water

MDPI

Article

Hybrid Reconstruction of Sea Level at Dokdo in the East Sea Using Machine Learning and Geospatial Interpolation (1993–2023)

MyeongHee Han and Hak Soo Lim *

Marine Natural Disaster Research Department, Korea Institute of Ocean Science and Technology, Busan 49111, Republic of Korea; mhan@kiost.ac.kr
* Correspondence: hslim@kiost.ac.kr

Abstract

Sea level variability in the East Sea (Sea of Japan) and the Northwest Pacific poses challenges for coastal risk management due to the scarcity of long-term observations at remote locations such as Dokdo (Dok Island). This study reconstructs a continuous monthly sea level record at Dokdo from 1993 to 2023 by imputing gaps in 13 nearby Permanent Service for Mean Sea Level tide gauge stations using eight machine learning models and geospatial interpolation methods. The ensemble mean of Machine Learning-based imputations produced physically realistic and temporally coherent timeseries, preserving both seasonal and interannual variability. Sea level at Dokdo, estimated via inverse distance weighting, aligned well with satellite altimetry from Copernicus Marine Service and exhibited strong regional coherence with nearby stations. These results demonstrate that a hybrid framework combining statistical imputation, Machine Learning, and inverse barometric correction can effectively reconstruct sea level in data-sparse marine regions. The methodology provides a scalable tool for monitoring long-term trends and validating satellite and model products in marginal seas like the East Sea.

Keywords: sea level reconstruction; Dokdo (Dok Island); east sea; permanent service for mean sea level; satellite altimetry; ML-based imputation

check for updates

Academic Editors: Chunpeng Chen, Ce Zhang and Wenting Wu

Received: 15 July 2025
Revised: 19 September 2025
Accepted: 15 October 2025
Published: 16 October 2025

Citation: Han, M.; Lim, H.S. Hybrid Reconstruction of Sea Level at Dokdo in the East Sea Using Machine Learning and Geospatial Interpolation (1993–2023). *Water* **2025**, *17*, 2989. https://doi.org/10.3390/w17202989

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Sea level change is a critical indicator of climate variability and oceanographic dynamics, particularly in marginal seas and coastal regions where anthropogenic and natural influences converge [1–3]. In the Northwest Pacific, the East Sea (Sea of Japan) and its surrounding areas, including Dokdo (Dok Island), Ulleungdo, and the southeastern coast of the Korean Peninsula, exhibit complex sea level variability influenced by monsoon forcing, typhoons, and large-scale circulation systems such as the Kuroshio Extension and the Tsushima Warm Current [4,5]. Monitoring and understanding sea level variability in this region is essential for coastal hazard assessment, maritime safety, and climate change adaptation strategies [6,7].

However, continuous long-term sea level observations in the open ocean near Dokdo are lacking due to the absence of long-term in situ tide gauges [8,9]. In contrast, the Permanent Service for Mean Sea Level (PSMSL) provides a network of tide gauge stations around the East Sea and the broader region, offering a valuable opportunity to estimate sea level at data-sparse locations through statistical interpolation and machine

learning (ML) techniques [10,11]. Recent advances in ML enable the reconstruction of missing or unavailable observations by leveraging spatial and temporal patterns from nearby stations [12–14].

In this study, we aim to reconstruct and analyze monthly sea level timeseries at Dokdo from 1993 to 2023 using a combination of geospatial interpolation, statistical models, and ML algorithms [15,16]. We begin by collecting and preprocessing monthly mean sea level data from the 13 selected PSMSL stations closest to Dokdo, applying outlier filtering and distance-based reordering. Two baseline methods, simple ensemble averaging and inverse distance weighting (IDW), are employed to generate initial predictions at Dokdo [17,18]. These are compared to results from a suite of ML imputation models applied to the 13 stations, including tree-based ensembles (Random Forest, XGBoost, LightGBM), regularized regression (Ridge, Bayesian Ridge), proximity-based methods (KNN), and neural networks [19–25].

Large-scale climate modes modulate sea level in the East Sea (Sea of Japan), introducing variability on interannual to multi-decadal timescales. The El Niño–Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and Indian Ocean Dipole (IOD) influence regional sea level through steric height changes, wind forcing, and basin-scale circulation, and can temporarily accelerate or decelerate the background rise [26–36]. Our 1993–2023 reconstruction overlaps major events such as the 1997–1998 El Niño and early-2000s PDO phase shifts, but the limited time span precludes a full resolution of longer climate cycles [37–39]. In this study, we focus on developing and validating a reproducible reconstruction framework for Dokdo during the satellite-altimetry era, while acknowledging that these climate modes provide essential context for interpreting regional sea-level change [30,32].

Our approach allows for the assessment of interannual to decadal sea level variability at Dokdo in the absence of direct long-term tide gauge measurements [40]. By integrating geodesic distance, regional coherence, and advanced imputation techniques, we provide a robust and reproducible methodology for reconstructing historical sea level at isolated marine locations [41]. The reconstructed records offer insights into regional oceanographic trends and provide a foundation for future studies on sea level rise and variability in the East Sea and surrounding maritime zones [42,43].

2. Data and Methods

This study investigates monthly sea level variability at Dokdo from 1993 to 2023 by utilizing observations from 13 selected nearby PSMSL tide gauge stations distributed across the East Sea and adjacent Northwest Pacific coastal regions (Figure 1). These stations, located within approximately 1000 km of Dokdo, form a spatially coherent observational network that captures regional oceanographic variability. Original PSMSL records were compiled for each station (Figure 2) and processed through a $\pm 3\sigma$ outlier filtering procedure to remove anomalous values. Tide-gauge records were corrected for the inverse barometer (IB) effect, which represents the dominant component of the Dynamic Atmospheric Correction (DAC) and ensures consistency with satellite altimetry products. Vertical land motion (VLM) corrections were not applied in this study, as Global Navigation Satellite System (GNSS)-constrained rates are not uniformly available for all stations in the East Sea region. Instead, we focus on reconstructing relative sea level, and we discuss the implications of VLM for tide-gauge/altimetry consistency in Section 3.4. Tide-gauge observations were obtained from the PSMSL [44], while satellite altimetry products were taken from the Copernicus Marine Service (CMS; [45]).

