Amrita Vishwa Vidyapeetham

Amrita School of Computing

Technical Report

Anomaly Prediction in Sea Surface Temperature using Time-Series Models

Team :

Roll No : AM.EN.U4AIE20126, Name: Gayathri P

Roll No : AM.EN.U4AIE20134, Name: Jayanth Santosh Varma

Roll No : AM.EN.U4AIE20159, Name: Rahul Yalavarthi

Project Guide: Dr.Gopakumar G

Project Coordinator: Dr.Akshara

Signature :

Signature :

November 20, 2023

Contents

Abstract

Forecasting oceanic phenomena is crucial for ensuring maritime safety, managing coastal areas, and being prepared for disasters, including severe weather events and tsunamis. Accurate measurements of parameters like sea surface height (SSH) and sea surface temperature (SST) are essential for developing ocean forecasting models. In the past, SST values were determined using methods such as bathymetric charts or on-site measurements taken by ships and buoys. However, these methods have several limitations, including limited spatial coverage, sparse temporal resolution, and the need for significant manpower and hardware. Nowadays, advanced satellite sensors are used to determine SST values, and anomalies in these values are predicted using various deep-learning techniques. This research has first delved into the usage of basic time series models like ARIMA and SARIMA, later on moving to a Long Short-Term Memory (LSTM) neural network to predict anomalies in temporal SST values, this model is later tested in comparison with another research and it shows that our model outperforms theirs. The performance metrics of our model are MSE(7.07×10^{-5} in the Gujarat Coastal Region), RMSE(0.01 in the Gujarat Coastal region), and MAPE(0.299 in the Korean Coastal region).

1 Introduction

The forecasting of oceanic phenomena is a critical component in ensuring maritime safety, managing coastal areas, and fortifying disaster preparedness measures. Forecasting anomalies in different oceanic factors like Sea Surface Temperature(SST), Sea Surface Height (SSH) is of great importance in this realm. Understanding these anomalies are essential to create and develop accurate ocean forecasting models that helps in anticipating climatic fluctuations, weather patterns, habitat preservation, erosion management, and effective disaster mitigation methods in coastal locations. The proposed model in this research work focuses on forecasting anomalies in Sea Surface Temperature.

Over the years, SST has exhibited an upward trend, with an average rise of 0.14°F per decade from 1901 to 2020[1]. This warming trend has resulted in elevated sea levels[2],posing a special threat to coastal nations like Bangladesh, China, India, the Netherlands, and others. For instance, warmer waters in the Indian Ocean have escalated the intensity of cyclones, hurricanes, and storm surges, as evidenced by the catastrophic Super Cyclone Amphan in 2020.

The catastrophic tropical cyclone Amphan, made landfall in May near the India-Bangladesh border, left a devastating mark on the North Indian Ocean region. According to the 'State of the Global Climate 2020' study, it was the most expensive tropical cyclone ever recorded in this region, causing enormous economic damages of almost USD 14 billion in India alone[3]. 2020 is one of the three warmest years on record, with a global average temperature that is almost 1.2 degrees Celsius higher than pre-industrial levels, despite the moderating La Niña event.The Cyclone Amphan and its

