Forecasting Sea Surface Temperature and Marine Weather Variables with ICOADS Time Series Data

Umaeshwer Shankar, Arin Paul, Alex Zhang, Ketan Totlani

Problem Definition:

The objective of our work is to build robust forecasting systems for maritime variables using time series modeling techniques and provide a reliable method for commercial shipping vessels to effectively plan and navigate their paths.

Description of Background:

The global shipping industry, valued at approximately 600 billion USD in 2025 [1], is the foundation of international trade. It is often the only possible method for transporting critical goods, including energy and bulk commodities like oil and natural gas reserves, in addition to large volumes of grains, ores, metals, coal, and manufactured products. But these routes are prone to frequent disruption due to weather conditions, leading to large-scale material and human losses. Addressing this issue requires an accurate prediction of maritime variables which play a critical role in storm developments, and global climate patterns, and their accurate prediction allows the shipping industry to optimize shipping routes, improve safety, reduce disruptions, and enhance efficiency.

There have been multiple approaches to solve this problem in the past: [3] uses a modified Dijkstra's algorithm, a classic heuristic search method, to optimize minimal-time weather routes by incorporating wave data, ship speed reductions, and navigational constraints in a graph-based network,[4] uses historical AIS and weather data to approximate vessel weather routes, balancing speed, fuel, and risk without full dynamic programming, and [5] filters unstable trajectories (e.g., 1-2 trip outliers) before deep learning, improving accuracy in gravity-inspired models for global maritime networks. Though these methods have been widely implemented and have been successful for real-time prediction of maritime variables, they do not incorporate exogenous variables for forecasting. Our preferred approach for this problem is to use time series analysis which is complex enough to model granger-causal relationships and we will be using multiple methods with a critical examination of their pros and cons.

Description of Dataset:

The International Comprehensive Ocean–Atmosphere Data Set (ICOADS) is a globally recognized repository of marine environmental observations compiled from ships, drifting buoys, moored buoys, and other ocean-based platforms. It provides one of the most extensive historical collections of ocean–atmosphere measurements, making it highly reliable and suitable for long-term climate and marine forecasting studies. The ICOADS dataset contains over 800 million records spanning from the mid-1600s to 2017. In our project, we accessed the ICOADS dataset using Kaggle's public BigQuery interface and saved the filtered subset into our own BigQuery tables for further processing. For computational efficiency, we filtered the dataset to include records from 2005 to 2017. The ICOADS dataset contains several key attributes that are essential for forecasting sea surface temperature (SST) and understanding the marine atmospheric conditions that influence it. Sea Surface Temperature (SST) is the primary target variable for our forecasting task, serving as a fundamental indicator of ocean heat content and climatic behavior. Air temperature is closely related to thermodynamic variables that help explain short-term and seasonal fluctuations in SST, as they reflect the exchange of heat between the ocean surface and the atmosphere. Sea-level pressure (SLP) captures large-scale atmospheric circulation patterns that drive wind, storms, and regional ocean dynamics, making it an important predictor for SST variability at synoptic and seasonal timescales. Wind speed plays a crucial role in mixing the upper ocean, influencing evaporation rates, heat flux, and local SST changes; although moderately correlated, it provides additional predictive value for capturing dynamic weather-driven variability. Spatial attributes such as latitude and longitude allow us to examine regional differences and construct geographically consistent time series, while timestamp fields (year, month, day) support seasonal decomposition and long-term trend analysis. These attributes collectively capture the physical processes of thermal exchange, atmospheric forcing, ocean mixing, and spatial variability that directly shape SST behavior. By leveraging these variables, the dataset provides the necessary information to address the core questions of the project: how SST evolves over time, how atmospheric and oceanic conditions influence SST, and how accurately future SST values can be forecast using multivariate time series models.

The central aim of our project is to leverage the comprehensive NOAA International Comprehensive Ocean-Atmosphere Data Set (ICOADS) to analyze and forecast SST and other essential marine weather variables like air temperature and wind speed.

Description of Methods:

Considering the size of the dataset and operations, we have performed the analysis entirely using Google BigQuery. BigQuery also allows us flexibility with data formats and has Colab notebooks built into it. We performed our data collection and selection through SQL operations and analysis through Python. Some of the packages and framework we used are: Pandas, Matplotlib, Statsmodels and PyTorch.