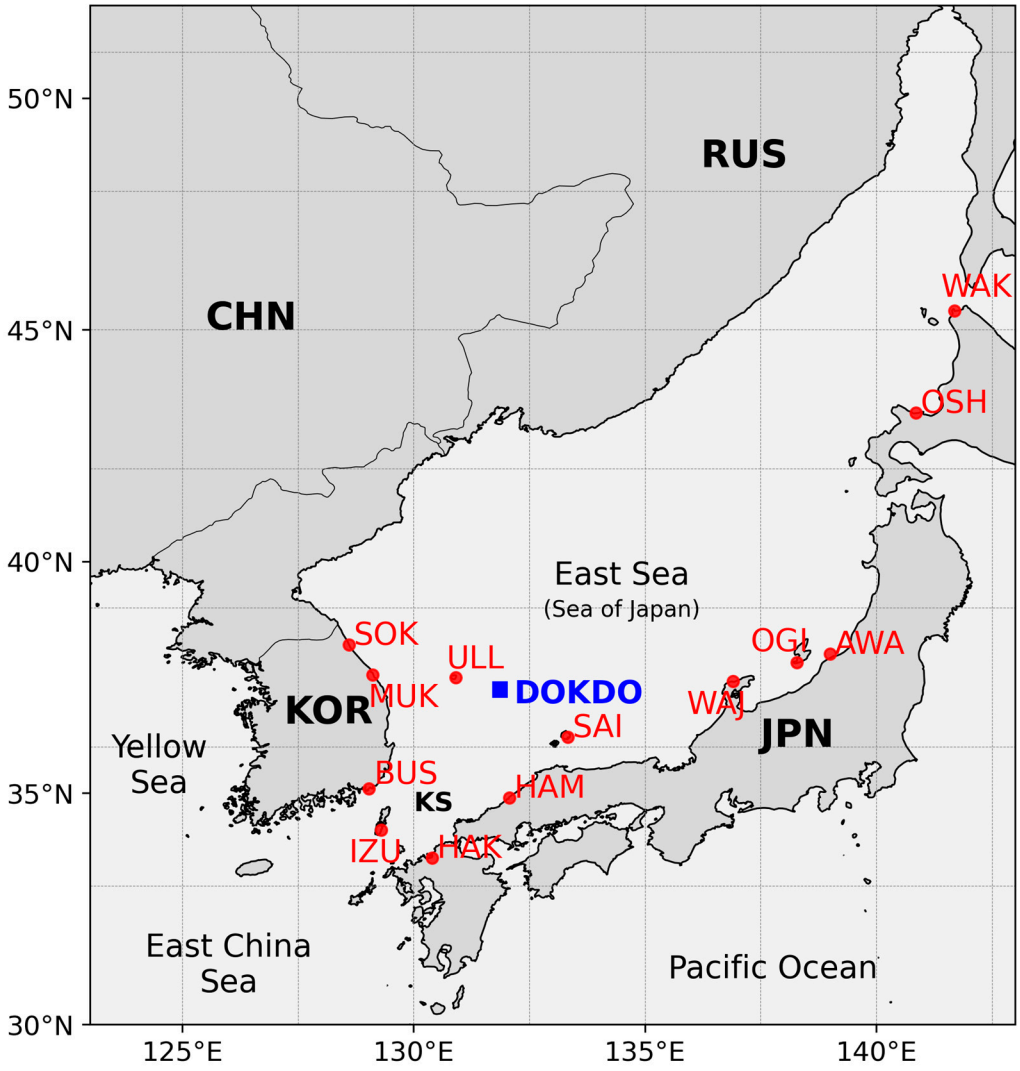


Figure 1. Map showing the geographical locations of the Dokdo (blue square) and surrounding selected PSMSL tide gauge stations (red circles) in the East Sea and adjacent regions. Station abbreviations correspond to the locations used in this study, covering coastal areas of Korea (KOR), Japan (JPN), and Russia (RUS). The map also denotes major surrounding seas, including the Yellow Sea, East China Sea, and Pacific Ocean. CHN and KS represent China and the Korea Strait. This regional configuration provides the basis for evaluating sea level variability and reconstructing sea level at Dokdo using neighboring tide gauge records.

All records were aligned to a common monthly time axis to enable consistent interstation comparison and reconstruction of missing data. To evaluate the completeness of the tide gauge records, we quantified the number and distribution of missing months for each station over 1993–2023 (Table 1). Most stations were highly complete, with <5% missing data (e.g., Saigo, Hamada II, Wakkanai), while a few exhibited longer gaps more than 19 months (e.g., Izuhara II, Hakata, Awa Sima). The presence of both short sporadic gaps and longer discontinuities motivated the use of ensemble-based machine learning imputation methods, which are more robust than single-model or linear interpolation approaches. By explicitly accounting for heterogeneous gap structures across stations, the framework ensures that reconstructions retain physical consistency and minimize biases from uneven data coverage.

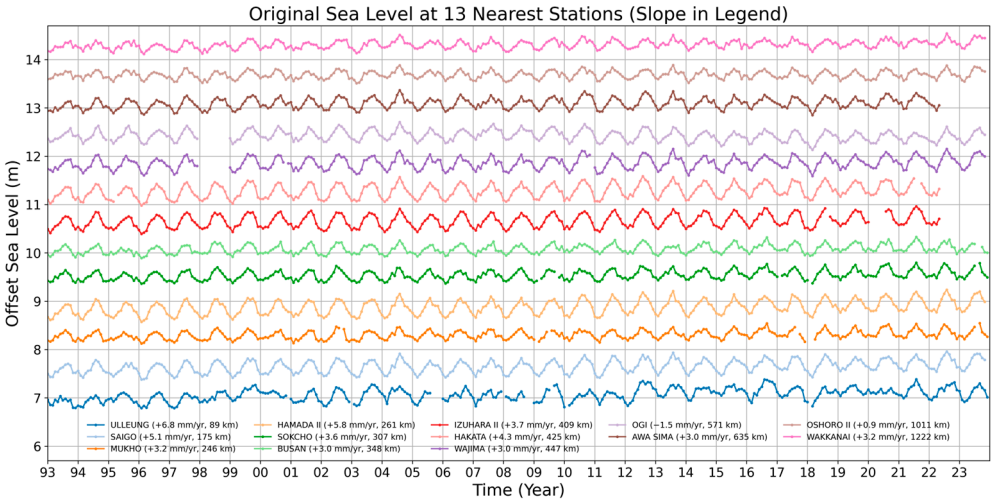


Figure 2. Monthly mean sea level at the 13 selected reference PSMSL tide gauge stations around Dokdo (1993–2023), shown with vertical offsets for clarity. Each station's linear trend (slope in mm/yr) and geodesic distance from Dokdo (in km) are included in the legend. Positive trends indicate long-term sea level rise. The data exhibit seasonal and interannual variability, with noticeable differences in trend magnitude and temporal coverage across stations.

Table 1. Data quality summary for the 13 tide gauge stations (1993–2023). Reported are the total record length (372 months), number of valid and missing months, fraction of missing data, total number of gaps, and maximum gap length (months). These statistics informed the choice of ensemble-based imputation methods.

Station	Total Months	Valid Months	Missing Months	Missing (%)	All Gaps	Max Gap Length
ULLEUNG	372	359	13	3.5	8	4
SAIGO	372	371	1	0.3	1	1
MUKHO	372	364	8	2.2	6	3
HAMADA II	372	371	1	0.3	1	1
SOKCHO	372	367	5	1.3	5	1
BUSAN	372	370	2	0.5	1	2
IZUHARA II	372	346	26	7.0	3	19
HAKATA	372	350	22	5.9	3	19
WAJIMA	372	357	15	4.0	4	12
OGI	372	358	14	3.8	3	12
AWA SIMA	372	353	19	5.1	1	19
OSHORO II	372	371	1	0.3	1	1
WAKKANAI	372	371	1	0.3	1	1

To independently validate the reconstructed sea-level time series at Dokdo, we used satellite altimetry data from the CMS. Specifically, we obtained the monthly gridded sea-level anomaly product [46], which provides global anomalies on a 0.25° latitude–longitude grid from January 1993 onward. The dataset is publicly available through the CMS portal [46]. To extract the altimetric sea-level anomalies near Dokdo, we identified the grid cell geographically closest to the Dokdo coordinates (37.241° N, 131.865° E) by minimizing absolute latitude and longitude differences. Because CMS provides relative anomalies, whereas tide-gauge composites carry absolute levels, we applied a mean-alignment procedure: The temporal mean of the extracted SLA series was removed, and the mean of the 13-station PSMSL ensemble (monthly, gap-filled) was computed. This scalar mean was added to the SLA series, producing a level-aligned product comparable with tide-gauge data. The aligned series represents CMS at Dokdo adjusted to the PSMSL reference level. Because tide-gauge and satellite altimetry observations reflect different

spatial scales—tide gauges capture nearshore processes such as coastal setup, local winds, and waves, whereas altimetry averages offshore conditions over larger areas—systematic offsets are expected. To reconcile these differences, we applied a linear ordinary least squares (OLS) regression between the CMS SLA and the tide-gauge reconstruction for their overlapping period. This adjustment removes systematic biases and scales the altimetric anomalies to the tide-gauge reference frame. We also constructed 95% prediction intervals around the adjusted altimetry series to quantify mapping uncertainty.

Validation skill was assessed using multiple metrics, root-mean-square error (RMSE), mean absolute error (MAE), correlation coefficient (r), and coefficient of determination (R^2), together with moving-block bootstrap confidence intervals (12-month blocks, 1000 replicates). Metrics were computed for the full overlapping record (1993–2023) and for targeted extreme-event windows, including the 1997–1998 El Niño and the 2003 typhoon season. This procedure ensured that the Dokdo tide-gauge reconstruction was validated against an independent large-scale dataset while properly accounting for local-to-regional observational differences.

