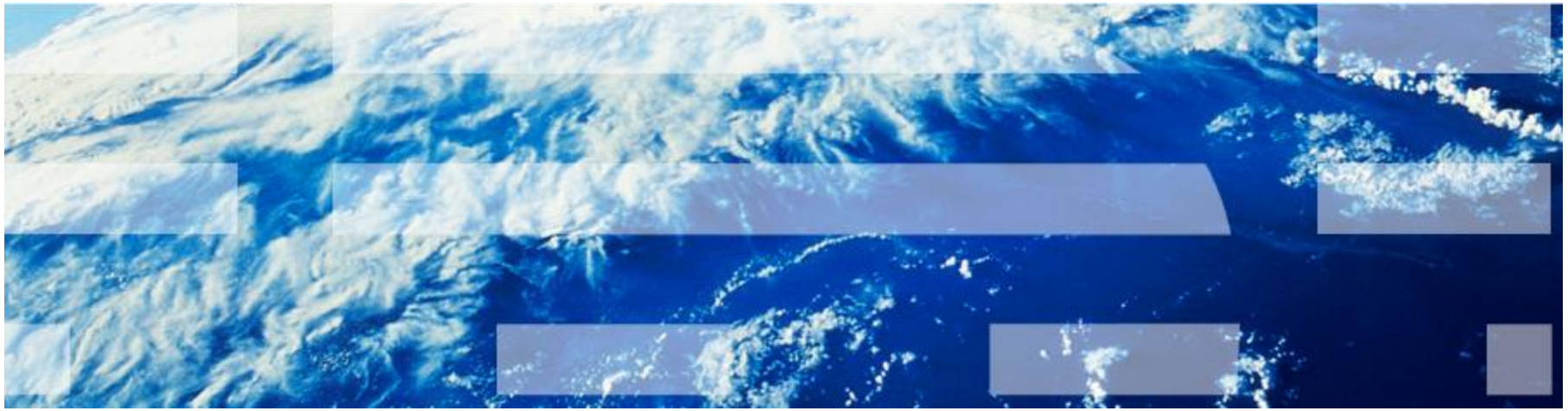
E6893 Big Data Analytics:

The Impact of Global Warming from Big Geographical Data

Team Members: Chandan Kanungo(ck2749)
Wei Zhang(wz2363)
Yizhou Shen(ys2840)



December 15, 2016

Outline

- Overview
- Technologies Used
- Dataset Overview
- Challenges / Struggles
- Architecture
- Dataset Visualization
- Sea Level Trend Visualization
- Demo
- Findings
- Conclusion
- Next Steps

Overview

The issue of Global Warming has been a controversial topic over the past decades. President Trump even claimed it is a made-in-China topic. The impact of global warming can truly be devastating to our planet. For example, 40% of the population in Netherlands are exposed to the risk of drowning. Growing sea level resulted from global warming can lead to submerging city's land like Manhattan. We used evidence from big geographical data and evaluated the impact of global warming. We predicted the global temperature and the resulting sea level trend around the whole world.

Technology Used

- Spark/Scala, ML-Lib

  We used Spark/Scala and ML-Lib to process data and produce models.

- Python/Matlab

  We used Python/Matlab script to handle NetCDF and RLR file and pre-process data.Python code is used to also find the location of the ocean among the entire lat/long data.

- Linear Regression

  We take the knowledge of linear regression as the foundation of training.

- Database Knowledge

  We used the concepts of cross-join and self-join to deal with each table of data.

