

# Climate change since 1750

## Proposal Paper

Kaixin Chen

Department of Computer Science  
University of Colorado Boulder  
1111 Engineering Dr  
Boulder, Colorado 80303  
changshawyuanzu@gmail.com

Xinyi Zhang

Department of Computer Science  
University of Colorado Boulder  
1111 Engineering Dr  
Boulder, Colorado 80303  
xizh3366@colorado.edu

Yonghao Zhang

Department of Computer Science  
University of Colorado Boulder  
1111 Engineering Dr  
Boulder, Colorado 80303  
yozh2004@colorado.edu

## 1 PROBLEM STATEMENT

In this paper, we will examine the trend of climate change since 1750 in different area and country all over the world. We map the climate data with average temperature, country, latitude, and longitude while also concerning the uncertainties of each numerical data. In the analyze part, we will explore the relationship between temperature and latitude and discover the land average high and low trends of the climate. We would also concern the effect of melting glaciers and the development of industry to have a more conclusive idea about the climate change. The results from this analysis could help those climate studies discovering the phenomenon of global warming and helping those people that are not familiar with the climate change get a better and straightforward understanding about the changes in climate these years.

## 2 LITERATURE SURVEY

There are many existing studies analyzing how global climate has changed, which includes analysis on temperature, human effects, and some natural factors. These studies also concluded how to solve or slow down the process of the climate change and what consequences the climate change might have.

### 2.1 EPA: United States Environmental Protection Agency

The report of EPA claimed that the primary cause of climate change is imbalance between energy entering and leaving the planet's system. As for we know, when incoming energy from the sun is

imbalance of energy is caused by the greenhouse effect, natural changes including changes within the sun and changes in Earth's orbit and changes in reflectivity. Greenhouse effect that causes the atmosphere to retain heat. Natural changes affect can affect the intensity of the sunlight that reaches Earth's surface that affects how much energy reaches Earth's system. Finally, changes in reflectivity also affect how much energy enters Earth's system because the amount of sunlight that reaches Earth can be reflected or absorbed that depends on Earth's surface and atmosphere. [1]

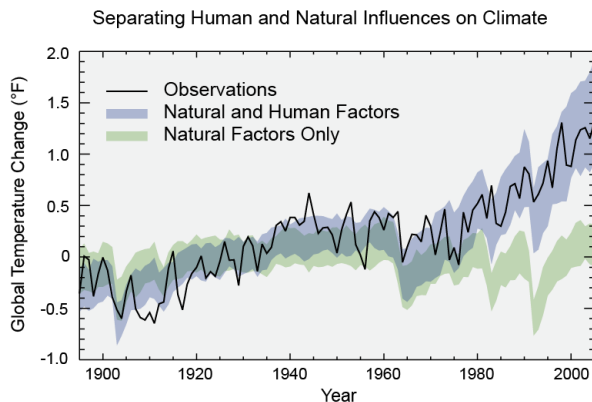
### 2.2 NASA: Climate Change and Global Warming

The NASA research about the climate change also concludes the global warming. The research conclusion of NASA could be regarded an authentic Wikipedia; in other words, the research includes evidence, causes, effects, and solutions. We might focus our attention on the Climate Resource Center, which including the analysis of the climate change through many aspects. These results are useful for us to prove the credibility of our project and could also help us conclude many other aspects we might need in the final conclusion part. [2]

## 3 PROPOSED WORK

We found a great data set from Kaggle as the raw data we tend to analysis and we divided the pre-processing work into three parts that containing data extraction, data cleaning, and data integration. [3] As this data set is kind of big for our test calculation at the beginning, we will extraction the first 1000 data points and store them as a temporary data set to make the coding part quicker and easier. And we also noticed that the land temperature data from 1750 to 1850 contains average temperature only instead of maximum temperature and minimum temperature, we will probably get rid of them while performing some analysis. As the time when we do the actual calculation and plot the graphs, we will make a throughout check to find the errors in the whole dataset. Data integration is one of the most important parts in our study. For some of our interesting questions, we need to compare the data between different latitude, different area, and different time period. We aim to answer the following questions in the course of the project:

- (1) Is Global Warming a fact or fiction?
- (2) Is temperature rising a global phenomenon or it only happens in the certain area?
- (3) What is the relationship between the climate change and altitude?



absorbed by the Earth system, Earth warms and vice versa. The

- (4) What is the relationship between the seasonality and climate trend?
- (5) What is the relationship between the highs and lows trend and climate trend?

