

Predicting Future Sea Level Variations Using Global Mean Sea Level Variations

Youssef Sharaf, Dr. Ghada Khoriba

Fall 2024 Semester

Table of Contents

Contents

[Abstract 3](#_Toc187187549)

[Introduction 3](#_Toc187187550)

[Related works 4](#_Toc187187551)

[Methods 7](#_Toc187187552)

[Results 9](#_Toc187187553)

[Discussion 13](#_Toc187187554)

[Conclusion 14](#_Toc187187555)

[References 15](#_Toc187187556)

[Appendix 15](#_Toc187187557)

# Abstract

This project aims to predict future global mean sea level (GMSL) variations by leveraging historical satellite altimetry data. As climate change accelerates, accurate forecasting of sea level trends is essential for coastal planning, environmental protection, and disaster preparedness. Using a time series forecasting approach, the study analyzes GMSL data from the TOPEX/Jason mission, incorporating both raw and smoothed sea level measurements, with and without Global Isostatic Adjustment (GIA). Several machine learning models, including Random Forest, Support Vector Regression (SVR), Linear Regression, Multi-Layer Perceptron (MLP), and Recurrent Neural Networks (RNN), were implemented and evaluated. These models were assessed for their ability to capture both short-term fluctuations and long-term trends in sea level. The results highlight that neural network-based approaches, particularly RNNs, excel at modeling temporal dependencies, while ensemble methods like Random Forest provide robust predictions. By enhancing sea level forecasting capabilities, this project contributes valuable insights for mitigating the impacts of rising sea levels on vulnerable communities and ecosystems.

# Introduction

Global mean sea level (GMSL) is a critical indicator of climate change, reflecting the cumulative impact of factors such as thermal expansion, melting ice sheets, and glaciers. The rise in GMSL poses significant risks to coastal communities, ecosystems, and infrastructure, necessitating accurate forecasting models to inform policy and adaptation strategies. Understanding and predicting sea level variations are essential for managing risks associated with flooding, storm surges, and coastal erosion.

Satellite altimetry missions, such as the TOPEX/Jason series, have provided valuable datasets for monitoring GMSL variations over the past few decades. These datasets include raw and smoothed measurements of sea level changes, with and without adjustments for Global Isostatic Adjustment (GIA), offering a comprehensive view of both short-term fluctuations and long-term trends. However, the complexity of the data and the intricate interactions between contributing factors make predicting future sea level variations a challenging task.

This study aims to address this challenge by employing machine learning models to forecast GMSL variations based on historical satellite altimetry data. By leveraging the predictive capabilities of models such as Random Forest, Support Vector Regression (SVR), Linear Regression, Multi-Layer Perceptron (MLP), and Recurrent Neural Networks (RNN), this research evaluates the performance of different methodologies in capturing temporal dependencies and trends in the data.

The findings of this study not only provide insights into the effectiveness of various machine learning approaches but also contribute to the broader goal of improving sea level prediction. This is particularly crucial for climate adaptation planning and the mitigation of risks to vulnerable communities and ecosystems. By advancing the state of predictive modeling for GMSL, this research underscores the importance of integrating data-driven techniques into environmental monitoring and decision-making processes.

# Related work

The prediction of sea level variations has been a focus of extensive research due to its importance in understanding and mitigating the impacts of climate change. Researchers have explored diverse machine learning (ML) models and methodologies to address this complex problem, leveraging different datasets, features, and prediction horizons. This section reviews relevant studies that highlight the progression and contributions in this field.

(Nur Alyaa Hazrin et al., 2023) conducted a comprehensive analysis of machine learning algorithms to predict sea level changes on a daily basis. Using data spanning from 1985 to 2018, the study evaluated various regression and neural network models across multiple locations in Malaysia, including Peninsular Malaysia and Sabah. The results indicated that model performance varied by location, with linear regression, Gaussian process regression, and trilayered artificial neural networks emerging as optimal choices for specific areas. Their findings also emphasized the critical role of temporal lag in enhancing model accuracy, particularly with a 7-day lag. The study underscored the utility of ML in modeling short-term regional coastal sea level changes, offering decision-makers valuable tools for climate change mitigation. Furthermore, the integration of physical variables, such as temperature, into ML models demonstrated the potential to bridge the gap between localized predictions and global climate adaptation strategies.

(Nieves et al., 2021) investigated the application of machine learning to predict sea level variations, focusing on the influence of ocean temperature as a proxy for the thermosteric component of sea level change. The study highlighted the regional complexities of sea level variations, driven by processes occurring across multiple timescales, with temperature-induced changes playing a dominant role on a multi-year scale. By incorporating key temperature estimates, the proposed machine learning models successfully captured coastal sea level variability and its associated uncertainties over timescales ranging from months to several years.

