B.Tech 2020-24 CSE- Project Phase 1

Proposal

I. Group No.: 42

Project Title.: Anomaly Prediction in Sea Surface Temperature using Time Series Models

Team members :

Rahul Yalavarthi AM.EN.U4AIE20159

Gayathri P AM.EN.U4AIE20126

Jayanth Santosh Varma AM.EN.U4AIE20134

II. 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 proposed model utilizes a Long Short-Term Memory (LSTM) neural network to predict anomalies in SST values. The model aims to capture both spatial and temporal variations in SST values to accurately forecast an SSTA map for a specific region of interest. The persisting challenges in this problem are those of the temperature points being very close which makes even a slight deviation a significant hindrance, also there are many other factors to take in whilst not only taking the past data but also other factors like icebergs in the region

III. Background Study

Title & Year	Problem	Contributions	Limitations	Open problems/ Future work
TITLE: Multilayer Fusion Recurrent Neural Network	The article addresses the need for accurate and holistic	This study introduces a novel MLFrnn model that fully	1) The model's performance may be impacted by data	1) Explore data augmentation techniques to

for Sea Surface Height Anomaly Field Prediction YEAR: 2022 JOURNAL: IEEE transactions on geoscience and remote sensing	prediction of sea surface height anomaly (SSHA) in the South China Sea using a multilayer fusion recurrent neural network (MLFrnn). The proposed model aims to capture long-term dependencies and spatial correlations among neighboring and remote regions, achieving high accuracy and stability for 21-day ahead SSHA predictions.	captures contextual and time-sequential information from neighboring and remote regions to improve sea surface height anomaly (SSHA) field prediction. The model integrates both spatial and temporal features, forecasting the entire SSHA map for the region in a single model. A multilayer fusion cell is designed to effectively combine global and local spatiotemporal features.	availability and quality. 2) The MLFrnn's computational complexity could hinder real-time applications. 3) The study focused on the South China Sea; generalization to other regions may require further investigation.	enhance model robustness. 2) Investigate model optimization for faster inference. 3) Extend the MLFrnn to analyze SSHA fields in different ocean regions for broader applicability.
TITLE : Anomaly Prediction With Hybrid Supervised/Unsupervised Deep Learning for Elastic Optical Networks: A Multi-Index Correlative Approach YEAR : 2022 JOURNAL: Journal of	The paper addresses the challenge of predicting network anomalies in complex optical network environments. It proposes a multi-index anomaly prediction scheme that combines supervised and unsupervised	The proposed anomaly detection scheme for elastic optical networks (EON) with multiple indicators utilizes correlative prediction and LSTM neural networks to forecast future time series. It classifies anomalies using	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) Explore methods to handle imbalanced anomaly data for better classification accuracy. 2) Investigate ensemble techniques to combine supervised and unsupervised models for improved

lightwave technology	deep learning techniques for elastic optical networks. The scheme aims to enhance anomaly prediction by selecting influential indicators, analyzing time series data, and establishing a deep neural network for anomaly classification. The feasibility of the proposed scheme is demonstrated using a real network dataset, showcasing its ability to predict future network anomalies accurately and improve network stability and robustness.	supervised DNN for labeled data and unsupervised DBSCAN for unlabeled data. Experimental results on a real network maintenance dataset demonstrate its ability to identify strongly correlated indicators, predict anomalies in advance, and enhance EON's QoS and stability.		anomaly detection.
TITLE : Robust Anomaly Detection for Multivariate Data of Spacecraft Through Recurrent Neural Networks and Extreme Value Theory	Spacecraft anomaly detection is crucial for avoiding catastrophic failures. Current systems require costly human expertise, facing challenges like large test data,	The paper proposes an unsupervised deep learning-based anomaly detection method for spacecraft telemetry data, combining stacked GRU-based	Limited labeled data for precise anomaly identification, complex model architecture, sensitivity to hyperparameters, variable effectiveness across diverse spacecraft, EVT	1)The proposed method can be extended to explore transfer learning techniques for improved efficiency and robustness across multiple spacecraft and machines.