We first aggregated the data from BigQuery public datasets and filtered the records between 2005 and 2017 for schema consistency and partitioned the table by year for read optimization. We also performed exploratory data analysis to better understand the relationship between the variables, the details of which will be provided in the next section. Moving onto the modeling part, we have divided our time series analysis into three classes of methods: statistical modeling, recurrent neural networks and transfer learning.

a. Statistical Modeling

During our exploratory data analysis, we observed that there are significant null counts in the exogenous variables. This made us question whether we can use auto-regressive methods to predict maritime variables (in specific, Sea Surface Temperature). Auto-regressive methods use lags and differencing to model the data and make predictions. They can be good baseline models since they can be built quickly and can model temporal dependencies.

b. Recurrent Neural Network Architecture

To capture nonlinear temporal dependencies in sea surface temperature (SST), we implemented a recurrent neural network (RNN) approach using a Long Short-Term Memory (LSTM) architecture. LSTMs are well suited for geophysical and environmental time series because they can represent long-range dependencies and mitigate vanishing-gradient issues commonly encountered in standard RNNs.

c. Transfer Learning

As part of our open-source modeling approaches, we implemented Facebook's Prophet model to forecast sea surface temperature. Prophet is a decomposable time series model that represents observations as an additive combination of trend, seasonality, and irregular components. Its design emphasizes interpretability, robustness to missing data, and strong performance on data exhibiting multiple seasonal patterns.

Experiment:

a. Exploratory Data Analysis

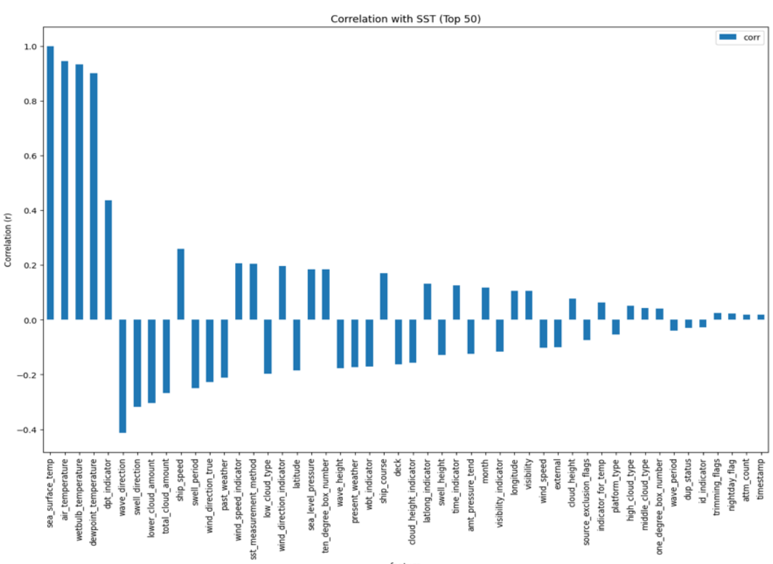

Figure 1:Top 50 variables most correlated (by magnitude) with SST

The dataset was first inspected to understand its schema, after which audit-related columns that were not relevant to the analysis were removed. Next, features exhibiting high correlation with SST were dropped (Fig. 1), including all other temperature variables except "dpt_indicator", which represents the dew point temperature at which air becomes saturated and water vapor starts to condense. Patterns of missingness were then explored using a pairwise missingness dependence plot (Fig. 2) and a missingness pattern heatmap (Fig. 3). Figure 2 indicates that the missingness of each variable depends on at least one other variable, implying that the data are not missing completely at random (MCAR).

