month, year and total data set. Once completed we will compare the results of **GSTemp** and **SSTemp**.

### Data Set:

- **The International Surface Temperature Initiative: (“GSTemp”)** which contains a monthly timescale mean, max, and min temperature from observations from the 1800s to the present [https://catalog.data.gov/dataset/international-surface-temperature-initiative-isti-global-land-surface-temperature-databank-sta#topic=coastalflooding\\_navigation](https://catalog.data.gov/dataset/international-surface-temperature-initiative-isti-global-land-surface-temperature-databank-sta#topic=coastalflooding_navigation)
- **Climate Data Record of Sea Surface Temperature: (“SSTemp”)** which contains data from January 1988 – December 2007 with a 3-hourly observation recording <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00887>

The main challenge with these data sets would be converting them into the same sample size. Meaning reducing GSTemp and adjusting SSTemp to be monthly from hourly. Reducing GSTemp shouldn't be too hard of a task, whereas SSTemp changing from hourly to monthly/yearly taking into account of leap years changes the calculations of an almost 20 year spread of data.

## Methods and Models:

- **SSTemp from hourly to monthly/yearly points of data**

Changing out SSTemp data from hourly to monthly/yearly may look like this: We are given data in 3hour chunks, 8 of these make a day, and we know there are 5 leap years in our data set so to create our average month data we must first find our average days in a month which would be 30.4375. This translate for every 1-243.5 points of data we have our month, to generate our average we just divide by 243.5. Based on this we can also find a year, with every 1-2922 points of data we have a year.

- **GSTemp Reduce Data Range**

This data incorporates years and months, I would start by removing the years not included in the SSTemp data such as all including and before 1987 and everything after and including 2008.

- **Calculating our Rate of Change**

Rate of change can be calculated by taking the Change in output divided by the Change of input. Or in math terms  $((y_2 - y_1) / (x_2 - x_1))$ . For example, calculated our data month to month would look like this, labeling each month 1 through 12.

$$\frac{\text{Degree2} - \text{Degree1}}{\text{Month2} - \text{Month1}}$$

$$\frac{82^{\circ} - 78^{\circ}}{2 - 1} = 4^{\circ} \text{ for a month to month}$$

Determining year to year will be very similar but the only difference will be using December month of a year – the January month of a year.

$$\frac{\text{Degree12} - \text{Degree1}}{\text{Month12} - \text{Month1}}$$

$$\frac{41^{\circ} - 38^{\circ}}{12 - 1} = 0.2727^{\circ} \text{ for the year}$$

Similarly, for good measure we will determine the rate of change or the entirety of the data set.

$$\frac{(\text{Degree12 from December 2007}) - (\text{Degree1 from January 1988})}{\text{Month467} - \text{Month1}}$$

With 467 being the number of months given from our date range within the data.

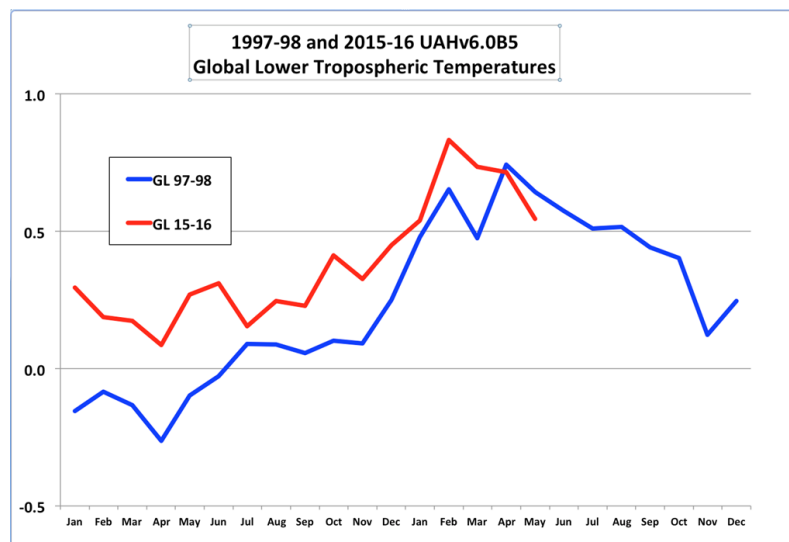
## Assessment:

With the data collected and calculated we can then produce visual data and view similarities or differences from the average temperatures of the sea and global surfaces. We can place our observation into a graph with two lines one for GSTemp and one for SSTemp for each of our calculated data, one that shows our average temperature per month, one that shows our average temperatures per year, and one that shows our rate of change per month and year. For our rate of change for the entire data sets we can use to compare which seems to be growing or shrinking the most. At this point we can start making our speculating and determine how one data set may affect the other or how they relate or contrasts. The benefits to these calculations may provide an algorithm that allows us to find the temperature of the surface to the sea based on our global surface temperature or vice versa.

## Presentation and Visualization:

I would present this data with several graphs with each having two lines that represent the two points of data's averages for each calculation. I may even include the differences in temperatures between the two data sets for each year's averages to show their possible relationships. As for the rate of change for the entirety of the data sets I would just show those numbers side by side.

("Not Actually data")



("Not Actually data")