The most noticeable difference between our study and what we have found previously in the literature survey is our project based on the database we choose to use. We not only compare the numerical data but also combine what we find through the data with other natural and human effects that have not been concluded in the dataset we will use. Also, we will only compare our conclusion with the already known analysis instead of using them as the basic theory of our study.

## 4 DATA SET

We obtained the five global temperatures since 1750 datas from the Kaggle.com. [3] The raw data comes from the Berkeley Earth data page, which is affiliated with Lawrence Berkeley National Laboratory. These five data sets repackaged 1.6 billion temperature reports from 16 pre-existing archives. In this dataset, we have include several files:

- **Date:** starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures
- **LandAverageTemperature:** global average land temperature in celsius
- **LandAverageTemperatureUncertainty:** the 95% confidence interval around the average
- **LandMaxTemperature:** global average maximum land temperature in celsius
- **LandMaxTemperatureUncertainty:** the 95% confidence interval around the maximum land temperature
- **LandMinTemperature:** global average minimum land temperature in celsius
- **LandMinTemperatureUncertainty:** the 95% confidence interval around the minimum land temperature
- **LandAndOceanAverageTemperature:** global average land and ocean temperature in celsius
- **LandAndOceanAverageTemperatureUncertainty:** the 95% confidence interval around the global average land and ocean temperature

Other files include:

- Global Average Land Temperature by Country (GlobalLandTemperaturesByCountry.csv)
- Global Average Land Temperature by State (GlobalLandTemperaturesByState.csv)
- Global Land Temperatures By Major City (GlobalLandTemperaturesByMajorCity.csv)
- Global Land Temperatures By City (GlobalLandTemperaturesByCity.csv)

## 5 EVALUATION METHODS

One of the most evaluation methods for our project is to check the credibility through comparing our results with the existing studies or conclusions. As the climate change is a relatively old and well-known topic that has been studied for many years, it would be not hard for us to find plenty of related studies. We will

not perform our project base on these already-completed results absolutely; however, it is necessary for our team to compare the results as the data set we decided to use contains only temperature issue while the climate change is a complicated process. Moreover, many existing studies are pretty authentic as many famous results have been accepted world-widely for many years. Furthermore, checking whether our data is useful to examine the question we are interested in and could our data provide the answer to these questions are necessary evaluation methods. Even if we obtain many results from the data set, we still need to make sure we do the correct and proper analysis that could provide the enough information for us to draw the conclusions.

## 6 TOOLS

- **SQL**  
Helping host the preprocessed database.
- **Matlab**  
Tool we might use at beginning to analysis the data and plot the graph. Might switch to Python if the complete data is too large for Matlab to compute efficiently.
- **Weka**  
Containing tools for data pre-processing, classification, regression, clustering, association rules, and visualization.
- **Python**  
Another tool to process the data efficiently.
- **Excel**  
The application that stores the original data.
- **Latex**  
Using  $\LaTeX$  to form the report.

## 7 MILESTONES

### 7.1 Previous Work

- **Data collection and extraction**  
First thing to do is extract data from an original source and store them.
- **Data pre-processing**  
Find the errors in data set and clean them. And then for some of our interesting question, we have to compare data between a different area in the certain area.
- **Process for derived data**  
We will only use the data points after 1850 because only data after 1850 has the minimum and maximum temperature for each country and each city.
- **Evaluate data and summary**  
Evaluate the credibility of the result and make conclusions.

### 7.2 What have achieved so far

We have finished the data extraction and data cleaning so far. As the date range in the most dataset is about 250 years, lacking values is a common phenomenon in many datasets. **Matlab** is our preferred tools for basic analyzing as we are most familiar with **Matlab** and it is much easier for us to analysis the basic data through **Matlab**. Through the data cleaning process, we used **Matlab** to read the raw data, which was transferred into table format finally, and calculated the mean values through “*nanmean*“, which was an effective tool

calculating the mean values while ignoring all the null values directly. We made a lot of effort during this process since we tried to optimize the input data in **Matlab** to make it more concise. In this way, the calculation would be swift and the code would be much easier to fix. Moreover, we found that most datasets were arranged in the same format. Instead of extracting the first 1000 data points as we planned before, we used one of the smallest datasets for our test calculation directly. This process helped us getting much more familiar with the mandatory attributes and saved a lot of time dealing with the differences between the sampling data and the raw datasets.