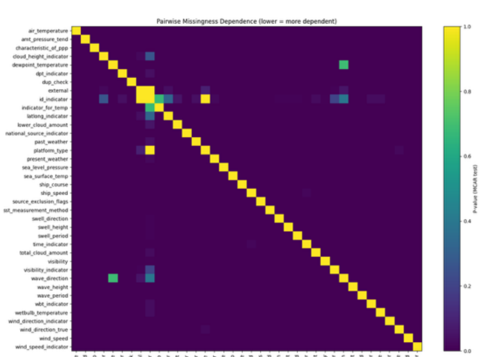

Figure 2:Pairwise MCAR test matrix

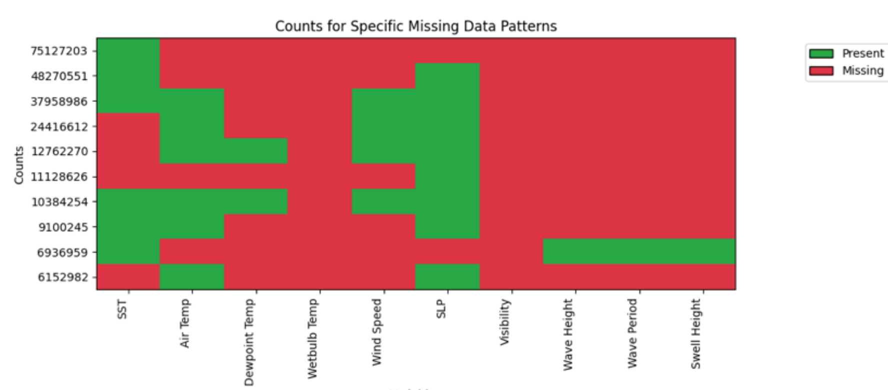

Figure 3:Patterns of missing data

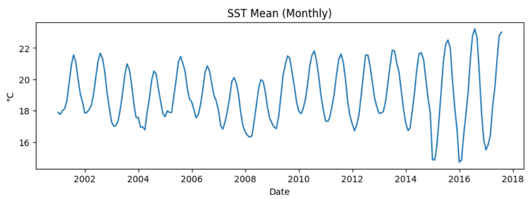

Figure 4(a): SST monthly mean

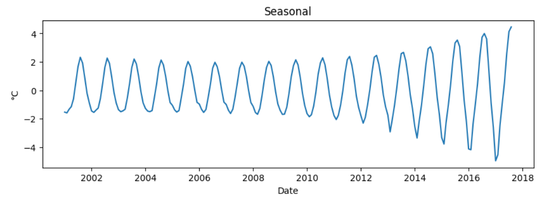

Figure 4(b): SST seasonal plot

Temporal structure in the sea surface temperature series was examined using time-series plots (Fig. 4). Finally, univariate and bivariate analyses were conducted to assess distributional properties of the variables.

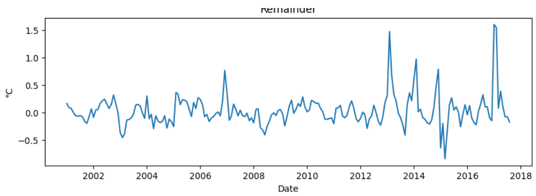

Figure 4(c): SST remainder time series plot

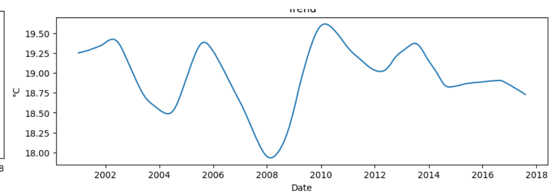

Figure 4(d): SST trend plot

For instance, wind speed exhibited a skewed distribution, so a log-normal transformation was applied to approximate normality, and the resulting distribution was checked for consistency with the earlier correlation structure (Fig. 5).