To reconstruct missing values in the monthly tide-gauge records, we applied an Iterative Imputer framework with eight different regression algorithms: K-Nearest Neighbors (KNN; [47]), Random Forest (RF; [48]), Extra Trees (ET; [49]), Bayesian Ridge (BR; [50]), Ridge Regression ([51]), Multilayer Perceptron (MLP; [52]), Extreme Gradient Boosting (XGBoost; [53]), and Light Gradient Boosting Machine (LightGBM; [54]). Each model was trained to iteratively predict missing entries from the available station network. To ensure comparability and avoid overfitting to any single method, all regressors were implemented with standardized, lightly tuned hyperparameters. For example, RF and ET were limited to 100 trees each, XGBoost and LightGBM to 100 boosting rounds, and the MLP to 1000 training iterations. KNN used three nearest neighbors, while Ridge and Bayesian Ridge followed their default regularization schemes. These choices balanced computational cost and predictive stability, while ensuring that no individual algorithm dominated due to excessive tuning. The ensemble mean across the eight imputed datasets was then used as the final filled product, which stabilizes the time series by reducing model-specific biases and noise. Table 2 summarizes the eight regression models embedded in the Iterative Imputer. Their complementary assumptions ensure that the ensemble average mitigates model-specific biases and enhances reconstruction robustness.

Table 2. Summary of regression models used in the Iterative Imputer framework for sea-level gap filling.

Model	Key Parameters (Fixed)	Strengths	Limitations
KNN	Uses the 3 closest neighbors (equal weights) for imputation	Simple, non-parametric; effective for local similarity	Sensitive to data sparsity and choice of number of neighbors
RF	100 decision trees, unlimited depth, minimum 2 samples per split	Captures nonlinear relations; robust to noise	May smooth extremes; computationally heavier
ET	100 randomized decision trees, use all features for splits	Fast, decorrelated trees; reduces variance	Less stable for small sample sizes
BR	Assumes linear relations with uninformative priors (flat assumptions on weights and noise)	Probabilistic, includes uncertainty estimates	Assumes linear relations
Ridge	Linear regression with moderate regularization strength (α = 1.0)	Stable under multicollinearity; interpretable	Limited to linear dependencies
MLP	One hidden layer with 100 nodes, ReLU activation, trained up to 1000 iterations	Flexible nonlinear mapping	Requires more data; sensitive to initialization
XGB	100 boosted trees, learning rate = 0.1, depth = 6, full sample usage	Handles missing values internally; strong predictive power	Risk of overfitting; tuning-sensitive
LGBM	100 boosted trees, learning rate = 0.1, 31 leaves per tree, unlimited depth	Efficient on large datasets; handles categorical splits	Requires careful parameter choices for stability

To ensure robustness, we clarify that the eight imputation methods were implemented to construct an ensemble, with a standardized set of hyperparameters applied across all models (Table 2). This approach emphasizes methodological diversity rather than fine-tuned optimization of individual algorithms. Each method contributes distinct strengths: tree-based approaches capture nonlinear dependencies, linear regressors provide interpretable baselines, and neural networks offer flexible mapping of complex relationships. Because no single method performed optimally across all stations and time periods, we combined their outputs into an ensemble mean. This averaging reduces model-specific biases while preserving the dominant physical and statistical signals, resulting in reconstructions that are both stable and transferable to other coastal regions with sparse tide-gauge coverage.

The resulting reconstructed timeseries for each model are shown in Figure 3. An ensemble mean across the eight ML outputs was computed to generate a robust, gap-filled dataset (Figure 4). This ensemble averaging helped minimize overfitting and model-specific noise, yielding a stable representation of sea level variability across the region. Using this ensemble dataset, monthly sea level at Dokdo was estimated using a spatial interpolation approaches: IDW, where geodesic distances from Dokdo to each station were used to assign weights (Figure 5 and Equation (1)). To reconstruct sea level $S(x)$ at location x (e.g., Dokdo) from surrounding stations:

$$S(x) = \frac{\sum_{i=1}^{N} \frac{S_i}{d_i}}{\sum_{i=1}^{N} \frac{1}{d_i}} \quad (1)$$

where S_i, d_i, and N are sea level at stations i, distance from station i to location x, number of nearby stations used (e.g., 13). Stations in closer proximity to Dokdo, such as Ulleung, Mukho, and Saigo, received greater influence in the IDW-based estimation. This dual-method approach provided two physically plausible reconstructions for Dokdo.

To assess the physical consistency of the reconstructed Dokdo sea level series, we compared both the Mean- and IDW-based estimates against satellite altimetry sea level anomalies from the CMS at the Dokdo location. All timeseries were shifted vertically to facilitate intercomparison (Figure 5). The reconstructions captured the seasonal cycle and interannual variability with strong coherence across the three datasets, especially from the early 2000s onward. Discrepancies in the early 1990s and post-2019 periods reflect a combination of factors, including early-year data sparsity, potential inconsistencies in vertical datum, and the coarser spatial resolution of gridded altimetry. Nonetheless, the strong agreement validates the statistical reconstructions and affirms their suitability for scientific analysis in data-sparse oceanic settings. To compare sea level variability at Dokdo with the 13 selected PSMSL tide gauge stations, we extracted satellite-derived sea level anomaly (SLA) from the CMS monthly product (1993–2023) at 13 PSMSL stations and the grid point closest to Dokdo (37.241° N, 131.865° E). All station series were resampled to a monthly resolution and visualized as vertically offset timeseries, using fixed offsets based on geodesic distance from Dokdo. A uniform negative offset was assigned to the Dokdo timeseries to clearly separate it from the others. The final figure includes station-specific labels and color coding, enabling visual comparison of seasonal and interannual variability patterns across the 14 series.

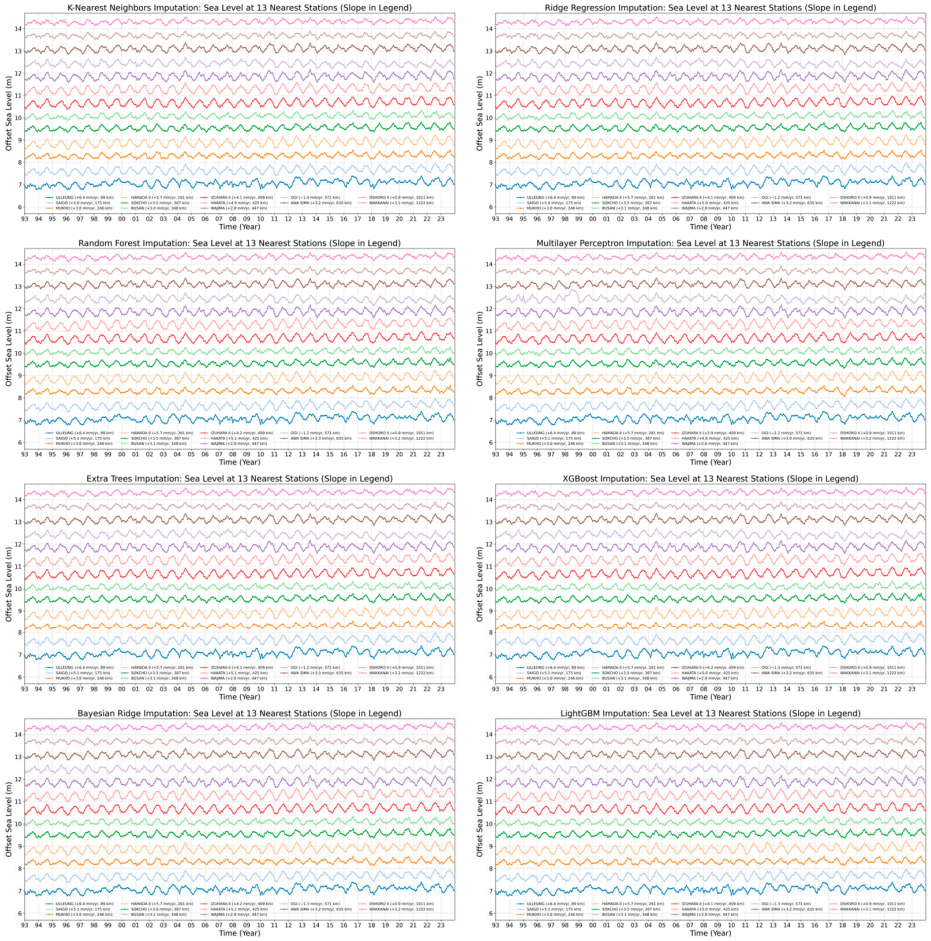


Figure 3. Sea level time series (1993–2023) at the 13 selected reference PSMSL tide gauge stations around Dokdo, reconstructed using eight ML imputation models: KNN, RF, ET, BR, Ridge, MLP, XGBoost, and LightGBM. Vertical offsets are applied for visual clarity. Each subplot shows the imputed series for all stations using one model, with the corresponding linear trend (slope in mm/yr) and geodesic distance to Dokdo (in km) shown in the legend. All models fill data gaps to produce continuous series, preserving interannual and seasonal variability while differing slightly in long-term trend estimates.