As we processing the data, we would be able to give a rough answer to the question we came up with before. We put our focus on the relationship between time period and global temperature currently. Through analyzing the dataset about global land and ocean temperature, we could give some results about the relationship between the highs and lows trend along with the global climate trend.

### 7.3 What remains to be done

We still have a lot of work needed to be processed obviously. We did many efforts on the global trend so far; however, in the proposal part, we put our focus on the regional analysis. For example, the question “*Is temperature rising a global phenomenon or it only happens in the certain area?*” and “*What is the relationship between the climate change and altitude?*” need more detailed analysis about the relationship between temperature and latitude and altitude. Also, even though we already have a basic concept of the structure of these datasets, we do want to visualize these datasets for a better understanding. This object is pretty tough for us currently due to the fact that we chose **Matlab** as the major analyzing tool. **Matlab** is a great calculate in many aspects. The concept of matrix helped us sort and process the unfamiliar raw data easily, but making a plot that contains a lot of information and elements in **Matlab** was not convenience. Our target was to make a global graph had major countries on it and we could mark the average temperature of each country through different colors. We still have problem plotting this graph as the graph could not be dragged no the report and **Matlab** could not process this kind of graph. To solve this problem, we probably would swift our major tool into Python later.

Another great challenge needed to be solved is the major target for the further analysis. As this dataset is about the detailed global temperature in 250 years, it would be tough for us to give a detailed yearly analysis about each city or country. Using time range is a better choice in this case. We need to choose several proper time point in this case. Another option might be putting our focus on a specific country, such as the United States. Choosing countries by their current developing status is also a factor we might consider. It is definitely helpful for us to compare the temperature status of developing countries and developed countries. Through the analysis of each country, we would be able to find what might be the factors that influence the temperature. Moreover, the growth rate of temperature in each region and each time range is a vital element we are looking for in order to answer the question “*Is Global Warming a fact or fiction?*”

Moreover, we need to add uncertainties into our calculation and plots later. In order to get familiar with the raw data easier, we

did not conclude all the uncertainties in our current calculation. The effect of uncertainty is not neglected for our analysis. And changes in uncertainty would also help us make more convincing conclusion.

## 8 RESULTS SO FAR

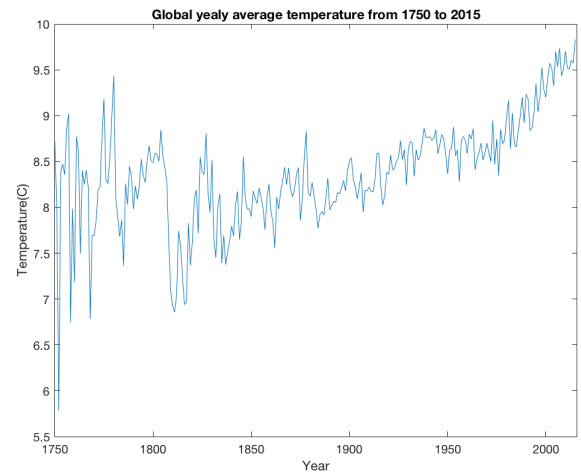
To solve the problem regarding “*What is the relationship between the seasonality and climate trend?*” we examined the climate trend from 2001-1-1 to 2013-09-01 in city Changchun, China. Our results show that the Global Warming is actually a fact that the average temperature from 1980 to 2013 in Changchun has increased by 1 Celsius. This conclusion is definitely not conclusive since we only analysis one major city.

We also analyzed the dataset about the global temperature form

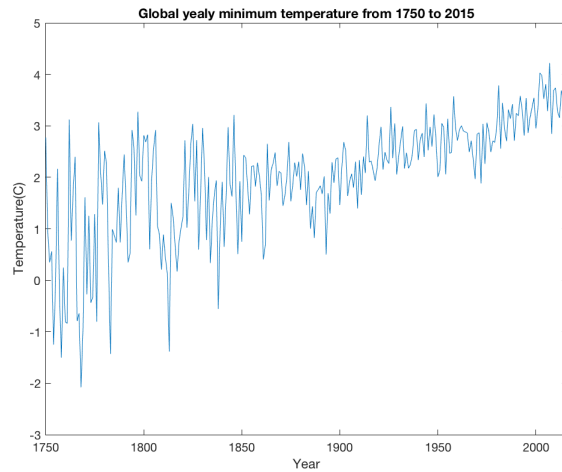
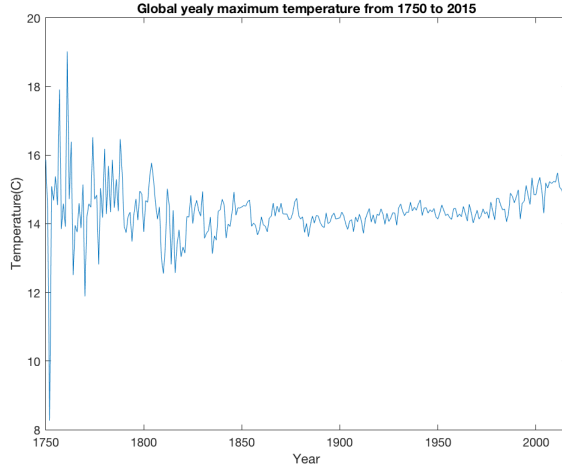
**Table 1: Average Temperature in Changchun**

