DSCI 550: Data Science at Scale

Fall 2025 Project

Project Proposal Report

Project Title

Forecasting and Trend Analysis of Sea Surface Temperature and Marine Weather Variables using NOAA ICOADS Time Series Data

Team members

1. Umaeshwer Shankar (umaeshwe@usc.edu)
2. Arin Paul (arinpaul@usc.edu)
3. Alex Zhang (alexyzha@usc.edu)
4. Ketan Totlani (totlani@usc.edu)

Project Idea

Using historical and modern **ICOADS** data, train and evaluate statistical and machine learning models that forecast the following key ocean atmosphere variables:

- **Sea Surface Temperature (SST),**
- **Air Temperature**
- **Wind Speed**

Description of dataset

The International Comprehensive Ocean-Atmosphere Data Set (**ICOADS**) is a global collection of ocean and weather data maintained by National Oceanic and Atmospheric Administration (**NOAA**). It combines information from ships, buoys, coastal stations, and other ocean platforms and consists of dated entries ranging from 1662 all the way up to the modern day. The dataset has over 30 million records and is frequently updated, meaning that it is an accurate representation of ocean conditions over time. This is the primary reason we chose it for our analysis.

Project Plan

1. Data Preprocessing
    - Query relevant ICOADS tables (SST, wind, pressure) via BigQuery Python client.
    - Filter by region and aggregate into regular intervals (daily/weekly/monthly).

- Handle missing data with interpolation or imputation and normalize variables.

2. Exploratory Analysis
    - Visualize trends, seasonal cycles, and autocorrelation.
    - Compute correlations and decompose time series into trend, seasonality, and residuals.
3. Modeling
    - Statistical Models: ARIMA/SARIMA for univariate SST forecasting.
    - Deep Learning Models: Prophet for capturing long-term dependencies.
    - Feature Engineering: Lag features, temporal features (month/season), and meteorological covariates.
4. Forecasting
    - Split data using rolling-window or time-based approach.
    - Evaluate with RMSE and generate 1-day, 7-day, and 30-day forecasts.
5. Visualization & Interpretation
    - Plot actual vs predicted values, residuals, and feature importance.
    - Create spatial maps of SST forecasts and compare model strengths and limitations.

Division of Work

Category	Assigned Member
Data Preprocessing	Arin Paul
Exploratory Analysis	Alex Zhang
Modeling & Forecasting	Umaeshwer Shankar
Visualization & Interpretation	Ketan Totlani

References

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control (4th ed.). John Wiley and Sons.

2. **Zhang, X., et al. (2020).** *Sea surface temperature prediction using deep learning methods. Remote Sensing*, 12(17), 2778.

3. **Mishra, N., & Ganguly, A. R. (2018).** *Long-term trends and variability in global ocean surface temperatures: A non-linear time series analysis approach.*

4. Adineh, A. H., Narimani, Z., & Satapathy, S. C. (2020). *Importance of data preprocessing in time series prediction using SARIMA: A case study*.

5. Gao, J., Hu, W., & Chen, Y. (2024). *Revisiting PCA for time series reduction in temporal dimension*