Dataset Overview

- Climatology Data from NOAA

  https://www.nodc.noaa.gov/access/index.html

- Global Surface Temperature Data from NASA

  http://data.giss.nasa.gov/gistemp/

- Sea Level Trends Data from NOAA

  http://tidesandcurrents.noaa.gov/sltrends/sltrends.html

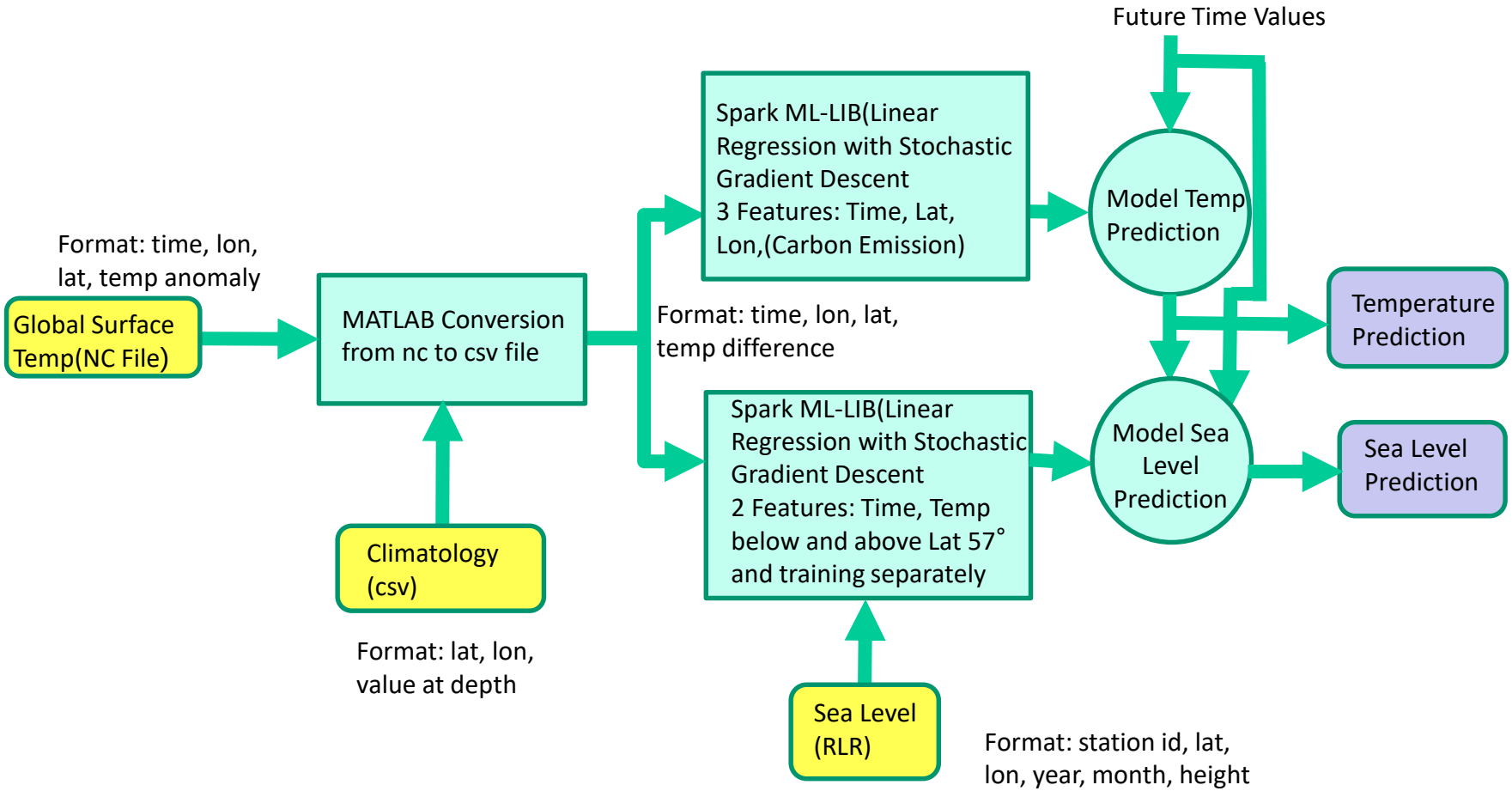
- Sea Level Trends Data From PSMSL

  http://www.psmsl.org/data/obtaining/

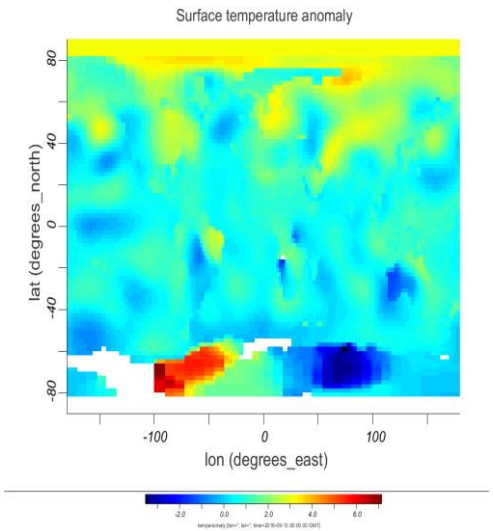
Challenges / Struggles

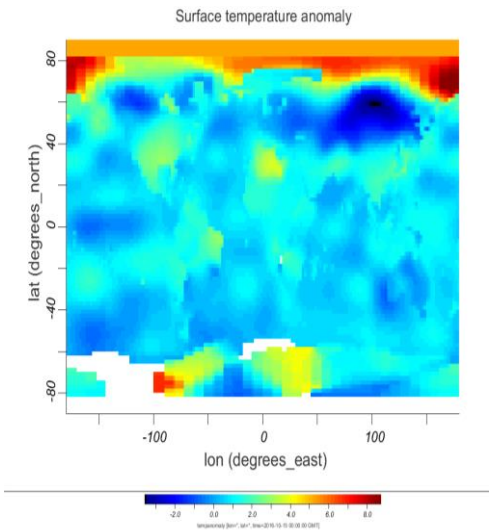
- Alignment of all dataset from different source. (Based on GPS location)
- Dealing with NetCDF file. Easy to visualize but hard to process.
- Prediction of temperature and regional relation between temperature and sea level.
- The size of data introduces some memory issues.