Their models demonstrated high accuracy, particularly in coastal regions heavily influenced by internal climate variability. Furthermore, the methodology proved effective for both regional assessments and broader global evaluations of rising and falling sea level patterns. Importantly, the study emphasized the practical implications of these predictions, offering valuable insights for short-term planning and coastal protection measures. This approach bridges the gap between observational data and actionable forecasting, providing a robust framework for near-term decision-making.

Sarmad et al. (2024) explored the performance of Support Vector Machine (SVM) and k-Nearest Neighbors (kNN) models for predicting sea level variations. The study involved training thirteen distinct models, thoroughly assessing their performance during both training and testing phases. It was observed that while SVM models demonstrated good performance during training, their testing performance was relatively poorer, suggesting potential overfitting issues. In contrast, kNN models exhibited consistent and reliable performance across both training and testing datasets.

A particular focus was placed on evaluating the effectiveness of different SVM kernels, with the Radial Basis Function (RBF) kernel identified as the most suitable for analyzing sea level rise data. The RBF kernel delivered acceptable values for common metrics, including RMSE, MAE, and R², making it a preferred choice for modeling such datasets. This study underscores the importance of carefully selecting algorithms and kernel functions to achieve robust and accurate predictions in sea level modeling tasks.

Tur et al. (2021) emphasized the importance of accurate sea level predictions for coastal infrastructure design and harbor operations, presenting a methodology that utilizes sea level height and meteorological observations. The study focused on two scenarios: SC1, which used lagged sea level observations as inputs, and SC2, which incorporated both lagged sea level and meteorological factors. A cross-correlation analysis was performed to optimize the input combinations for each scenario.

Predictive models were developed using multiple linear regression (MLR) and adaptive neuro-fuzzy inference system (ANFIS) techniques. The evaluation metrics included RMSE, MAE, scatter index (SI), and Nash Sutcliffe Efficiency (NSE). The findings demonstrated that incorporating meteorological factors into the input parameters enhanced MLR model accuracy by up to 33% for short-term predictions. Furthermore, ANFIS consistently outperformed MLR across both SC1- and SC2-based input combinations, highlighting its superior capability for capturing complex nonlinear relationships in sea level prediction tasks.

Ayinde et al. (2024) provide a comprehensive review of methodologies for forecasting sea level change (SLC) using machine learning (ML) models, emphasizing their potential for improving coastal management and resilience in the face of climate change. The study highlights the advantages of artificial neural networks (ANNs), particularly deep learning models and their hybrid variants, over traditional regression methods and simpler ML approaches for short-term sea level anomaly prediction. According to the authors, supervised learning approaches dominate the field, with semi-supervised methods excelling in short-term projections.

The paper further examines the performance of simpler models, such as regression and support vector machines (SVM), which are noted to perform adequately with sufficient training data but often struggle with the complexity of non-linear scenarios. The importance of selecting appropriate input variables, such as atmospheric, oceanic, and geological factors, is emphasized, as this significantly impacts the model's accuracy. Additionally, the authors stress the need to maintain a balance between training and testing data to avoid the risks of overfitting and underfitting.

The studies reviewed in this section underscore the growing importance of machine learning techniques in addressing sea level change. While advancements have been made in short-term anomaly prediction, challenges remain in improving long-term forecasting accuracy.

# Methods

This study employs a machine learning-based approach to forecast global mean sea level (GMSL) variations using historical satellite altimetry data. The methodology is structured into several steps, including data preprocessing, model implementation, training, and evaluation.

**1. Data Collection and Preprocessing**

The dataset used in this study is derived from the TOPEX/Jason satellite altimetry missions, containing measurements of GMSL variations. The data includes 13 columns, such as altimeter type, cycle number, year, number of observations, and various GMSL measurements (both raw and smoothed, with and without Global Isostatic Adjustment).

Preprocessing steps included:

* **Feature Engineering**: Selecting relevant input features, including altimeter type, cycle number, year fraction, and various GMSL columns.
* **Scaling**: Standardizing the input features using StandardScaler to normalize the data for machine learning models.
* **Encoding**: Label encoding was applied to categorical features like altimeter type.