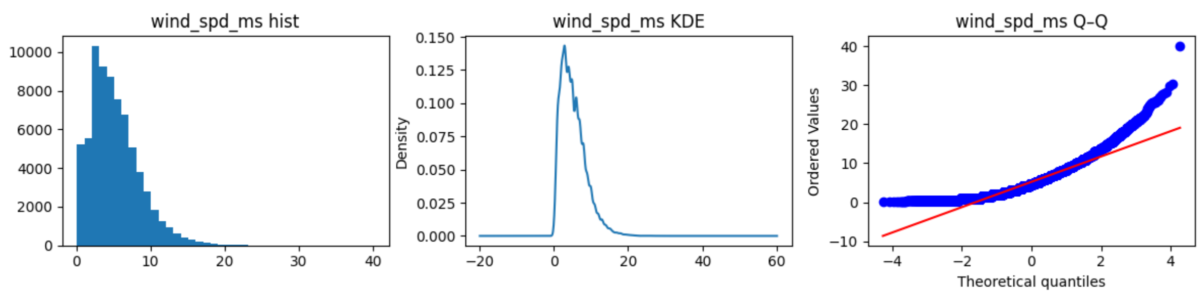

Figure 5: Distribution and QQ plot for wind speed

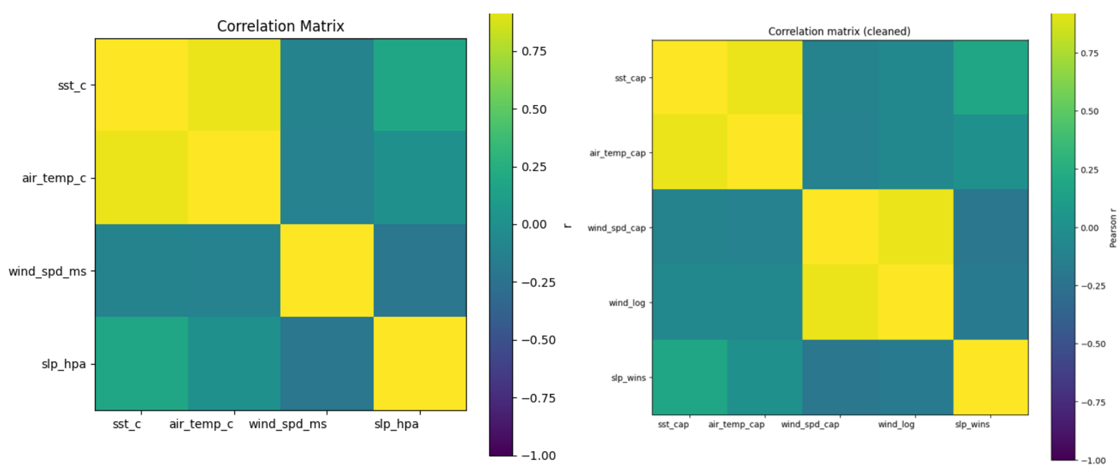

Figure 6: Correlation matrix between variables before (left) and after (right) log-transformation

Since there was a high correlation between SST and other temperature variables, we dropped these variables in order to remove bias in the model.

b. Statistical Modeling

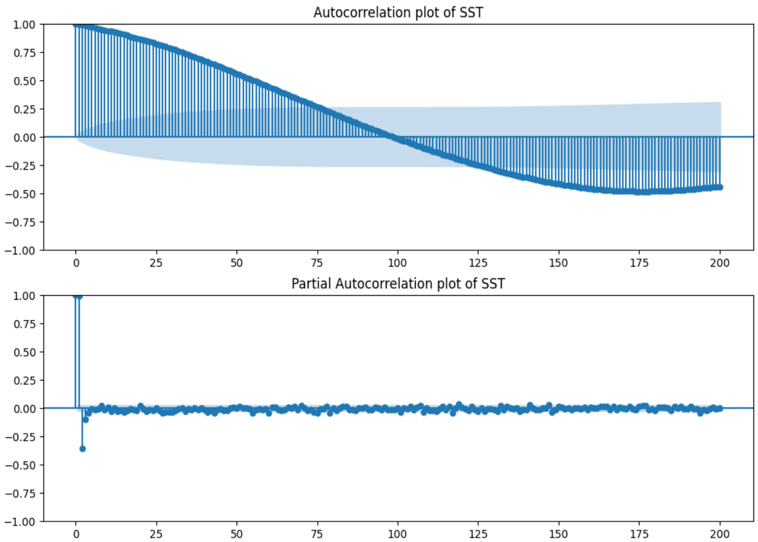

Figure 7: Autocorrelation plots of SST

Since latitude and longitude are not time-dependent variables, we queried the records to find out the Top 5 locations where SST readings were taken and zeroed in on the best location (37.81 N, 122.47 W) for the analysis. We then checked for null counts in the exogenous variables and identified that there is a significant margin of nulls in them. This prompted us to use univariate autoregressive models for analysis, which only require the predictor variable and the timestamp. An analysis of the data revealed that the data is available at an hourly granularity and there is no significant difference for these readings every 24 hours. So, we resampled the dataset to change the granularity to daily.

We then looked at the autocorrelation and partial autocorrelation plots to understand the temporal dependencies in the data (Fig. 7). While the partial autocorrelation plots die down to zero after the first few lags, the autocorrelation plots suggest another story. So, we need to take the differences of SST to observe patterns.