effects on the area were significantly connected to the higher Sea Surface Temperature (SST) that year. Increased sea surface temperatures (SSTs), a result of warming seas, are recognised to play a role in the intensification of tropical storms. In the case of Amphan, the Bay of Bengal's warmer SSTs supplied the energy needed for the cyclone to intensify and reach its full destructive power upon landfall. Even though millions of people in India and Bangladesh were evacuated during the storm, the cyclone resulted in a tragic loss of 129 lives across both countries,massive infrastructure damage and extensive flooding, highlighting the relentless force of nature and the vulnerabilities of coastal communities in the face of such extreme weather events[3].
Another notable cyclone that is linked to an increased sea surface temperature is the Tropical Cyclone Yasi (TC Yasi)[4]. It is considered one of the most powerful tropical cyclones to strike the Queensland coast, made landfall on February 3, 2011, near Mission Beach, as a category 5 cyclone. Its occurrence during a period marked by SSTs in the Coral Sea that were 1–2 degree celcius higher than average due to a La Niña event. This raised questions about the influence of warmer SSTs on its track, intensity, size, and the associated storm surge and rainfall[4].
Motivated by these broad-reaching implications of SST changes on coastal regions, our initiative seeks to understand, predict, and prepare for the potential impacts.Leveraging technological advancements, particularly satellite measurements, allows for the acquisition of precise and real-time SST data. However, comprehending the complex fluctuations in SST demands sophisticated modeling techniques. In the existing works in this domain, there are several persisting challenges to be addressed among which an important one is the need to consider seasonality while predicting anomalies. Moreover, navigating the landscape of specialized frameworks posed a notable challenge. Several frameworks utilized in seminal works were inaccessible for free public use. In this work, we have utilized three time-series models including two statistical models - Autoregressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average(SARIMA) and a deep learning LSTM neural network.
The contributions of this work are:

- An Accurate Model for Predicting Anomalies in Sea Surface Temperature (SST): Developing an accurate model for predicting anomalies in sea surface temperature is a fundamental contribution of the project. By accurately forecasting these anomalies, the model aids in providing insights into potential variations or irregularities in sea temperatures.

- Crucial Precautions in Coastal Regions During Disasters: The accurate prediction of SST anomalies has significant implications for disaster preparedness and risk mitigation in coastal regions. SST anomalies often precede or accompany severe weather events such as tropical cyclones, hurricanes, or storms. By being able to forecast these anomalies accurately, authorities and coastal communities can take proactive measures and precautions well in advance of impending disasters. This includes timely evacuations, reinforcement of infrastructure, allocation of resources, and coordinated emergency responses.

- Addressing the Disparity Between Anomaly Prediction and Detection Research: While anomaly detection (identifying anomalies after they occur)has garnered more attention than anomaly prediction. By emphasizing prediction over detection, the project not only fills this research void but also underscores the importance of proactive forecasting and prevention in the domain of SST anomalies.

2 Literature Survey

Exploring the dynamics of oceans and their impact on climate involves examining various key elements, including sea surface height (SSH), sea surface temperature (SST), atmospheric pressure, density fluctuations and wind patterns. Analyzing these factors plays a pivotal role in comprehending how oceans function and their influence on weather patterns.
Yuan Zhou et al. in their work ”Multilayer Fusion Recurrent Neural Network for Sea Surface Height Anomaly Field Prediction” [5], introduces a novel approach called the Multilayer Fusion Recurrent Neural Network (MLFrnn) designed specifically for predicting Sea Surface Height Anomaly (SSHA) fields. The model proposed by the authors uses a multilayer fusion cell, an architectural unit that has the ability to effectively comprehend both temporal dependencies within the SSHA time series and spatial correlations across neighboring and remote regions. Unlike methods that forecast SSHA values for single points and require training the model multiple times for an entire area, the MLFrnn forecasts the entire SSHA map for a region with just a single model calculation, reducing storage consumption and retraining time.
”Anomaly Prediction With Hybrid Supervised/Unsupervised Deep Learning for Elastic Optical Networks: A Multi-Index Correlative Approach” by Hui Yang et al.[6], focuses on predicting anomalies in complex optical network environments using a hybrid deep learning approach. The research addresses the challenge of anticipating network anomalies in elastic optical networks (EONs), which are highly influenced by multiple indicators. the paper introduces a multi-index anomaly prediction scheme that involves three phases: selecting influential indicators using Spearman’s rank correlation coefficients, predicting indicator time series with LSTM neural networks, and establishing a model for anomaly classification.
Qi Shao et al. in ”Mid-Term Simultaneous Spatiotemporal Prediction of Sea Surface Height Anomaly and Sea Surface Temperature Using Satellite Data in the South China Sea” [7], proposes a data-driven method, empirical orthogonal function of multivariate, complete ensemble empirical mode decomposition, and a multilayer perceptron (MEOF-CEEMD-MLP), for simultaneous mid-term prediction of Sea Surface Height Anomaly (SSHA) and Sea Surface Temperature (SST) using satellite data in the South China Sea (SCS). While traditional numerical models have limitations in time validity and computational complexity, data-driven methods offer promise by leveraging historical data to make predictions. In the proposed model in this research, MEOF