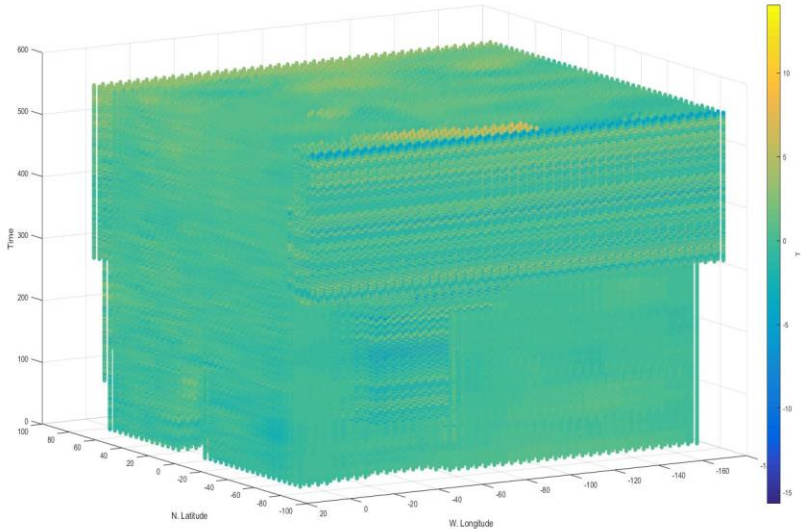
Architecture



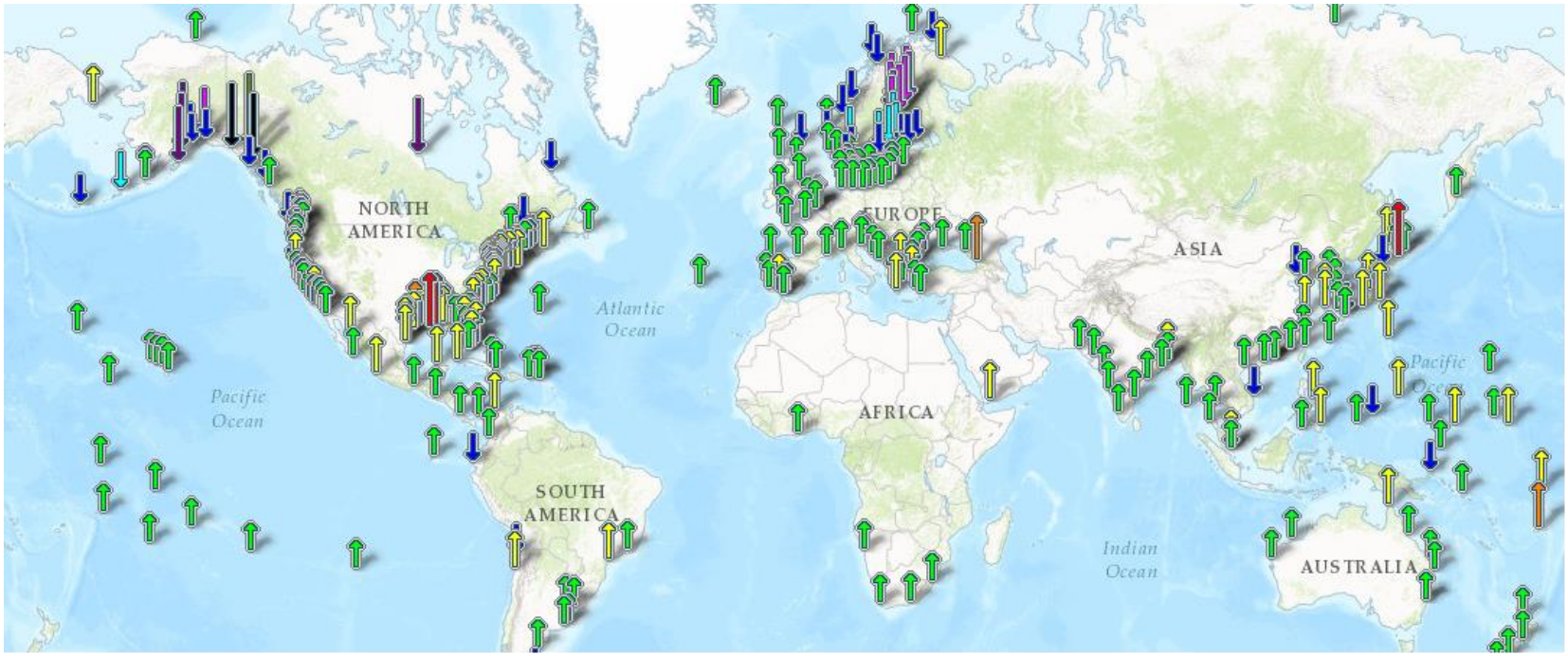
Dataset Visualization



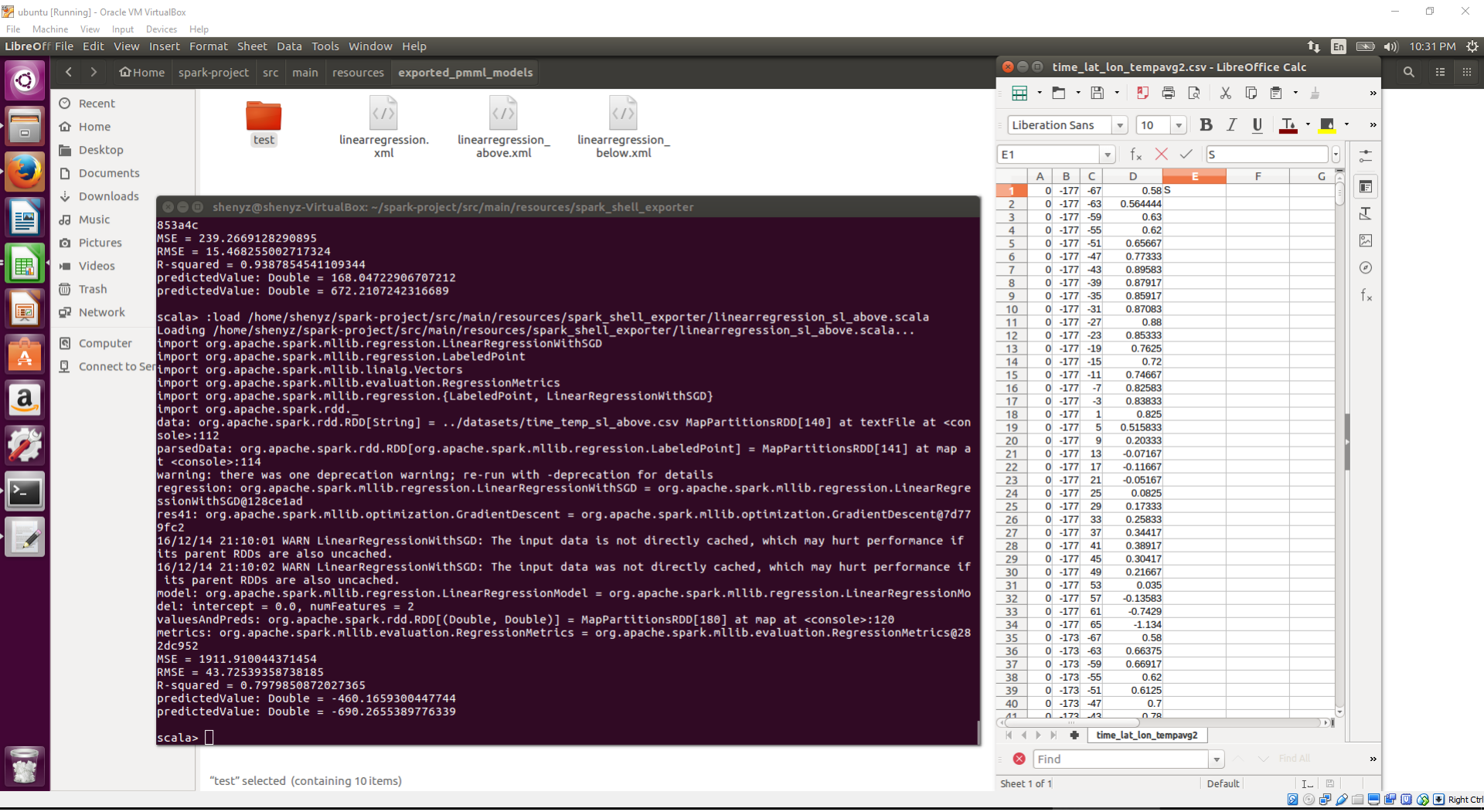




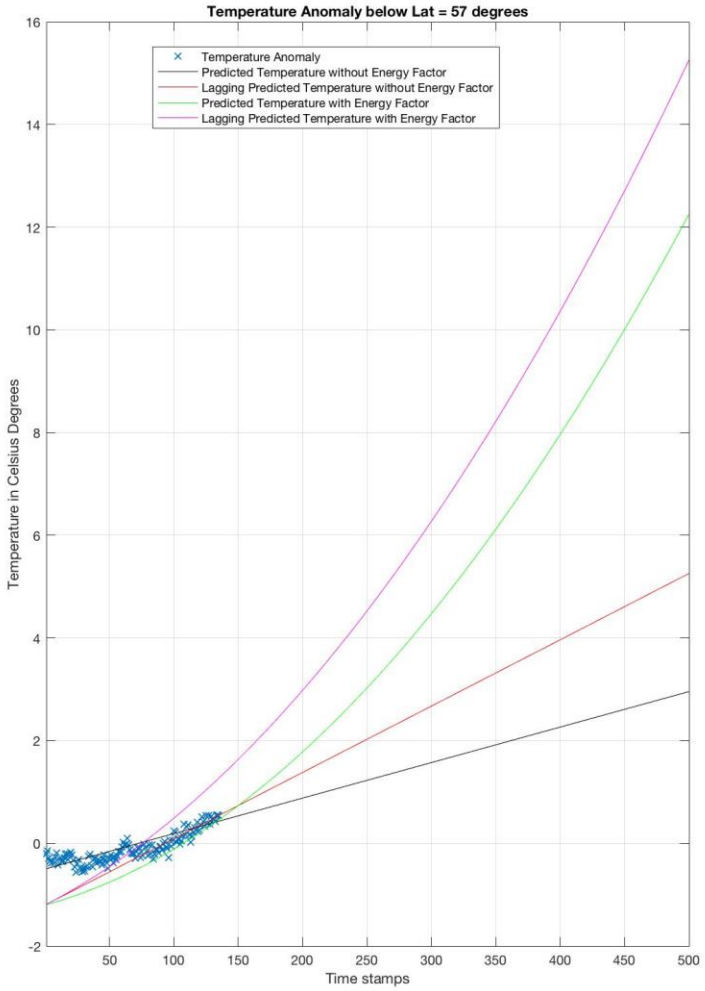
Sea Level Trend Visualization

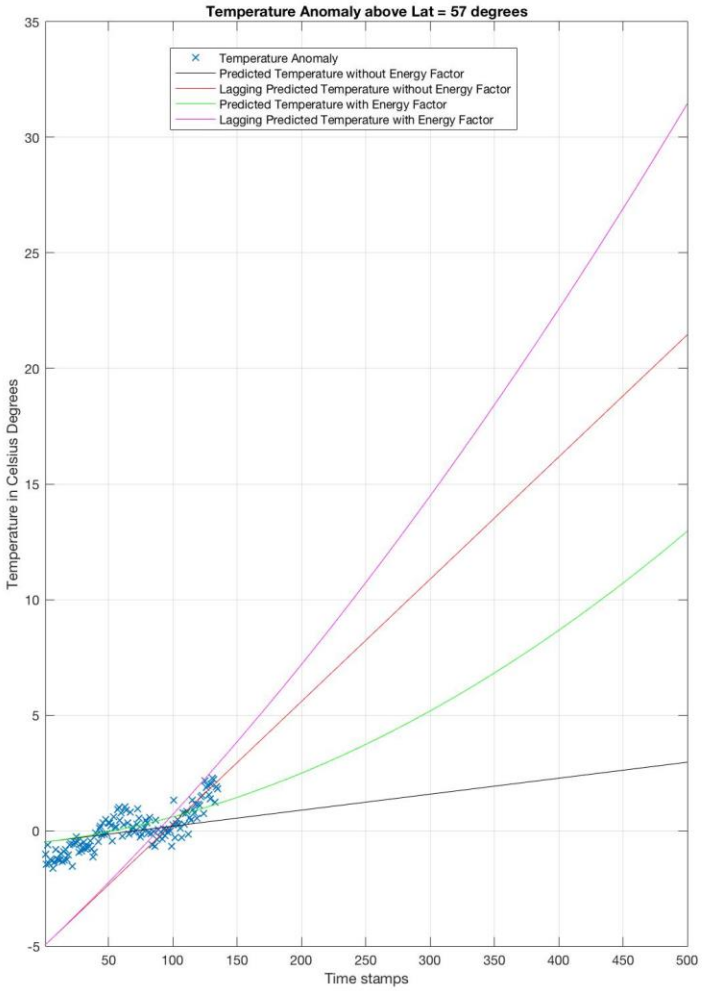


Demo

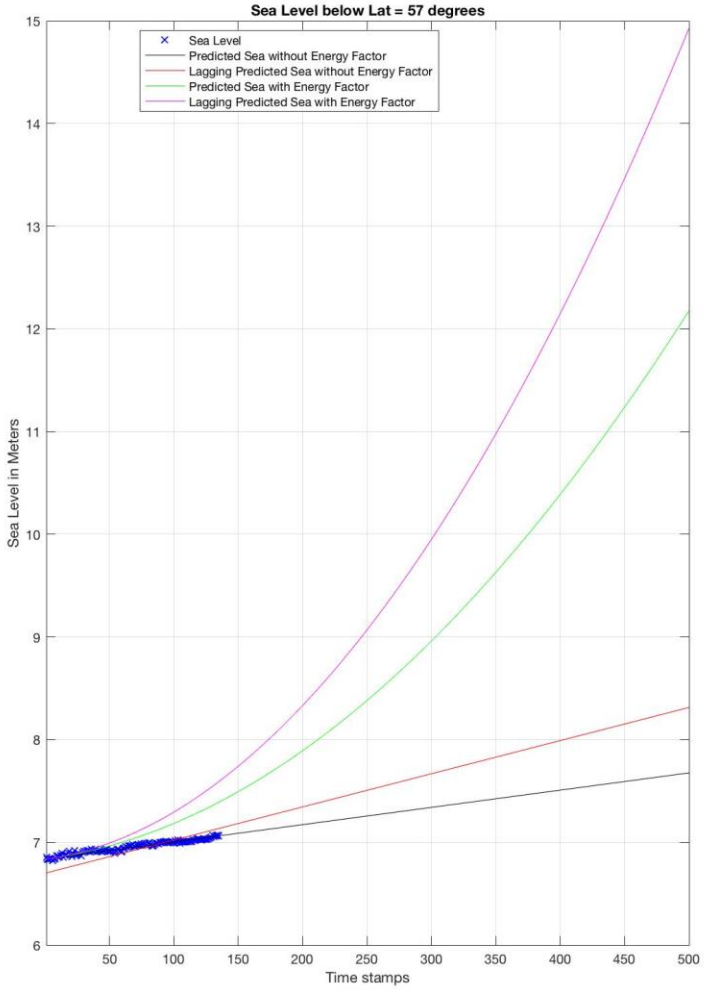


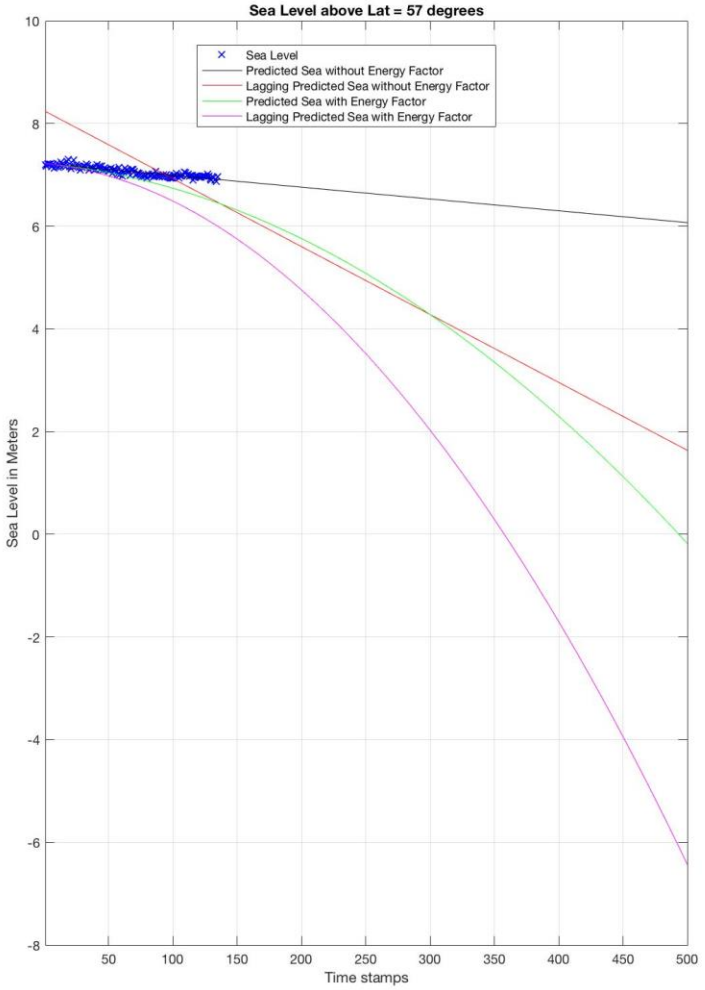
Findings—Temperature Anomaly





Findings—Sea Level





Conclusion