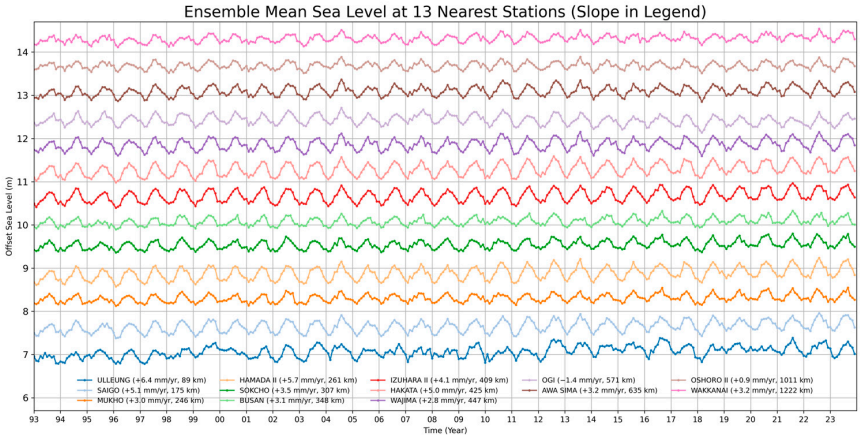


Figure 4. Ensemble mean sea level at the 13 selected PSMSL stations around Dokdo from 1993 to 2023, reconstructed using eight ML imputation models. Vertical offsets are applied for visual clarity. The legend shows each station's linear trend (slope in mm/yr) and geodesic distance to Dokdo (in km). The ensemble mean preserves seasonal cycles and reveals consistent long-term sea level rise at most stations, with ULLEUNG showing the highest trend (+6.4 mm/yr) and OGI the lowest (−1.4 mm/yr). These results serve as the basis for subsequent sea level reconstruction at Dokdo.

To evaluate the reliability of the reconstructed sea level data and assess consistency with independent observations, we compared the inverted barometric effect (IBE)-corrected PSMSL reconstructions with satellite altimetry estimates from the CMS (Equation (2) and Figure 6). To correct the sea level for mean sea level atmospheric pressure:

$$\Delta h = \frac{P_{obs} - P_{ref}}{\rho g} \tag{2}$$

where P_{obs}, P_{ref}, ρ, and g are observed mean sea level atmospheric pressure, reference pressure (e.g., 1013.25 hPa), seawater density, and gravity. Monthly CMS SLA data from 1993 to 2023 were extracted at the coordinates of the 13 selected PSMSL stations and at Dokdo, using nearest-neighbor sampling from the 0.25° × 0.25° gridded product (Figure 7). For Figure 8, we constructed an offset timeseries plot comparing the monthly sea level at each of the 14 stations (13 PSMSL + Dokdo) from both PSMSL and CMS. PSMSL records were first gap-filled using the ensemble mean of eight ML models and corrected for the IBE using ERA5 mean sea level atmospheric pressure data. Each station's PSMSL timeseries was plotted with a solid line, and the corresponding CMS altimetry timeseries with a dashed line. To ensure visual clarity and facilitate inter-station comparison, vertical offsets were applied based on the geodesic distance from each station to Dokdo, with Dokdo plotted at the bottom using a fixed offset. Colors and line styles were standardized across both datasets for station-level coherence. The legend was simplified to include only station names (e.g., Ulleung, Saigo, …, Dokdo) in geodesic order. For Figure 9, we specifically focused on Dokdo to compare the reconstructed sea level series derived from the IDW-weighted average of the 13 filled PSMSL stations (after IBE correction) with the CMS altimetry record at the same location. The IDW method applied geodesic distance-based weights to each contributing station. Both series were plotted together as monthly timeseries, using a red solid line for the reconstructed PSMSL (IDW) estimate and a black dashed line for the CMS altimetry product. This comparison enabled a direct assessment of temporal coherence, seasonal cycle fidelity, and interannual variability between the in situ-derived and satellite-based records.

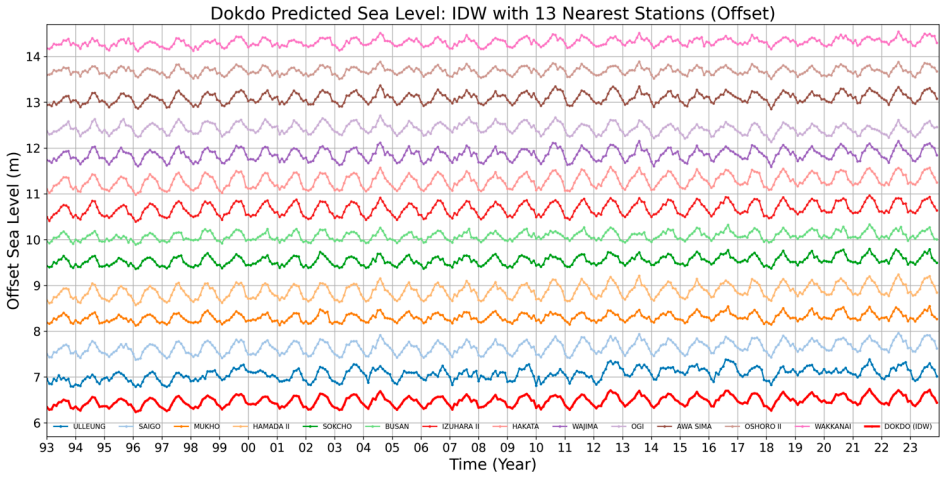


Figure 5. Offset timeseries of monthly mean sea level (1993–2023) from the 13 selected PSMSL stations used to predict sea level at Dokdo, along with the IDW-based estimate. Each curve is vertically shifted for clarity. The red line represents the IDW prediction at Dokdo, highlighting consistent seasonal cycles and multiyear variability compared to surrounding stations.

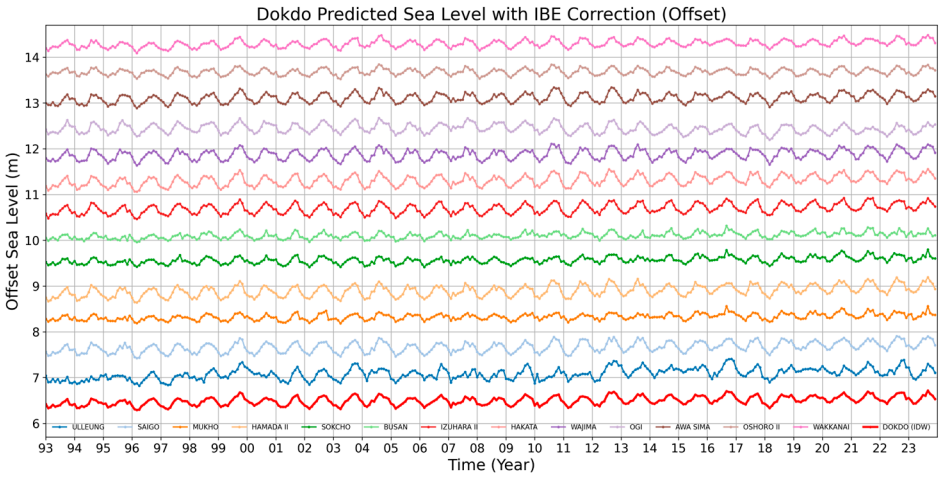


Figure 6. Offset timeseries of monthly mean sea level (1993–2023) from the 13 selected PSMSL stations after applying IBE correction, along with the IDW-based prediction at Dokdo. Each station's timeseries is vertically offset for visual clarity. The red line shows the IBE-corrected sea level at Dokdo estimated using IDW, capturing consistent seasonal and interannual variability in alignment with surrounding stations.