**2. Model Implementation**

Six machine learning models were implemented to predict GMSL variations:

* **Random Forest (RF)**: An ensemble learning model used to capture nonlinear relationships in the data.
* **Support Vector Regression (SVR)**: A kernel-based model, leveraging the radial basis function (RBF) kernel to model complex patterns.
* **Linear Regression (LR)**: A baseline model to establish the performance of simple linear relationships.
* **Multi-Layer Perceptron (MLP)**: A feedforward neural network with one hidden layer, designed to model nonlinear relationships.
* **Recurrent Neural Network (RNN)**: A deep learning model leveraging temporal dependencies, with the SimpleRNN architecture trained on time-series reshaped input.
* **Decision Tree**: A non-parametric model that splits the data into subsets based on feature values, recursively partitioning the data into homogeneous groups.

**3. Training and Testing**

The dataset was split into training (80%) and testing (20%) subsets to evaluate model performance. For models requiring sequential data (e.g., RNN), the input was reshaped into a 3D format with time steps and features.

**Hyperparameter Tuning:** The RNN used a hidden layer size of 50 and the Adam optimizer, while the Random Forest used 100 estimators.

**4. Evaluation Metrics**

Model performance was evaluated using:

* **Mean Squared Error (MSE)**: To quantify the average squared difference between predicted and actual values.
* **R² Score (Coefficient of Determination)**: To measure the proportion of variance in the target variable that is explained by the model.
* **Visualization**: Scatter plots of predicted vs. actual values to assess model accuracy visually.

**5. Implementation Details**

The models were implemented in Python using libraries such as scikit-learn, TensorFlow, and pandas. The SVR and MLP models required standardized inputs, while RNN training involved reshaped data to capture temporal patterns.

**6. Visualization and Analysis**

Results were visualized using scatter plots and bar charts, providing a comparative analysis of actual versus predicted GMSL variations. This aided in interpreting model predictions and identifying trends in the data.

By implementing and comparing these models, the study identifies the most effective approaches for forecasting GMSL variations, contributing to better predictions and climate adaptation strategies.

# Results

Below, figures (1, 2, 3, 4, 5, 6) show a scatter plot for each of the random forest, decision tree, SVR, RNN, linear regression, and MLP models, respectively. Figures (7, 8) show the mean squared error and r^2 accuracy score for each model.