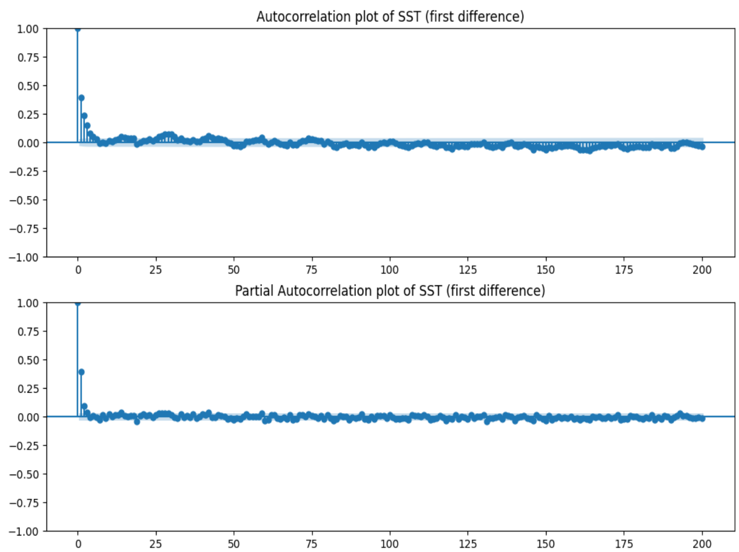

Figure 8: Autocorrelation plots of SST (first difference)

We further confirm this with the autocorrelation plots for the first difference of SST. This time we see a clear impact of autoregressive models as the correlations drop to zero after the second reading. So, we used an AR model with p = 1 or 2 and d=1.

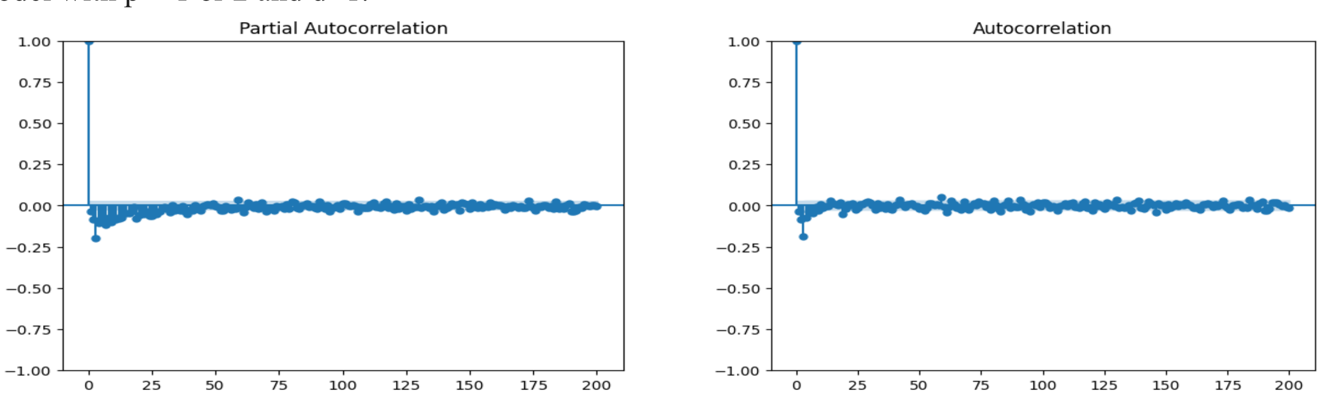

Figure 9: Autocorrelation plots of the residuals

We compared the results of both the models and chose AR(2,1,0) as it performed better on the parameters log-likelihood, AIC, BIC and HQIC. The autocorrelation plots of the residuals also revealed that there is no pattern in the residuals, which means that the autoregressive model fit the temporal dependencies well (Fig. 11).

c. Recurrent Neural Network Architecture

For the deep learning–based forecasting approach, a Long Short-Term Memory (LSTM) network was used to model nonlinear temporal dependencies between sea surface temperature (SST) and key atmospheric variables at the selected location (37.81 N, 122.47 W). The input features included latitude, longitude, sea level pressure, wind speed and cyclic encodings of month and hour (sine and cosine terms), while SST served as the prediction target. Before modeling, missing values were handled with forward–backward fill, all features were coerced to numeric types, and both inputs and target were standardized using a StandardScaler to stabilize training and make the network less sensitive to scale differences across variables.

To expose the LSTM to temporal context, the continuous time series was transformed into overlapping sliding windows of length 48 time steps, where each window's 48-step history of features was mapped to the SST value at the subsequent time step. The data was then split chronologically into training (70%), validation (20%), and test (10%) sets, and wrapped in PyTorch sequence datasets and data loaders to enable mini-batch training and efficient iteration over temporal sequences. This formulation allows the model to learn how short-term and intra-day variability in atmospheric conditions and cyclic time patterns jointly influence SST evolution.