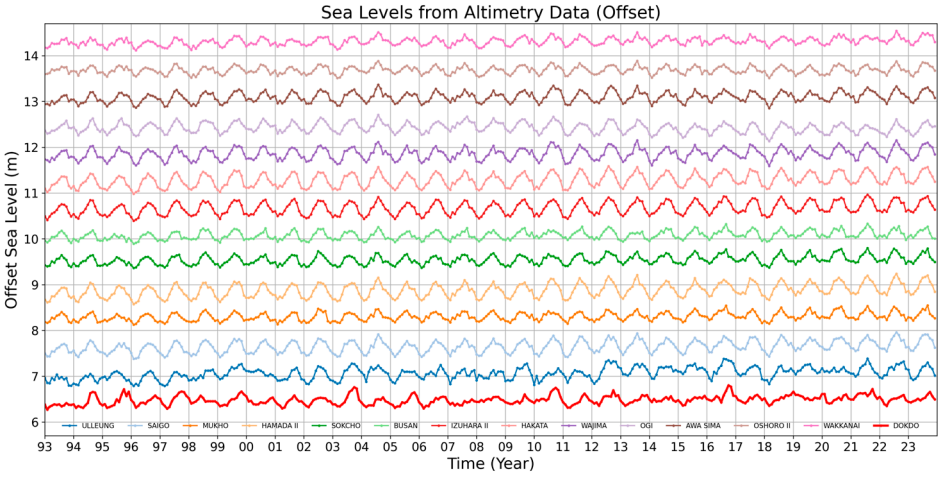


Figure 7. Monthly sea level timeseries (1993–2023) across 13 PSMSL tide gauge stations and Dokdo (red line) from satellite altimetry, shown with vertical offsets. The altimetry-derived SLA at Dokdo was obtained from the CMS dataset, re-referenced to match the ensemble mean of the 13 station records. The offset layout allows for visual comparison of seasonal and interannual variability across locations. Dokdo is placed at the bottom for emphasis and clarity.

To assess uncertainty in the reconstructed sea-level record at Dokdo, we implemented a Monte Carlo bootstrap framework with 1000 replicates. In each replicate, both the imputation models (drawn from the set of eight ML algorithms) and the contributing tide-gauge stations (drawn with replacement and inverse-distance weighting) were resampled. This procedure produced an ensemble of Dokdo reconstructions, from which we derived 95% confidence intervals (CIs) for the monthly time series and the linear trend. For validation against CMS altimetry, we employed a 12-month moving-block bootstrap applied to paired IDW–altimetry series. This approach accounts for temporal autocorrelation and yielded 95% CIs for skill metrics, including RMSE, MAE, correlation, and R^2. To test robustness during periods of sparse input data, we computed performance metrics for the Dokdo reconstruction against CMS altimetry in both the full overlap period (1993–2023) and the early period (1993–2000). Metrics were evaluated using RMSE, MAE, correlation, and R^2, with uncertainty estimated from a 12-month moving-block bootstrap (N = 1000) to account for temporal autocorrelation.

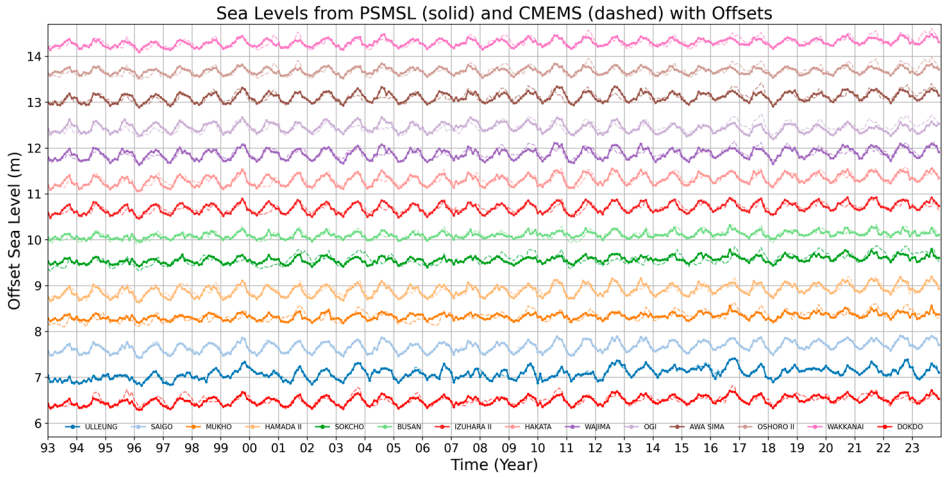


Figure 8. Monthly sea level timeseries (1993–2023) from 13 PSMSL tide gauge stations and reconstructed Dokdo sea level, compared with collocated satellite altimetry (CMS) estimates. Solid lines represent PSMSL (or IDW-based Dokdo) sea level data, while dashed lines show CMS sea level anomalies extracted at the same locations. Each station is vertically offset for clarity, with Dokdo plotted at the bottom in red. The agreement in seasonal and interannual variability highlights the consistency between in situ and altimeter observations across the Northwest Pacific.

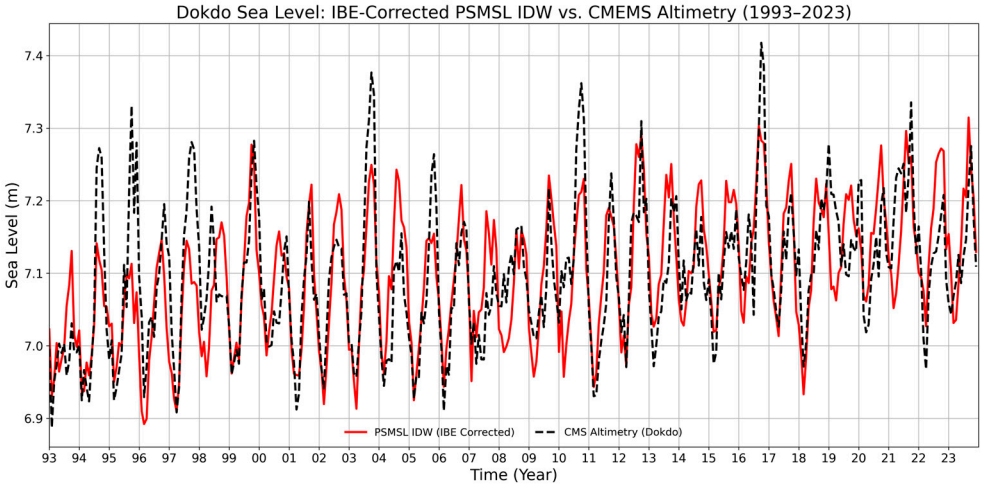


Figure 9. Monthly sea level at Dokdo (1993–2023) reconstructed using IBE-corrected PSMSL data with inverse distance weighting (IDW, red line) compared with CMS satellite altimetry (black dashed line). Linear trend lines and corresponding statistics (slope and coefficient of determination, R^2) are shown in the legend. The IDW reconstruction shows a stronger long-term trend (4.64 mm/yr, R^2 = 0.21) than CMS (3.41 mm/yr, R^2 = 0.10), highlighting the influence of nearshore spatial interpolation and the limitations of satellite altimetry in resolving localized signals. Seasonal and interannual variability are well captured in both datasets.

3. Results

The original monthly sea level records from the 13 selected PSMSL tide gauge stations to Dokdo (Figure 2) reveal a pronounced seasonal cycle, with both annual and semiannual components, and notable interannual fluctuations. However, these series also contain heterogeneous data gaps of varying lengths and temporal distribution, particularly during the early 1990s. Distant stations, such as Izuhara II and AWA SIMA, show more pronounced discontinuities and variability in amplitude, underlining the need for systematic gap-filling prior to spatial analysis and prediction efforts.