A graph with blue dots

Description automatically generated

Figure 1

A graph of a tree

Description automatically generated with medium confidence

Figure 2

A graph with blue dots

Description automatically generated

Figure 3

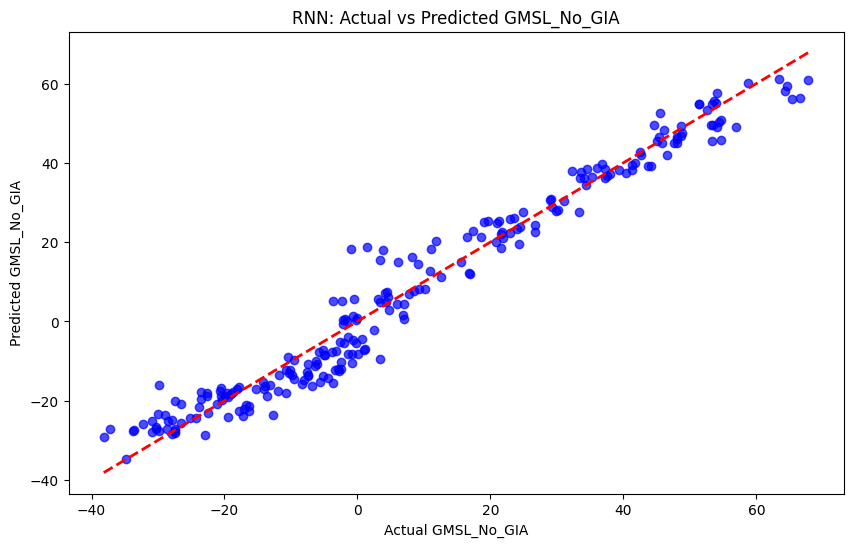


Figure 4

A graph of a line graph

Description automatically generated with medium confidence

Figure 5

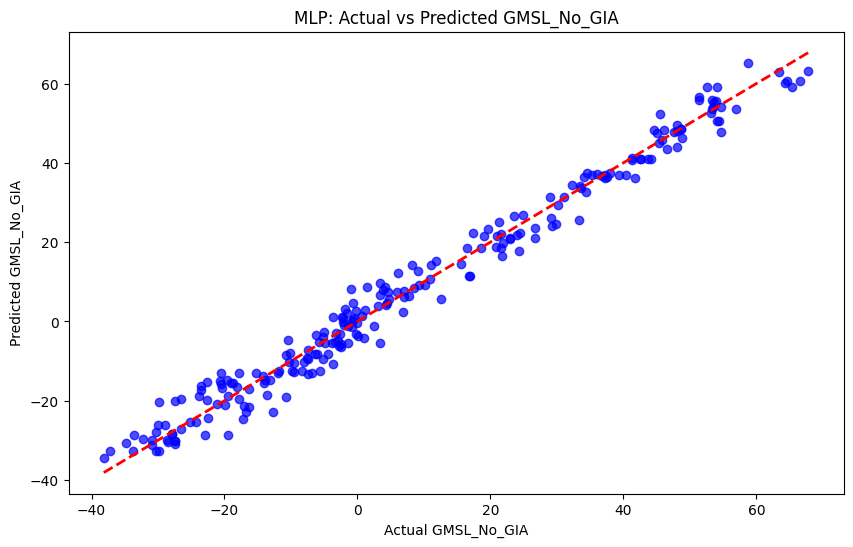


Figure 6

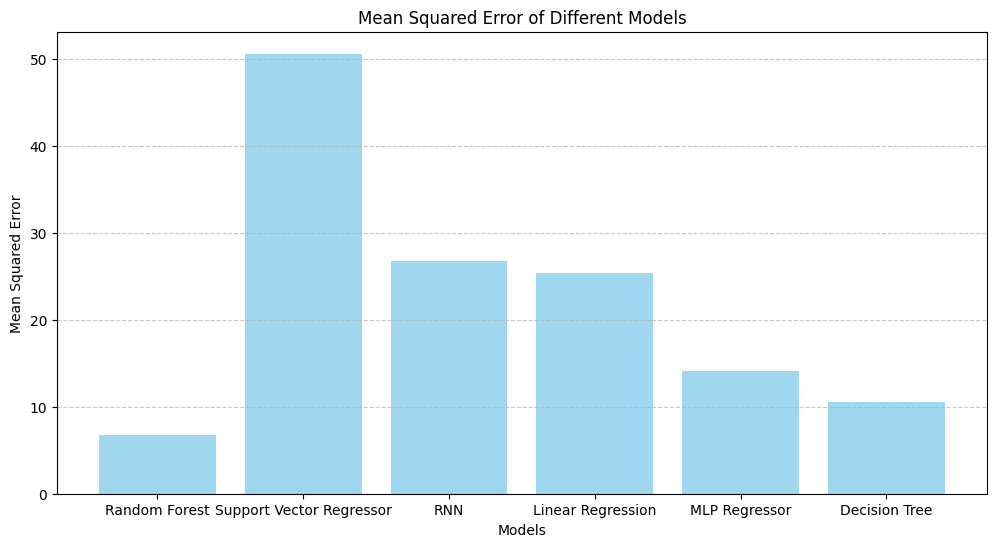


Figure 7

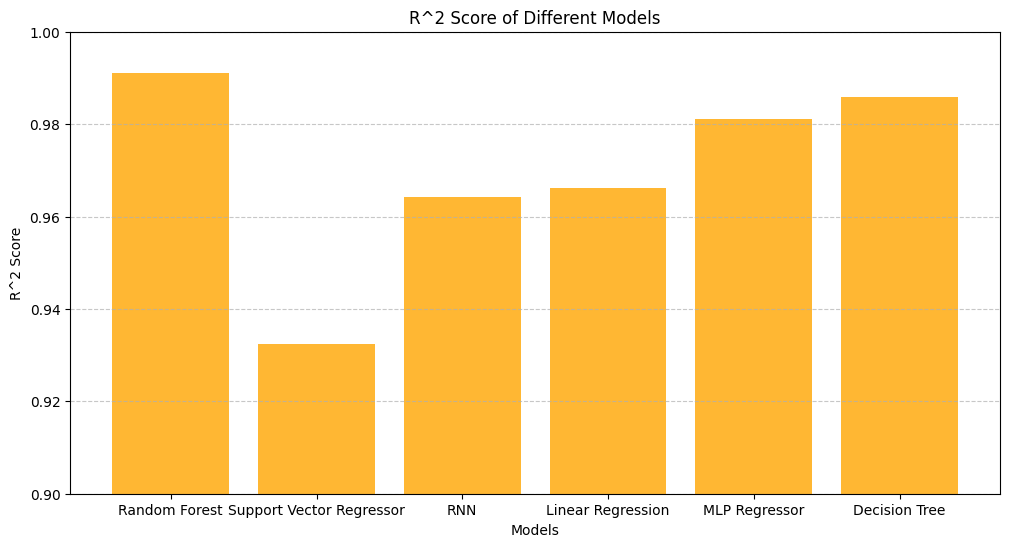


Figure 8

# Discussion

**Random Forest**

The Random Forest model demonstrated the best performance among all models, achieving the lowest MSE (6.73) and the highest R² score (0.99). This indicates its ability to handle complex relationships and interactions in the dataset effectively. Its ensemble nature makes it robust to overfitting, contributing to its superior accuracy.