Table 1: *Summary of the Related works*

Paper	Title/Year	Problem Addressed	Contributions	Limitations	Open Problems
1	Multilayer Fusion Recurrent Neural Network for Sea Surface Height Anomaly Field Prediction, 2022	Addresses the need for accurate and holistic prediction of SSHA in the South China Sea using a Multi-layer Fusion Recurrent Neural Network.	Introduces a novel MLFrnn model that captures spatial and temporal features to improve accuracy Forecasts the entire SSHA map for the region MLFrnn combines global and local spatiotemporal features.	1) The MLFrnn's computational complexity could hinder real-time applications. 2) The study focused on the South China Sea;generalization to other regions may require further investigation.	1) Investigate model optimization for faster inference. 2) Extend the MLFrnn to analyze SSHA fields in different ocean regions for broader applicability.
2	Anomaly Prediction With Hybrid Supervised/ Unsupervised Deep Learning for Elastic Optical Networks: A Multi-Index Correlative Approach, 2022	Addresses the challenge of predicting network anomalies in complex elastic optical network (EON) environments..	Proposes a multi-index anomaly prediction model that combines supervised and unsupervised deep learning techniques for EON. Utilizes correlative prediction and LSTM .	1) Limited dataset size may affect the generalization of the model. 2) The scheme's performance on highly dynamic network environments needs further investigation.	1)Use data augmentation for improving model.
3	Mid-Term Simultaneous Spatiotemporal Prediction of Sea Surface Height Anomaly and Sea Surface Temperature Using Satellite Data in the South China Sea, 2020	The primary objective of this paper is to tackle the difficulty of forecasting daily SSHA and SST in the marine environment for a period.	Proposes a novel data-driven method that combines empirical orthogonal function of multivariate, empirical mode decomposition,and a multilayer perceptron (MEOF-EMD-MLP). It takes into account the correlation, temporal and spatial relationship between SSH and SST.	The model may not be well-suited for predicting abrupt oceanic changes caused by external forces like storms, which can significantly impact the marine environment.	1) Developing a data-driven based ocean- atmosphere coupled model. 2) Extending the model's forecastingcapabilities beyond the 30-days period to address long-term predictions. 3) could also focus on incorporating external forces, such as wind conditions.
4	Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies, 2021	To evaluate the predictability of seasonal SST and SSH anomalies over the ice-free global ocean and to compare the ensemble-mean hindcast skill of a Linear Inverse Model (LIM) with that of the North American Multi-Model Ensemble (NMME) for the period 1982–2010.	Developed a global LIM that can predict monthly mean SST and SSH anomalies over the ice-free global ocean. The LIM's ability to capture predictable patterns of climate variability, such as ENSO, PDO, and AMO, contributes to its potential as a valuable tool for seasonal prediction.	1) The LIM used in the study is a coarse-grained model, which may limit its ability to capture fine-scale coastal ocean features and local variations. 2) The hindcast period used for evaluation (1982-2010) may be relatively short for comprehensive skill assessment.	1) Future work could explore the LIM's skill in predicting other essential oceanic and atmospheric variables. 2) Extending the hindcast period beyond 2010 could provide a more comprehensive assessment of the LIM's skill for longer-term climate modes and variability.
5	Deep-learning model for sea surface temperature prediction near the Korean Peninsula, 2023	To build an LSTM model which could provide 1-7 days predictions and would help in forecasting high water temperatures in Korean peninsula	Build an LSTM model with good performance which could forecast data upto seven days for the Korean peninsula region.	1) The model explores only the usage of LSTM for this purpose. 2) The model has reduced accuracy after a five day period	1) Can build a better model to accommodate better one day prediction results. 2) Can try to improve their model after the five-day span predictions.

facilitates understanding the correlation between SSHA and SST, establishing their temporal and spatial relationship. CEEMD addresses nonlinearity and nonstationarity, enhancing predictability across different scales. Finally, MLP is employed for spatiotemporal prediction.