The LSTM architecture consisted of a two-layer LSTM with 64 hidden units and dropout of 0.2, followed by a fully connected output layer that maps the final hidden state to a single-step SST forecast. Training used mean squared error (MSE) as the loss function and the Adam optimizer with a learning rate of 0.001, run for 10 epochs with mini-batches of size 64 while monitoring validation loss each epoch to track generalization. The evolution of training and validation loss showed gradual reduction in training error with relatively stable validation loss, indicating that the model was able to learn temporal structure without severe overfitting over this horizon.

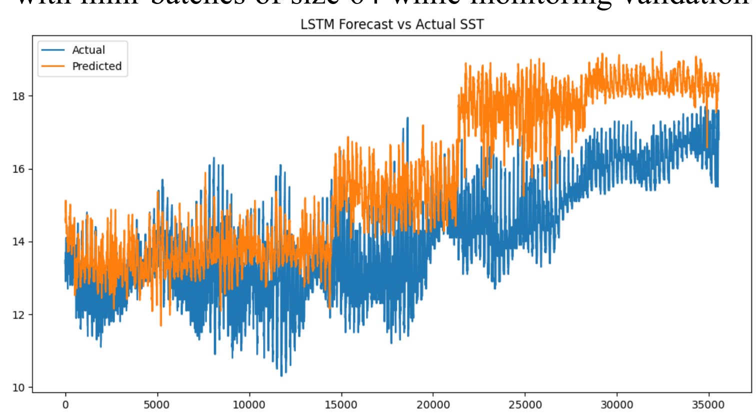

Figure 10: Actual v/s predicted values for SST using LSTM

On the held-out test set, predictions were inverse-transformed back to the physical SST scale using the fitted target scaler and evaluated using mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The final LSTM achieved an MSE of approximately 4.19, an RMSE of about 2.05, and an MAE of about 1.71, demonstrating that the recurrent deep network can capture nonlinear relationships and produce competitive single-step SST forecasts relative to simpler autoregressive baselines. Visual comparison of the predicted and actual SST trajectories as shown below, on the test period shows that the LSTM closely tracks the main temporal patterns, including local fluctuations, though some high-variance excursions remain under or over-estimated, reflecting the inherent complexity and noise in marine weather dynamics.

d. Transfer Learning

To complement the autoregressive and LSTM-based forecasting approaches, we also implemented a Prophet model to capture trend, seasonality, and holiday-like effects in the sea surface temperature (SST) series at the selected location (37.81° N, 122.47° W). Prophet is well-suited to geophysical time series because it decomposes the signal into an interpretable sum of components trend, seasonal cycles, and irregular fluctuations while remaining robust to missing data and outliers.

As with the statistical and neural models, the SST records were resampled to daily frequency to remove high-frequency noise and stabilize the temporal structure. The time series was reformatted into Prophet's required schema with the timestamp stored in a `ds` field and SST stored in a `y` field. No external regressors were included due to the significant volume of missingness observed in the atmospheric covariates during the exploratory analysis, matching the assumptions used in the autoregressive analysis.

Prophet automatically estimates a smooth nonlinear trend using piecewise linear segments. To accommodate the seasonal cycles visible in the SST decomposition plots, yearly seasonality was enabled by default, and weekly seasonality was disabled due to the lack of weekly periodic structure in ocean temperatures. Additional Fourier terms were allowed to ensure sufficient flexibility in capturing the annual SST oscillation. The model was trained on chronologically earlier observations (2005–2014) and evaluated on the (2015-2017) period reserved as the test set, consistent with the LSTM model splitting strategy.

Quantitatively, the Prophet model achieved an RMSE of approximately 4.51, an MAE of 3.85, and an MAPE of 28.68 on the test set. These results place Prophet between the autoregressive benchmark and the LSTM model in terms of predictive accuracy: it outperformed the AR(1,1,0) model on long-term seasonal behavior but lagged behind the LSTM in capturing rapid short-term variability driven by atmospheric dynamics.