YEAR: 2021 CONFERENCE: IEEE ACCESS	imbalanced distributions, and scarcity of labeled samples. This work proposes an unsupervised anomaly detection algorithm combining GRU-based RNN and EVT. It introduces a novel ensemble learning framework and automatic anomaly threshold setting through EVT. The proposed method outperforms state-of-the-art approaches in model performance and robustness.	networks and EVT. The approach overcomes data challenges and sets dynamic thresholds for prediction errors. It outperforms state-of-the-art methods in large datasets, marking the first exploration of GRU-based neural networks and EVT in spacecraft anomaly detection. Future work will focus on improving efficiency and robustness through transfer learning across related spacecraft and machines.	assumptions, scalability challenges, suitability for real-time applicability, and lack of interpretability in critical situations.	2) Additionally, investigating interpretability methods to enhance trust in critical decision-making would be beneficial.
TITLE: Mid-Term Simultaneous Spatiotemporal Prediction of Sea Surface Height Anomaly and Sea Surface Temperature Using Satellite Data in the South China Sea	The primary objective of this paper is to tackle the difficulty of forecasting daily sea surface height anomaly (SSHA) and sea surface temperature (SST) in the marine environment for	The paper proposes a novel data-driven method that combines empirical orthogonal function of multivariate, complete ensemble empirical mode decomposition, and a multilayer	1) The proposed model's performance might be specific to the South China Sea region where it was tested. Generalizing the model to other marine environments may require additional	1)Developing a data-driven-based ocean-atmosphere coupled model could be explored to capture the relationship between the ocean and the atmosphere, leading to more

YEAR : 2020 JOURNAL : IEEE Geoscience and Remote Sensing Letters	a mid-term period.	perceptron (MEOF-CEEM D-MLP). The model takes into account the correlation between SSHA and SST, and the temporal and spatial relationship between discrete points allowing more accurate predictions.	validation and tuning. 2)The model does not consider wind conditions in its predictions. 3) 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.	accurate predictions. 2)Extending the model's forecasting capabilities beyond the 30-day mid-term period to address long-term predictions. Future work could also focus on incorporating external forces, such as wind conditions.
TITLE: Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies YEAR: 2021 JOURNAL: Geophysical Research Letters	To evaluate the predictability of seasonal sea surface temperature (SST) and sea surface height (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	Developed a global Linear Inverse Model (LIM) that can predict monthly mean sea surface temperature (SST) and sea surface height (SSH) anomalies over the ice-free global ocean. This model outperforms operational numerical models for SSH skill in the Atlantic and US East Coast tide	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, especially for modes of variability with	1)Future work could explore the LIM's skill in predicting other essential oceanic and atmospheric variables, such as currents, precipitation, and atmospheric circulation patterns. 2)Extending the hindcast period beyond 2010 could provide a more comprehensive assessment of the LIM's skill, especially for

	period 1982–2010.	gauge stations.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.	decadal or longer time scales. 3)The NMME GCMs used for sea surface height (SSH) hindcasts do not assimilate satellite altimetry information in their initializations, which could contribute to relatively poorer SSH prediction skill.	longer-term climate modes and variability. 3)explore the impact of data assimilation techniques on improving the skill of NMME models.

IV. Challenges

1. Dataset size: the dataset selected for the project is only across 2 years and has approximately 700 points deemed too short for analysis, this problem is solved by considering various window sizes and training the data points.
2. Dataset format: the given file is in the format of .nc which is just a masked array collection of various parameters, this unlike usual csv files is harder to train and preprocess.
3. Range of values in dataset: the values in the dataset are very close to each other varying only in half a degree usually and 4-5 degrees at most, so although the predictions show slight deviation its an indication of higher error.
4. Availability of frame works: the frameworks used in some of the base papers are not even published for free public use, example multilayer fusion Rnn. also the existing ConvLSTM doesn't support .nc files as an input.

V. Deliverables of Phase I

- Phase I
- Benchmark existing sequential model on the sea surface temperature dataset, and identify the best one
- To build the best network to predict sea surface temperatures at a particular point and assess its generalization capability on similar dataset.

- Phase II
- With the knowledge gained, design better model/architecture
- To use the further built model to predict temperatures at other points in the dataset.
- To build more advanced models for other parameters and new datasets including sea surface heights, sea wave heights and other marine parameters.(future scope)

Outcomes/Deliverables

- A model which can accurately predict temperatures with a maximum error of only 2 degrees.
- Another model which can predict with similar performance as the above model but this time over a region rather than a single point. Using advanced LSTM systems.
- More models which can even predict parameters like sea surface heights, sea wave heights and other marine parameters.(future scope)