The research ”Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies” by Sang-Ik Shin et al. [8],focuses on assessing the predictability of sea surface temperature (SST) and sea surface height (SSH) anomalies across global coastal regions, particularly North America. The study emphasize the LIM's (Linear Inverse Model) competency in predicting large-scale climate variability patterns like ENSO, PDO, and AMO up to nine months in advance. It demonstrates similar or superior performance to NMME in several areas, particularly in SSH predictions for the Atlantic. The study's insights also emphasize the relevance of the LIM in diagnosing deficiencies in coupled models and supporting coastal ocean seasonal prediction.

Hey-Min-Choi et al. [9] in their research ”Deep-learning model for sea surface temperature prediction near the Korean Peninsula”. This study addresses the escalating sea surface temperatures (SSTs) near the Korean Peninsula due to global warming

and proposes a predictive method using time series SST data and a Long Short-Term Memory (LSTM) network. The model identifies SSTs exceeding 28 °C, the threshold for high water temperature warnings set by the Korean government. Evaluation metrics include coefficients of determination (R2), root mean square error (RMSE), mean absolute percentage error (MAPE), and F1 scores. Results show high accuracy, with the 1-day prediction model achieving an R2 of 0.985, RMSE of 0.14 °C, MAPE of 0.38%, and an F1 score of 0.963. The 7-day prediction model demonstrates slightly reduced accuracy but remains effective with an F1 score of 0.739.
Summary of the background study is presented in Table 1.

3 Methodology

3.1 Dataset

There are two datasets that are used in this project. First is the sea surface temperature data across 40 years along the Gujarat Coastal region in India. This was a data collection according to a custom region from the databank Marine Copernicus. The features in this dataset are (Year, Month, Day, Mean temperature in kelvin, Mean temperature in Celsius, Mean temperature uncertainty, Fraction of sea ice covered ocean).
The second dataset chosen is the SST along the Korean coastal region. This dataset was used by Hey-Min-Choi et al. [9]. This dataset was chosen to asses our model performance compared to that of theirs. This dataset is of netcdf format and it is described as dimensions(sizes): (longitude(57), latitude(37), time(14976)), variables(dimensions): (float32 longitude(longitude) , float32 latitude(latitude), int32 time(time), int16 sst(time, latitude, longitude))

3.2 ARIMA

For the first dataset concerning SST of Gujarat Coast, the first approach used was the simpler ARIMA model [10]. Autoregressive Integrated Moving Average (ARIMA) models constitute a prominent class of time series analysis tools widely employed in diverse fields, ranging from finance to meteorology. At the heart of ARIMA lies a comprehensive framework for capturing and forecasting temporal dependencies in sequential data. The model comprises three key components: autoregressive (AR) terms, differencing operators, and moving average (MA) terms.
The AR terms account for the linear correlation between past and present observations, the differencing operators handle non-stationarity by transforming the time series into a stationary one, and the MA terms capture the influence of past white noise on the current state. This intricate interplay allows ARIMA to adeptly model a wide array of time series phenomena, making it a versatile tool for understanding and predicting complex temporal patterns.

Algorithm 1 ARIMA Algorithm

Input: Time series data Y_t, order of differencing d, order of autoregression p, order of moving average q
Output: ARIMA model parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q)$
Step 1: Differencing
for $i \leftarrow 1$ to d **do**
  $Y_t \leftarrow Y_t - Y_{t-1}$
end for
Step 2: Autoregression
for $i \leftarrow 1$ to p **do**
  Compute autocorrelation function (ACF) at lag i, denoted as ρ_i
  Fit autoregressive (AR) model with lag i, $\phi_i = \rho_i$
end for
Step 3: Moving Average
for $i \leftarrow 1$ to q **do**
  Compute partial autocorrelation function (PACF) at lag i, denoted as α_i
  Fit moving average (MA) model with lag i, $\vartheta_i = \alpha_i$
end for
Step 4: Model Fitting
Fit ARIMA model with parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q)$
Return: ARIMA model with parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q)$