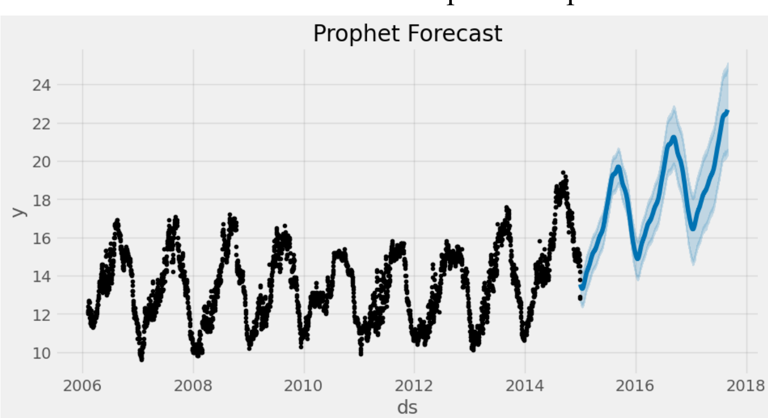

Figure 11: SST Forecast from Prophet Model

The visual comparison of the predicted and actual SST curves for 2015-2017 showed that Prophet closely tracks the annual rise and fall of SST but underestimates some of the sharper local oscillations observed in the real data as shown below. This behavior is expected Prophet's smoothing prior favors interpretable seasonal components rather than highly reactive short-horizon dynamics. The confidence intervals widen appropriately during parts of the year with greater variance, indicating that the uncertainty estimates are consistent with the underlying heteroscedasticity of the SST series.

Overall, Prophet provides a strong interpretable baseline that successfully models long-range trend and seasonality in SST. While less sensitive to short-term fluctuations than the LSTM, its decomposition into identifiable components makes it a valuable diagnostic tool and a complementary model to the autoregressive and deep learning approaches.

Observations and Conclusions:

Across the full analysis, several key patterns emerged regarding the behavior and forecastability of sea surface temperature (SST) within the ICOADS dataset. Exploratory data analysis revealed strong correlations between SST and other temperature-related variables, prompting their removal to prevent multicollinearity, while patterns of structured missingness led to restricting later models to univariate SST data. Temporal decomposition showed pronounced annual seasonality and a smooth long-term trend, supported by autocorrelation plots that indicated strong lag dependence and the need for first differencing to achieve stationarity. Statistical calculations using ACF/PACF guided the selection of an AR(2) model with d=1, which provided a well-behaved residual structure and competitive baseline performance. In contrast, the LSTM architecture leveraged multivariate predictors and sliding-window sequences to model nonlinear temporal dynamics, achieving an RMSE of roughly 2.05 after scaling corrections and demonstrating superior short-term responsiveness. Prophet's decomposable framework captured annual seasonality effectively and produced moderate accuracy (RMSE ~4.51), though it smoothed rapid local fluctuations. Overall, the calculations across models consistently showed that SST is both strongly seasonal and moderately predictable, with simpler statistical models capturing long-term structure and deep learning methods providing more accurate short-horizon forecasts despite increased complexity.

References:

[1] "Global Maritime Freight Transport Market | 2022 - 27 | Industry Share, Size, Growth - Mordor Intelligence," www.mordorintelligence.com. https://www.mordorintelligence.com/industry-reports/global-maritime-freight-transport-market

[2]G. Houtven, M. Gallaher, J. Woollacott, and E. Decker, "Act Now or Pay Later: The Costs of Climate Inaction for Ports and Shipping," 2022. Available: https://www.edf.org/sites/default/files/press-releases/RTI-EDF%20Act%20Now%20or%20Pay%20Later%20Climate%20Impact%20Shipping.pdf

[3]D. Sen and C. P. Padhy, "An approach for development of a ship routing algorithm for application in the North Indian Ocean region," *Applied Ocean Research*, vol. 50, pp. 173–191, Mar. 2015, doi: https://doi.org/10.1016/j.apor.2015.01.019.

[4]C. Gkerekos and I. Lazakis, "A novel, data-driven heuristic framework for vessel weather routing," *Ocean Engineering*, vol. 197, p. 106887, Feb. 2020, doi: https://doi.org/10.1016/j.oceaneng.2019.106887.

[5]R. Song, G. Spadon, R. Pelot, S. Matwin, and A. Soares, "Enhancing global maritime traffic network forecasting with gravity-inspired deep learning models," *Scientific Reports*, vol. 14, no. 1, Jul. 2024, doi: https://doi.org/10.1038/s41598-024-67552-2.