**Support Vector Regressor (SVR)**

The SVR model had the highest MSE (50.63) and the lowest R² score (0.93). Despite being a strong candidate for smaller datasets or simpler relationships, the complexity of the GMSL dataset and potential sensitivity to hyperparameters may have limited its performance. This suggests that SVR might require more careful tuning or may not be as suitable for this problem.

**Recurrent Neural Network (RNN)**

The RNN achieved an MSE of 26.77 and an R² score of 0.96, indicating its ability to capture temporal dependencies in the data. While its performance was commendable, it lagged behind simpler models like Random Forest and Decision Tree, potentially due to challenges in hyperparameter optimization or the relatively small dataset size.

**Linear Regression**

Linear Regression performed moderately well, with an MSE of 25.35 and an R² score of 0.97. This suggests that while linear models can approximate the GMSL trends to some extent, they are limited in capturing nonlinear relationships inherent in the data.

**Multi-Layer Perceptron (MLP) Regressor**

The MLP Regressor exhibited strong performance, with an MSE of 14.08 and an R² score of 0.98. Its ability to model nonlinear relationships and flexibility in learning patterns highlights its suitability for the dataset. However, its performance fell short of Random Forest and Decision Tree, possibly due to overfitting or inadequate optimization.

**Decision Tree**

The Decision Tree model performed exceptionally well, with an MSE of 10.53 and an R² score of 0.98, closely rivaling Random Forest. Its interpretability and ability to handle nonlinearity make it a compelling choice for GMSL prediction. However, its susceptibility to overfitting compared to ensemble methods like Random Forest must be considered.

# Conclusion

The results highlight the effectiveness of ensemble methods (Random Forest, Decision Tree) and neural networks (MLP, RNN) in predicting GMSL variations. Ensemble methods consistently outperformed standalone models like Linear Regression and SVR, underscoring their robustness in capturing complex patterns. Neural networks like RNN and MLP showed promise but require larger datasets and optimized architectures to fully leverage their potential.

Future work should focus on refining neural network models, exploring hybrid approaches, and leveraging larger datasets for improved accuracy. The consistent performance of Random Forest and Decision Tree suggests their reliability for immediate application in sea level prediction tasks.

# References

GSFC. 2021. Global Mean Sea Level Trend from Integrated Multi-Mission Ocean Altimeters TOPEX/Poseidon, Jason-1, OSTM/Jason-2, and Jason-3 Version 5.1. Ver. 5.1 PO.DAAC, CA, USA. Dataset accessed 2025-1-7 at <https://doi.org/10.5067/GMSLM-TJ151>.

Nur Alyaa Hazrin, Chong, K. L., Huang, Y. F., Ahmed, A. N., Ng, J. L., Koo, C. H., Tan, K. W., Sherif, M., & El-shafie, A. (2023). Predicting sea levels using ML algorithms in selected locations along coastal Malaysia. *Heliyon*, *9*(9), e19426–e19426. <https://doi.org/10.1016/j.heliyon.2023.e19426>

Nieves, V., Radin, C. & Camps-Valls, G. Predicting regional coastal sea level changes with machine learning. *Sci Rep* **11**, 7650 (2021). <https://doi.org/10.1038/s41598-021-87460-z>

Sarmad Dashti Latif, Mohammad Abdullah Almubaidin, Chua Guang Shen, Sapitang, M., Birima, A. H., Ali Najah Ahmed, Sherif, M., & El-Shafie, A. (2024). Improving sea level prediction in coastal areas using machine learning techniques. *Ain Shams Engineering Journal*, 102916–102916. https://doi.org/10.1016/j.asej.2024.102916

‌ Tur, R., Tas, E., Haghighi, A. T., & Mehr, A. D. (2021). Sea Level Prediction Using Machine Learning. *Water*, *13*(24), 3566. https://doi.org/10.3390/w13243566

‌ Ayinde Akeem Shola, Huaming, Y. U., & Kejian, W. U. (2024). Review of machine learning methods for sea level change modeling and prediction. *The Science of the Total Environment*, *954*, 176410–176410. https://doi.org/10.1016/j.scitotenv.2024.176410

# Appendix

https://github.com/Elsha2023/Machine-intelligence