One of the notable features of ARIMA is its adaptability to various data patterns through the selection of appropriate model orders. The user specifies the degree of differencing (d), the number of autoregressive terms (p), and the number of moving average terms (q). This flexibility enables analysts to tailor the model to the specific characteristics of the data under consideration. Despite its effectiveness, ARIMA does have limitations, particularly in handling non-linear and irregular patterns, which has led to the development of more advanced time series models. Nonetheless, ARIMA remains a foundational methodology in time series analysis, providing valuable insights into temporal dependencies and serving as a benchmark for more complex modeling approaches. The steps involved in an ARIMA model is illustrated in 1.

3.3 SARIMA

Further after ARIMA implementation, we observed that the seasonality parameter of the dataset can be utilized to further enhance the prediction, hence we shifted to the SARIMA approach where we took the seasonality parameter as 3 as it is the average number of months a season lasts in India.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model stands as a sophisticated extension of the classical ARIMA framework, specifically designed to capture and forecast time series data exhibiting seasonal patterns [11]. SARIMA in-

Algorithm 2 SARIMA Algorithm

Input: Time series data Y_t, order of differencing d, order of autoregression p, order of moving average q, seasonal differencing D, seasonal order of autoregression P, seasonal order of moving average Q, length of the seasonal cycle s
Output: SARIMA model parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q, \Phi_1, \Phi_2, ..., \Phi_P, \Theta_1, \Theta_2, ..., \Theta_Q)$
Step 1: Seasonal Differencing
for $i \leftarrow 1$ to D **do**
    $Y_t \leftarrow Y_t - Y_{t-s}$
end for
Step 2: Differencing
for $i \leftarrow 1$ to d **do**
    $Y_t \leftarrow Y_t - Y_{t-1}$
end for
Step 3: Autoregression
for $i \leftarrow 1$ to p **do**
    Compute autocorrelation function (ACF) at lag i, denoted as ρ_i
    Fit autoregressive (AR) model with lag i, $\phi_i = \rho_i$
end for
Step 4: Moving Average
for $i \leftarrow 1$ to q **do**
    Compute partial autocorrelation function (PACF) at lag i, denoted as α_i
    Fit moving average (MA) model with lag i, $\vartheta_i = \alpha_i$
end for
Step 5: Seasonal Autoregression
for $i \leftarrow 1$ to P **do**
    Compute seasonal autocorrelation function (SACF) at lag is, denoted as Φ_i
    Fit seasonal autoregressive (SAR) model with lag is, $\Phi_i = \rho_i$
end for
Step 6: Seasonal Moving Average
for $i \leftarrow 1$ to Q **do**
    Compute seasonal partial autocorrelation function (SPACF) at lag is, denoted as Γ_i
    Fit seasonal moving average (SMA) model with lag is, $\Theta_i = \Gamma_i$
end for
Step 7: Model Fitting
Fit SARIMA model with parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q, \Phi_1, \Phi_2, ..., \Phi_P, \Theta_1, \Theta_2, ..., \Theta_Q)$
Return: SARIMA model with parameters $(\phi_1, \phi_2, ..., \phi_p, \vartheta_1, \vartheta_2, ..., \vartheta_q, \Phi_1, \Phi_2, ..., \Phi_P, \Theta_1, \Theta_2, ..., \Theta_Q)$

tegrates additional parameters to account for seasonal trends, encompassing both autoregressive and moving average components, thereby enhancing its ability to model complex temporal dependencies. By incorporating seasonal differencing, SARIMA addresses periodic variations in the data, offering a robust solution for applications where seasonality plays a crucial role. This model is particularly invaluable in fields such as economics, meteorology, and epidemiology, where the interplay of seasonally influenced factors significantly impacts the observed time series. SARIMA's adaptability to diverse seasonal patterns and its capacity to capture both short-term fluctuations and long-term trends make it a valuable tool in the arsenal of time series analysts and researchers seeking accurate and comprehensive modeling of intricate temporal dynamics. The SARIMA algorithm is represented clearly in 2.