To address these issues, eight ML models, KNN, RF, ET, BR, Ridge, MLP, XGBoost, and LightGBM, were applied using a unified Iterative Imputer framework. As shown

in Figure 3, the reconstructed timeseries across all stations were continuous, physically realistic, and retained seasonal cycles. Minor differences among models were observed, especially in earlier periods where training data were sparse. Nonetheless, all models succeeded in reproducing coherent interannual patterns and reducing noise inherent in the original records.

The ensemble mean of these imputed series (Figure 4) provided a robust representation of regional sea level variability. By averaging across multiple model reconstructions, the ensemble approach smoothed model-specific idiosyncrasies while preserving meaningful signals. The timeseries clearly reflected multiyear trends and recurring seasonal fluctuations, reinforcing the reliability of the ensemble dataset for further spatial interpolation and analysis.

Using the ensemble-filled data from the 13 stations, sea level at Dokdo was reconstructed via a complementary approach: an IDW interpolation (Figure 5). Both reconstructions yielded consistent seasonal and interannual signals, with prominent peaks during late summer (August–September) and troughs in winter, likely reflecting regional steric sea level responses to thermal expansion. The IDW-based estimate emphasized contributions from geographically closer stations such as Ulleung and SAIGO, leading to slightly enhanced short-term variability compared to the broader regional averaging in the mean-based approach. Despite this, both methods captured the long-term trend and seasonal modulation, supporting the validity of proxy-based Dokdo sea level estimation using surrounding tide gauge records.

3.1. Original Sea Level Records at 13 Selected Stations

Figure 2 presents the original monthly sea level records from 1993 to 2023 for the 13 tide gauge stations obtained from the PSMSL data. These stations were not chosen solely based on geographic distance to Dokdo, but rather because they are located in the surrounding region where vertical land motion is relatively small, making them more suitable as references for regional sea-level reconstruction. Geodesic distances from each station to Dokdo were computed and are annotated in Figure 2. Each time series is vertically offset for clarity. The records exhibit consistent seasonal cycles across stations, underscoring strong regional oceanographic connectivity. However, significant temporal gaps, particularly during the 1990s and early 2020s, limit the direct calculation of long-term trends. These discontinuities were most frequent in the more distant and northern stations, highlighting the challenge of relying solely on raw observational datasets for regional sea level assessments.

3.2. Gap-Filled Ensemble Sea Level at 13 Stations

To address the limitations posed by missing data, we applied a suite of ML methods, including KNN, RF, ET, BR, Ridge, MLP, XGBoost, and LightGBM, to reconstruct the sea level time series. As shown in Figures 3 and 4, the ensemble mean of these model outputs provides a continuous and smoothed representation of sea level variability across all stations. Each reconstructed series preserves the original seasonal patterns while eliminating temporal gaps. The imputed datasets capture both the local variability and the broader climatological signals associated with steric and mass-driven processes in the East Sea and East China Sea. Importantly, slope estimates (in mm/yr) derived from the filled time series are indicated in the legends.

3.3. Dokdo Sea Level Prediction Using Regional Observations

Dokdo's monthly sea level timeseries for 1993–2023 was estimated using the gap-filled records from the surrounding 13 stations through an interpolation strategy: inverse distance weighting. Figure 5 compares these reconstructions. While both approaches

reproduced the regional seasonal cycle and long-term variations, subtle differences emerged. The IDW prediction, favoring closer stations like Ulleung and Saigo, introduced higher-frequency fluctuations, whereas the mean-based prediction smoothed out local anomalies by averaging over all stations equally. These distinctions underscore the importance of methodological transparency in regional sea level predictions, especially when local processes may diverge from basin-wide trends.

Together, the results demonstrate that ML-enhanced imputation and ensemble interpolation offer a reliable framework for reconstructing realistic and continuous sea level timeseries in data-sparse locations like Dokdo. The predicted series preserves key regional signals and serves as a basis for future studies of coastal sea level change, hazard assessment, and climate variability.

Figure 6 provides validation using CMS satellite altimetry data at Dokdo, with all timeseries normalized and corrected for the IBE. The reconstructed Dokdo series—especially the IDW version—closely tracks altimetric variability, confirming physical realism and capturing regional seasonal dynamics. Figure 7 illustrates the monthly sea level variability from 1993 to 2023 across 13 PSMSL tide gauge stations and Dokdo form CMS altimetry. After aligning the altimetry-derived SLA to the mean of the PSMSL ensemble, the Dokdo timeseries was plotted alongside the others with vertical offsets for clarity. All stations exhibit strong annual cycles and coherent interannual variability, suggesting regionally consistent oceanographic forcing. The altimetry-derived sea level at Dokdo shows a well-defined seasonal signal and multi-year fluctuations that align closely with nearby stations such as Ulleung, Saigo, and Mukho. This agreement supports the physical consistency of the CMS product and validates the use of satellite altimetry in reconstructing sea level trends at locations without in situ observations.

3.4. Comparison Between PSMSL and CMS Sea Levels

Figure 8 presents the monthly sea-level time series from 13 PSMSL tide-gauge stations and their corresponding CMS altimetry values, along with the reconstructed sea level at Dokdo shown at the bottom. For visualization, all timeseries were vertically offset, with each station represented by two curves: solid lines for IBE-corrected, gap-filled PSMSL records and dashed lines for CMS altimetry. The comparison demonstrates strong seasonal consistency across datasets, with annual and semiannual cycles well captured. However, differences emerge in interannual variability and long-term trends at several stations, particularly at more distant sites such as Wakkanai and Oshoro II. In contrast, stations geographically closer to Dokdo, including Ulleung and Saigo, show higher consistency between the two datasets, affirming their reliability as predictors in the reconstruction framework.

A key factor underlying residual differences between tide-gauge (TG) and satellite altimetry (SA) trends is Vertical Land Motion (VLM). Tide gauges measure relative sea level, which can be affected by subsidence, uplift, or tectonic processes, whereas altimetry provides absolute sea level referenced to a global geodetic datum. Without correcting for VLM, discrepancies between TG and SA records may reflect crustal motion rather than true oceanographic signals. Foundational global syntheses and reviews demonstrate that VLM can either amplify or attenuate relative sea-level change, underscoring its central role in reconciling TG–SA trends [55,56]. More recent studies highlight the added complexity of nonlinear VLM processes: probabilistic reconstructions reveal that tectonic activity, groundwater extraction, and surface loading can strongly modulate regional projections, altering relative sea level by up to 50 cm and increasing projection uncertainty by nearly 1 m in projections extending to the year 2150 [57,58]. At the same time, advanced noise reduction techniques applied to TG and SA records, combined with GNSS-based VLM

corrections, have been shown to substantially improve consistency and reduce long-term uncertainties [59]. Although VLM rates are relatively modest in the East Sea compared to tectonically active margins, explicitly accounting for them remains important when reconciling TG- and SA-based records. While the present study focuses on methodological development and validation, future extensions should systematically incorporate GNSS-constrained VLM corrections wherever available to achieve absolute consistency between tide-gauge reconstructions and altimetry-derived trends.

3.5. Comparison of Reconstructed and Satellite-Derived Sea Level at Dokdo

Figure 9 directly compares the IBE-corrected monthly sea level at Dokdo reconstructed via IDW from 13 surrounding PSMSL tide gauge stations with CMS satellite altimetry data at the Dokdo location for the period 1993–2023. Both timeseries exhibit consistent seasonal cycles, confirming the reliability of the interpolation and the physical basis of the seasonal signal. However, distinct differences emerge in interannual variability, particularly during the 1990s, early 2000s, and post-2015, where the CMS record shows greater amplitude fluctuations compared to the smoother PSMSL-derived reconstruction. These discrepancies are likely attributable to higher noise levels or transient mesoscale oceanographic features captured by satellite altimetry but unresolved by the IDW method, which emphasizes spatially averaged signals.