VI. Assumptions/Declarations:

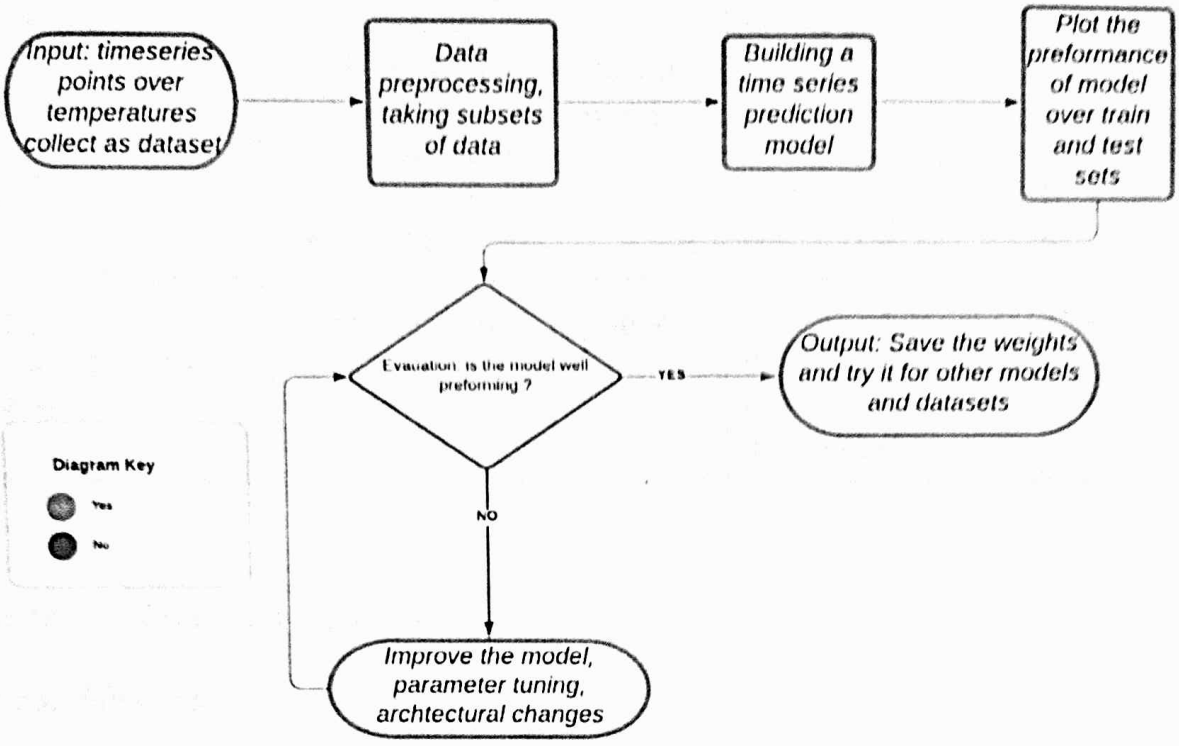
The assumptions for the dataset we are currently working are that:

- The relation to sea surface height is only dependent on previous temperatures and the analysed ssd_ice factor in the dataset.
- For now the LSTM model built on one point to predict temperatures is used is assumed to work on other points of the dataset also.

VII. Tools to be used

Software/Hardware Tools	Specifications
Python	Programming Language and inbuilt libraries to implement the models and test their working.
Kaggle	A platform to perform all the programming operations and measure the results.
Copernicus Marine Data Store	Dataset Accumulation
Copernicus MOTU tool	Accessing subsets of dataset

VIII. High Level Design



IX. Timeline

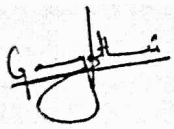
Week	Task Planned	Outcome	Challenge
Week 1-2: August 2nd - 3rd Week	Accumulating datasets and pre built models	A collection of datasets and pre built models	Time series models are underexplored hence lesser models compared to others.
Week 3-4: August 4 - September 1st week	Experimenting with • LSTM • Bi-directional LSTM • SciNet • ARIMA/SARIMA • Vector Autoregressive Model (VAR) • Vector Error Correction Model (VECM) • Empirical Mode Decomposition (EMD) • Prophet model by Facebook	A benchmark model for predicting sea surface temperatures	Might have anomalies so predictions are challenging
Week 5-6:	Experimenting various other	A benchmark	Might have

September 2nd - 3rd week	datasets like: • Sea Surface Wave Height • Sea Surface Height • Sea Surface Wind and Stress across benchmark model obtained earlier	model capable to predict multiple sea factors	anomalies so predictions are challenging
Week 7: September 4th week	Documentation and paper draft	A report detailing the work done and a latex document with a draft	None

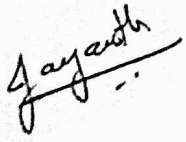
Students' Name and Signature:

Rahul Yalavarthi

Gayathri P



Jayanth Santosh Varma



Guide's Signature :