3.4 Long Short Term Memory model

Upon delving into these moving averages models for time series prediction, we chose to use Long Short Term Memory(LSTM) models since they are the best in time series prediction, for our model to be better we had to take a comparative model which could do LSTM predictions for a particular region, here came in the second dataset for comparative purposes. We have trained the LSTM model after all the values were normalized so that the predictions would be really accurate.

Algorithm 3 LSTM Cell

Input: Input sequence X_t, previous hidden state h_{t-1}, previous cell state c_{t-1}
Output: New hidden state h_t, new cell state c_t
Step 1: Input Gate
$i_t = \sigma(W_{ii}X_t + b_{ii} + W_{hi}h_{t-1} + b_{hi})$
Step 2: Forget Gate
$f_t = \sigma(W_{if}X_t + b_{if} + W_{hf}h_{t-1} + b_{hf})$
Step 3: Cell Gate
$\tilde{c}_t = \tanh(W_{ig}X_t + b_{ig} + W_{hg}h_{t-1} + b_{hg})$
Step 4: Update Cell State
$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$
Step 5: Output Gate
$o_t = \sigma(W_{io}X_t + b_{io} + W_{ho}h_{t-1} + b_{ho})$
Step 6: New Hidden State
$h_t = o_t \odot \tanh(c_t)$
Return: New hidden state h_t, new cell state c_t

Long Short-Term Memory (LSTM) networks represent a pivotal advancement in the realm of sequence modeling and temporal data analysis [12]. As a type of recurrent neural network (RNN), LSTMs excel in capturing intricate long-term dependencies within sequential data, mitigating the vanishing gradient problem that often hinders traditional RNNs. The LSTM architecture introduces memory cells equipped with gating mechanisms, enabling the network to selectively retain or discard information over extended time intervals. This unique capability allows LSTMs to effectively model and predict complex temporal patterns, making them particularly well-suited for a diverse array of applications, including natural language processing, finance, and healthcare.

The ability to learn and leverage both short-term and long-term dependencies grants LSTMs a distinctive advantage in capturing nuanced temporal relationships, positioning them as a cornerstone technology in the arsenal of researchers and practitioners working with sequential data at varying time scales. The working of LSTM model is depicted through an algorithm in 3.

4 Results

After thorough testing, the evaluation metrics decided were that of MSE and RMSE for the Gujarat Coastal Region and RMSE and MAPE Scores for the Korean Coastal Region.

1. MSE: The Mean Squared Error is a widely used metric in research for assessing the accuracy of predictive models. It quantifies the average squared difference between predicted and actual values, providing a measure of the overall model performance. The ARIMA model is found to have a MSE of 0.076 and SARIMA with a MSE OF 0.078 for the Gujarat SST dataset. On the other hand, LSTM model has a MSE of 7.07×10^{-5} for the Gujarat dataset and a value close to zero for the Korean Coastal Region Dataset.

2. RMSE: Root Mean Squared Error, an extension of MSE, offers an intuitive measure of prediction accuracy by taking the square root of the MSE. This metric is especially useful for providing interpretable results in the same units as the target variable, facilitating a more direct understanding of the magnitude of prediction errors. Both ARIMA and SARIMA models are found to have a RMSE of 0.28 while LSTM is found to have a RMSE of 0.01 for the Gujarat dataset.

3. MAPE: In research contexts, the Mean Absolute Percentage Error serves as a crucial metric for evaluating the accuracy of forecasting models, especially in fields such as time series analysis. MAPE calculates the average percentage difference between predicted and actual values, offering insights into the relative magnitude of errors. datasets and domains.

The evaluation metric values indicating the performance of the model with Gujarat Coast Dataset (Dataset 1) shown in Table.2. Table.3 shows the comparison of the performance of our LSTM model and the LSTM model used in [9] on the Korean Coastal Region dataset.

Fig.1, Fig.2 shows the visual plot of ground truth values and the forecasted SST values using ARIMA and SARIMA respectively for the SST value dataset along Gujarat Coastal Region. Fig.3 shows the performance of the LSTM model on Gujarat Coastal Region dataset and Fig.4 shows the performance of the model on the Korean Coastal Region dataset.