- By analyzing the data and using regression to predict the future values of temperature and sea level, it can be concluded that Mr President Elect should reconsider his stance as his statement do not aligns with the scientific data that has been generated by the National Oceanography agencies.

- The temperature will show an increase of an average by 3-15° C in the next 500 years for the latitude below 57 and for latitude above 57, the temperature will rise up to 30° C. Also, the temperature and other factors have lead to decrease of the sea level where Latitude is greater than 57° and increase of the sea level in rest of the planet. Over the next 500 years, the temperature and other factors will lead to a 2-8 meters decrease of the sea level where Latitude is greater than 57° C and up to 15 meters increase of the sea level in rest of the planet.

- Also, our project is of high commercial as well as moral values. We as humans have great responsibility towards the wellbeing of our planet as some call it “Mother Earth”, and without it we cannot have any future as a species. According to Mr. Stephen Hawking we have just 1000 more years on the planet. Huge commercial values must be associated on this area so that many people start valuing this fragile ecosystem that is still the only known life supporting cocoon in the entire universe.

Next Steps

- The impact from iceberg melt is not completely taken into account but should be.
- The consumption of natural resource should also be taken into account.
- Thus, the model should be more complex than what it is now.
  - It is possible a more complicated non-linear regression
  - Inference technique can be more sophisticated such as using variational inference
  - Model parameter can be fine-tuned better such as using cross-validation
- Datasets should be more comprehensive since currently we still miss some data due to accessibility issue.

Thank You