Average temperature from 1820-08-01 to 2013-09-01			
Winter	Spring	Summer	Autumn
-14.3	6.0	22.0	5.9
Average temperature from 1980-01-01 to 2013-09-01			
-12.74	7.5	22.5	6.6
Average temperature from 2000-01-01 to 2013-09-01			
-13.12	7.7	22.7	6.9
Relationship with season			
Season	Total Month	Higher Average Month	Percent
Winter	38	23	60%
Spring	39	23	59%
Summer	39	28	72%
Autumn	37	23	62%

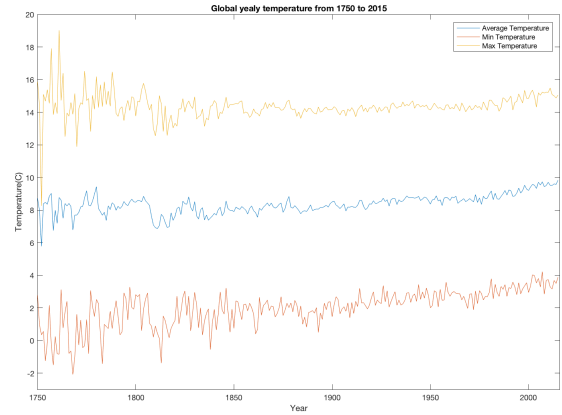
1750 to 2015. In this calculation process, we combined all the daily average data into a mean value each year in order to obtain the yearly average data.



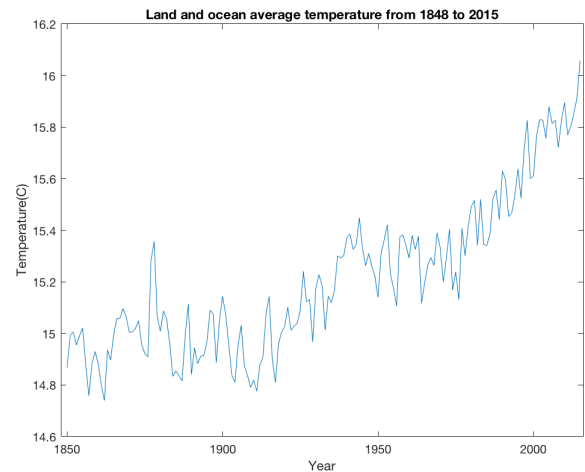
It is clear that the global yearly average temperature has a trend of increasing. Also the differences between each year became smaller and smaller.