Table 2: *Performace Analysis of Models on Gujarat Coast Dataset*

Model Name	MSE	RMSE
ARIMA	0.076	0.28
SARIMA	0.078	0.28
LSTM	7.07×10^{-5}	0.01

Table 3: *Performace Analysis of our LSTM Model and [9]'s LSTM model on Korean Coast Dataset*

Model Name	MSE	MAPE
LSTM-based SST model 1 day prediction by Authors	0.119	0.362
LSTM-based model by us(1 day prediction)	$\sim$0.00	0.299

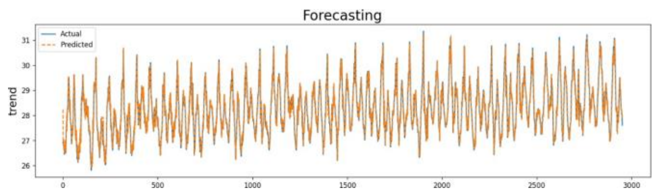

Figure 1: *Actual vs Predicted using ARIMA for Gujarat Coastal Region Dataset*

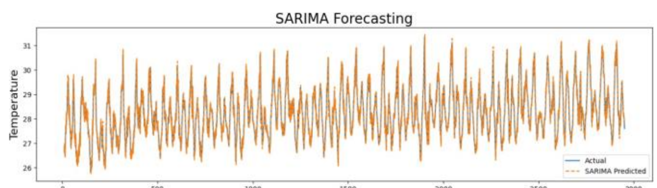

Figure 2: *Actual vs Predicted using SARIMA for Gujarat Coastal Region Dataset*

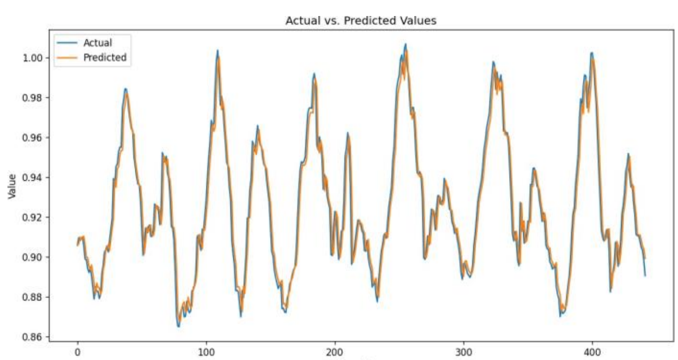

Figure 3: *Actual vs Predicted using LSTM for Gujarat Coastal Region Dataset*

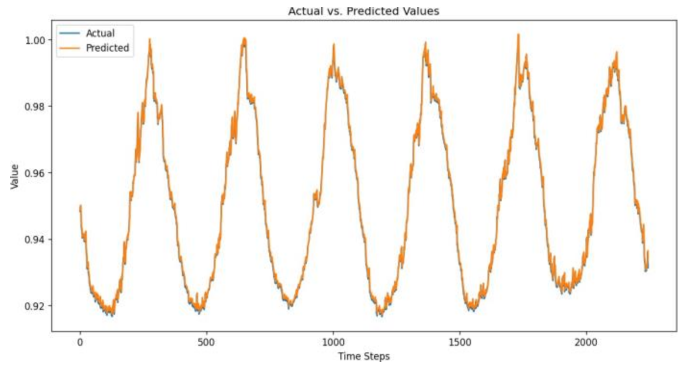

Figure 4: *Actual vs Predicted using LSTM for Korean Coastal Region Dataset*

5 Discussions

5.1 Gujarat Coastal Region

The results show that the predictions on the Gujarat coastal section were most accurately done by the LSTM model whilst the ARIMA and SARIMA are just a notch behind, the evaluation metric results in Table.2 conveys the same.