Despite these variations, the overall agreement between the two datasets supports the use of the IBE-corrected and ML-gap-filled PSMSL data as a realistic proxy for sea level variability at Dokdo. Trend analysis indicates that the reconstructed IDW series shows a slightly stronger rise of 4.64 mm/yr (R^2 = 0.21) compared to 3.41 mm/yr (R^2 = 0.10) in the CMS dataset. The lower R^2 values reflect the influence of strong intra- and interannual variability, yet the close alignment of long-term trends reinforces the robustness of the reconstruction. These findings highlight the potential of tide gauge–based regional interpolation, especially when satellite records are limited by coastal proximity, data gaps, or geophysical corrections.

3.6. Uncertainty of Sea Level at Dokdo

The Monte Carlo bootstrap analysis (N = 1000) indicates that the typical monthly reconstruction uncertainty is ±3.3 cm, with narrower ranges in later years (median half-width: 3.7 cm in 1993–2000 vs. 3.2 cm in 2007–2023). The reconstructed 1993–2023 trend at Dokdo is 4.64 mm yr^{-1} (95% CI: 3.57–5.20 mm yr^{-1}). Validation against CMS altimetry (Table 3) shows an RMSE of 0.067 m (95% CI: 0.061–0.075 m), an MAE of 0.053 m (0.048–0.059 m), and a correlation of 0.743 (0.674–0.794) over the full 1993–2023 overlap, confirming robust agreement within the expected uncertainty bounds. Performance remains stable across decadal sub-periods, with RMSE values ranging from 0.064 to 0.073 m and correlation coefficients from 0.667 to 0.778, demonstrating that overfitting is minimal even in early sparse data windows.

To assess robustness under extreme conditions, we evaluated the reconstruction against CMS altimetry during (i) the strong 1997–1998 El Niño and (ii) the August–October 2003 typhoon season (Table 4). During El Niño, the reconstruction achieved an RMSE of 0.096 m (0.092–0.116) and correlation r = 0.572 (0.060–0.897). During the 2003 typhoon season, the short evaluation window yielded higher variance but still strong correlation (RMSE = 0.110 m, MAE = 0.108 m, r = 0.916). These results suggest that the hybrid framework maintains predictive skill even during anomalous sea-level states. Taken together, the quantitative evaluation confirms that the Dokdo reconstruction achieves consistent agreement with altimetry (RMSE ≈ 0.07 m, MAE ≈ 0.05 m, r ≈ 0.7), with bootstrap confidence intervals confirming the stability of the metrics. The moving-block bootstrap further

ensures robustness to temporal dependence, while Monte Carlo resampling of gauges and models demonstrates that no single predictor dominates the reconstruction.

Table 3. Performance statistics of CMS-reconstructed sea level against PSMSL (IBE-corrected) observations at Dokdo. Metrics include root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (r), and coefficient of determination (R^2). All metrics are reported with 95% confidence intervals derived from 1000 moving-block bootstrap resamples (block length = 12 months) to account for temporal dependence. Results are presented for the full overlapping period (1993–2023) and three decadal sub-periods, thereby providing both an assessment of reconstruction skill and an explicit quantification of uncertainty.

Window	N (Months)	RMSE (m) [95% CI]	MAE (m) [95% CI]	r [95% CI]	R^2 [95% CI]
1993–2023	372	0.067 (0.061–0.075)	0.053 (0.048–0.059)	0.743 (0.674–0.794)	0.46 (0.28–0.55)
1993–2002	120	0.073 (0.059–0.092)	0.054 (0.044–0.072)	0.703 (0.588–0.794)	0.21 (−0.51–0.53)
2003–2012	120	0.064 (0.056–0.073)	0.051 (0.044–0.061)	0.778 (0.648–0.846)	0.48 (0.22–0.57)
2013–2023	132	0.065 (0.057–0.074)	0.053 (0.046–0.061)	0.667 (0.486–0.766)	0.32 (−0.04–0.52)

Table 4. Performance statistics of the Dokdo reconstruction during extreme sea level events. Shown are the RMSE, MAE, r, and R^2 between PSMSL-based reconstructions and CMS altimetry at Dokdo, each reported with 95% confidence intervals obtained from 1000 moving-block bootstrap resamples. Results are presented for the 1997–1998 El Niño and the August–October 2003 typhoon season, highlighting the robustness of the methodology under anomalous sea-level conditions.

Event	Window	RMSE (m) [95% CI]	MAE (m) [95% CI]	r [95% CI]	R^2 [95% CI]
El Niño 1997–1998	1997-01 to 1998-12	0.096 (0.092–0.116)	0.081 (0.078–0.106)	0.572 (0.060–0.897)	−0.49 (−2.98–−0.49)
Typhoon ASO 2003	2003-08 to 2003-10	0.110 (0.110–0.110)	0.108 (0.108–0.108)	0.916 (0.916–0.916)	−8.12 (−8.12–−8.12)

4. Discussion

This study demonstrates a comprehensive framework for reconstructing sea-level time series in data-sparse coastal environments, combining tide-gauge observations, inverse barometric correction, ML-based imputation, ensemble averaging, and geospatial interpolation. Applied to Dokdo in the East Sea, the framework successfully reproduced monthly sea level for 1993–2023 and showed robust agreement with satellite altimetry, including during anomalous events such as the 1997–1998 El Niño and the 2003 typhoon season.

Placing these results in a broader climate context is essential. While the 30-year window captures prominent events such as the 1997–1998 El Niño and the early-2000s PDO phase shift, it is too short to resolve full multi-decadal variability. Longer tide-gauge records from the Northwest Pacific (100–150 years) show rise rates broadly consistent with our Dokdo trend of 4.6 mm yr^{-1} (95% CI: 3.6–5.2 mm yr^{-1}), but also reveal accelerations and slowdowns linked to ENSO, PDO, and IOD. Embedding our reconstruction within this broader variability highlights the need to extend the framework to longer records and integrate explicit climate-mode predictors.

Our results therefore provide two key insights: (i) the proposed hybrid framework is robust and transferable to data-sparse coastal sites, even under stress conditions such as ENSO anomalies or typhoon seasons, and (ii) integration with centennial tide-gauge reconstructions and reanalysis-based predictors will be critical to disentangle background

sea-level rise from climate-driven variability, ultimately improving projections for marginal seas like the East Sea.

4.1. Ensemble-Based Reconstruction of Incomplete Sea Level Records

The diversity of missing data patterns across the 13 tide gauge records necessitated a robust and flexible imputation strategy. We implemented eight ML models, ranging from tree-based regressors (Random Forest, XGBoost, Extra Trees) to neural networks (MLP) and linear models (Bayesian Ridge, Ridge), within an IterativeImputer framework to reconstruct monthly sea level data. All models captured seasonal and interannual variability to varying degrees, and the ensemble mean across models further stabilized the timeseries by mitigating individual model biases and noise. As shown in Figure 4, the ensemble retained physical realism and reproduced dominant regional signals while minimizing overfitting.

4.2. Spatial Interpolation and Sea Level Estimation at Dokdo

The imputed and ensemble-averaged timeseries were then used to estimate sea level at Dokdo, a location lacking long-term direct observations. An interpolation method was applied: IDW. The consistency between the two Dokdo reconstructions underscores the spatial coherence of regional sea level, with nearby stations like Ulleung and SAIGO exerting the strongest influence. Notably, the IDW reconstruction was more sensitive to local extremes (e.g., the sharp September 2023 peak), which aligns with its design to emphasize proximal stations. This trade-off between smoothing and localized sensitivity highlights the importance of interpolation choice for marginal-sea reconstructions.

4.3. Comparison with CMS Altimetry and Post-Correction Residuals

To validate the reconstructed sea level at Dokdo, we compared the IBE-corrected timeseries derived from PSMSL-based interpolation with satellite-derived sea surface height anomalies from CMS altimetry (Figure 9). While overall agreement was strong, especially after IBE correction, residual amplitude differences remained. These stem from the inherent characteristics of each dataset: tide gauge-based reconstructions reflect nearshore processes, including coastal setup and wave-driven effects, while satellite altimetry provides spatially averaged offshore conditions. Consequently, discrepancies in timing and amplitude during high-energy periods reflect both observational scale differences and the smoothing effect of altimetric measurements. We note that applying the full DAC field, which accounts for both atmospheric pressure and wind-driven high-frequency ocean responses, can further improve the consistency of tide-gauge and satellite altimetry time series. In our study, the IB correction was applied as the dominant DAC term, but DAC fields (e.g., from PSMSL) were not available for all stations. Recent studies confirm that incorporating DAC improves tide-gauge/altimetry consistency and reduces uncertainty in long-term sea-level trends [59]. Future extensions of our framework should therefore explicitly integrate DAC-corrected products to refine regional reconstructions in the Sea of Japan.