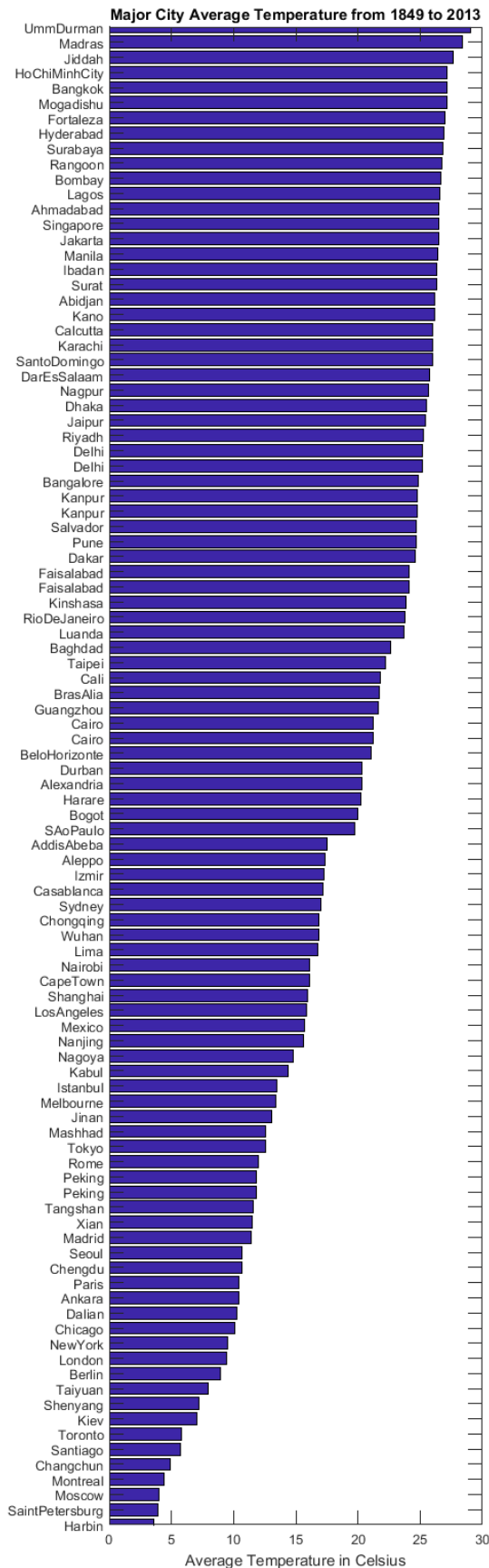


Moreover, we also generated graphs about the global maximum yearly temperature and minimum yearly temperature in the same time range. We can find both maximum and minimum temperature around 1750 was much lower than those temperatures of nowadays. This seems reasonable but we have to consider the fact that we have not include uncertainties in this graph and the uncertainties around 1750 are much greater according to the dataset. Moreover, when we combined these three graphs together, it was not hard to find that the global temperature had a trend of increasing overall. We could not conclude that global warming is a fact simply according to this analysis. However, these graphs provide us with the most directly overall idea about which factors need to be discovered in order to answer those questions we came up with.



Moreover, the land and ocean average temperature from 1848 to 2016 provided us with the same conclusion as the previous graphs while also proving a more significant change in the average temperature. In this case, we need to consider why the water would have greater change than the land. The specific heat capacity might cause this fact; however, we also need to consider other factors that might cause this difference.





The Histogram of Major City Average Temperature is the analysis of Average temperature of over one hundred cities from 1849 to 2013. Basically, we summarized the yearly average temperature of each major city from 1839 to 2013, which is about two hundred years, and then we used **Matlab** to calculate the average of their summation. The method we used is probably not very accurate because summarizing average temperature of major cities can just show the average temperature of different countries or even different continents. But we can use these outcomes to calculate the standard deviation of the average temperature of each major city or country or continents and then conclude which city or which country or which continent get the greatest variance. Besides, we can compare latitudes and longitudes of each major city to see how location influence the average temperature and we will try to figure out some specific coefficients. Finally, we will use this data set to calculate the average growth rate of the temperature of each major city by regressing the data to a linear function.

We all know the longitude and latitude of each major city inevitable associate the average temperature of each city. But there are something we usually ignore. For instance, the emission of greenhouse gas can not only damage the Ozone layer but also influence the global precipitation. The damage of the good Ozone layer will reduce the protection of lives on Earth from the sun's harmful ultraviolet (UV) rays and will increase the absorb of heat from the sun. On the other hand, the location of precipitation changes also influence the global surface temperature. For example, if the precipitation can't reach on the land, then the heat absorb from the sun could not be dissipate by evaporation. Therefore, these two aspects will be treated as the most important goal for as to find out in the next few weeks.

## REFERENCES

- [1] United States Environmental Protection Agency. Climate Change Science. (????).
- [2] NASA's Jet Propulsion Laboratory. 2017. Global Climate Change: Vital Signs of the Planet. (2017).
- [3] Berkeley Earth Data Website. 2016. Climate Change: Earth Surface Temperature Data. (2016).

## A THE BODY OF THE PAPER

### A.1 Problem Statement

### A.2 Literature survey

A.2.1 EPA: United States Environmental Protection Agency.

A.2.2 NASA: Climate Change and Global Warming.

### A.3 Proposed Work

### A.4 Data set

### A.5 Evaluation Methods

### A.6 Tools

### A.7 Milestones

### A.8 Results so far

### A.9 Reference