5.2 Korean Coastal Region

Our model outperformed the model done by Hey-Min-Choi et al. in [9] with significant improvement over single-day prediction scores which is depicted in Table.3. MSE obtained for our model was very close to 0 while the authors's model had a MSE of 0.119. MAPE score of our LSTM model on Korean Dataset is found to be 0.299 on the other hand, the MAPE of authors's LSTM model was 0.362.

5.3 Reason for better performance

The better performance of our model is due to normalization utilized for the dataset which made the calculations more efficient thus making the predicted values so accurate with the actual values.

6 Conclusion

This research shows our comprehensive exploration of anomaly detection in two distinct coastal regions, namely Gujarat and Korea, revealed notable insights into the predictive capabilities of various time series models. The LSTM model demonstrated superior accuracy in forecasting anomalies in the Gujarat coastal region, outshining traditional ARIMA and SARIMA models by a narrow margin. Additionally, our model surpassed the predictive performance of a prior study conducted by Hey-Min-Choi et al[9]. in the Korean coastal region, showcasing substantial advancements, particularly in single-day prediction scores. These findings underscore the effectiveness of LSTM in capturing complex temporal patterns inherent in coastal environmental data. However, the competitive performance of ARIMA and SARIMA models suggests their continued relevance, especially in scenarios with limited computational resources. This research contributes valuable benchmarks for anomaly detection in coastal regions, emphasizing the significance of choosing appropriate models based on the characteristics of the data.

References

[1] U. S. E. P. Agency. (2021) Climate change indicators: Sea surface temperature. [Online]. Available: https://www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature

[2] N. Mimura, "Sea-level rise caused by climate change and its implications for society," *Proceedings of the Japan Academy, Series B*, vol. 89, no. 7, pp. 281–301, 2013.

[3] T. E. Times. (2021) Cyclone amphan of 2020 resulted in 14*bneconomiclossesinindia* : *Unreport.*[*Online*].*Available* : *https* : *//economictimes.indiatimes.com/news/economy/indicators/cyclone* − *amphan* − *of* − 2020 − *resulted* − *in* − 14 − *bn* − *economic* − *losses* − *in* − *india* − *un* − *report/articleshow/*82157875.*cms*

[4] S. L. Lavender, R. K. Hoeke, and D. J. Abbs, "The influence of sea surface temperature on the intensity and associated storm surge of tropical cyclone yasi: a sensitivity study," *Natural Hazards and Earth System Sciences*, vol. 18, no. 3, pp. 795–805, 2018. [Online]. Available: https://nhess.copernicus.org/articles/18/795/2018/

[5] Y. Zhou, C. Lu, K. Chen, and X. Li, "Multilayer fusion recurrent neural network for sea surface height anomaly field prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[6] H. Yang, Y. Wan, Q. Yao, B. Bao, C. Li, Z. Sun, H. Wang, J. Zhang, and M. Cheriet, "Anomaly prediction with hybrid supervised/unsupervised deep learning for elastic optical networks: A multi-index correlative approach," *Journal of Lightwave Technology*, vol. 40, no. 14, pp. 4502–4513, 2022.

[7] Q. Shao, W. Li, G. Hou, G. Han, and X. Wu, "Mid-term simultaneous spatiotemporal prediction of sea surface height anomaly and sea surface temperature using satellite data in the south china sea," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020.

[8] S.-I. Shin and M. Newman, “Seasonal predictability of global and north american coastal sea surface temperature and height anomalies,” *Geophysical Research Letters*, vol. 48, no. 10, p. e2020GL091886, 2021.

[9] H.-M. Choi, M.-K. Kim, and H. Yang, “Deep-learning model for sea surface temperature prediction near the korean peninsula,” *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 208, p. 105262, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967064523000127

[10] R. H. Shumway, D. S. Stoffer, R. H. Shumway, and D. S. Stoffer, “Arima models,” *Time series analysis and its applications: with R examples*, pp. 75–163, 2017.

[11] P. Dabral and M. Z. Murry, “Modelling and forecasting of rainfall time series using sarima,” *Environmental Processes*, vol. 4, no. 2, pp. 399–419, 2017.

[12] K. Smagulova and A. P. James, “A survey on lstm memristive neural network architectures and applications,” *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.