4.4. Benefits of ML and Ensemble Approaches

ML-based imputation proved particularly beneficial in reconstructing records from the early 1990s, a period with more frequent data gaps. Models such as Random Forest and XGBoost performed well in reproducing known values but risked overfitting post-2000, when fewer gaps remained. The ensemble mean not only generalized better but also reduced high-frequency noise and outlier sensitivity. This approach reinforces the value of ensemble learning in geophysical timeseries reconstruction, especially in semi-enclosed seas where station coverage is sparse and environmental variability is high.

4.5. Broader Implications and Methodological Transferability

The methodology developed in this study is adaptable to other coastal or marginal seas with limited tide gauge coverage. By combining data-driven reconstruction with physically informed corrections (e.g., IBE), the framework supports the generation of proxy sea level datasets where direct measurements are unavailable. These reconstructions can aid climate monitoring, coastal hazard forecasting, and satellite product validation. Furthermore, the ensemble-based workflow allows for modular updates as new stations, satellite datasets, or reanalysis products become available.

4.6. Limitations and Future Directions

While the proposed framework demonstrates strong utility, several limitations remain. Reconstruction reliability is inherently dependent on the density and quality of available tide-gauge records. In its current form, the method produces deterministic outputs without explicit uncertainty bounds. To address this, we implemented a Monte Carlo bootstrap (N = 1000) to quantify uncertainty in both the monthly reconstructions and the long-term Dokdo trend, yielding an estimate of 4.6 mm yr^{-1} with a 95% confidence interval of 3.6–5.2 mm yr^{-1}. Nevertheless, more advanced approaches, such as Bayesian ensembling, dropout-based uncertainty estimation, or integrated bootstrap schemes, should be explored in future work.

A further opportunity lies in expanding the predictor set beyond sea-level observations. Physically relevant variables such as sea surface temperature, wind stress, and steric height from ocean reanalysis products could improve physical realism, particularly during extreme events. While our framework is robust in its current form, incorporating such predictors may provide incremental benefits and should be prioritized in future reconstruction studies across marginal seas.

4.7. Toward Hybrid Sea Level Digital Twins

The convergence between corrected in situ reconstructions and satellite observations highlights the potential of hybrid approaches for digital twin development in coastal monitoring systems. By blending observational datasets, physical corrections, and ML-driven inference, this study lays the groundwork for scalable and interpretable sea level digital twins. Such systems can not only reconstruct historical records but also support real-time forecasting and long-term trend analysis in vulnerable regions like the East Sea.

Figure 10 summarizes the end-to-end workflow of this study, illustrating how machine learning-based imputation, IBE correction, and spatial interpolation were combined to reconstruct sea level at Dokdo in the absence of direct tide gauge measurements. The integration of eight ML models enabled robust reconstruction of missing values in the PSMSL dataset, while the IDW method leveraged spatial coherence from nearby stations. By applying IBE correction and validating the reconstructed Dokdo series against satellite altimetry, the framework not only ensures physical consistency but also demonstrates the viability of ML-driven coastal monitoring in data-sparse regions. Finally, while our hybrid reconstruction–altimetry framework demonstrates strong skill at the event to decadal scale, it is equally important to situate these results within longer-term climate variability, which we address in Section 4.8.

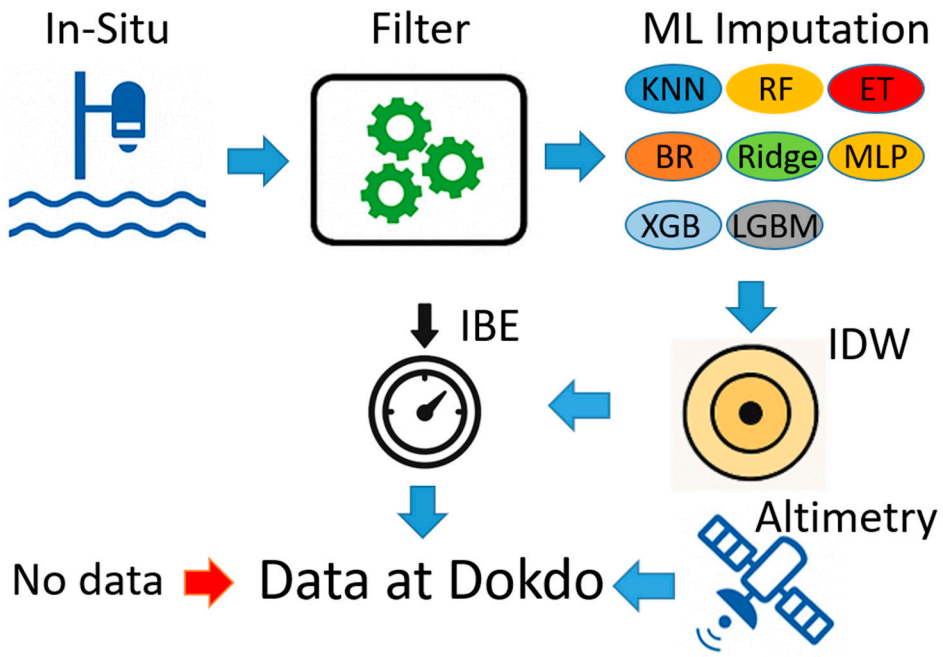


Figure 10. Schematic diagram of the sea level reconstruction workflow at Dokdo (Dok Island) using in situ PSMSL tide gauge observations and ML-based imputation models. After preprocessing and outlier filtering, missing data were filled using eight machine learning algorithms (KNN, RF, ET, BR, Ridge, MLP, XGB, LGBM), and ensemble averaging was applied. An inverse barometric effect (IBE) correction was incorporated. Since no tide gauge exists at Dokdo, sea level at Dokdo was estimated using inverse-distance weighting (IDW) from the 13 selected stations. The reconstructed timeseries was validated against satellite altimetry data at the Dokdo location.

4.8. Long-Term Climate Context

While this study focuses on a 31-year reconstruction (1993–2023), it is important to situate the results within the broader context of centennial- and millennial-scale sea-level variability. Long PSMSL tide-gauge records from the Northwest Pacific and global reconstructions extending back to the late 19th and early 20th centuries indicate rise rates of ~1.1 ± 0.7 mm yr^{-1} during 1880–1935, increasing to ~1.8 ± 0.3 mm yr^{-1} during 1936–2009, with an overall 20th-century trend of ~1.7 ± 0.2 mm yr^{-1} (Figure 5 of [60]). These centennial-scale rates are broadly consistent with our Dokdo estimate of 4.6 mm yr^{-1} (95% CI: 3.6–5.2 mm yr^{-1}) for 1993–2023, reflecting both the recent acceleration and the superimposed influence of multi-decadal climate modes such as the PDO, IOD, and ENSO, which can modulate local sea-level trends by several millimeters per year [36,61].

Our 1993–2023 window captures only part of this variability, including the strong 1997–1998 El Niño and the early-2000s PDO phase shift, but is too short to resolve complete multi-decadal cycles. Recent reconstructions (e.g., [36]) highlight how advanced statistical methods can separate long-term trend signals from climate-driven variability with ENSO- and PDO-related fluctuations capable of altering apparent sea-level rise by up to ~1 mm yr^{-1} over 5-year windows. Paleo-sea-level reconstructions and geological evidence further suggest that the background rise over the past millennium has been slower but punctuated by episodes of natural variability, underscoring the need to link satellite-era trends with centennial- and millennial-scale perspectives.

Related efforts include the study of Lee et al. [62], who developed a neural-network operator combining LSTM and data assimilation to reconstruct short-term ($\leq$72 h) gaps in Korean tide-gauge records, achieving sub-centimeter accuracy. Their framework demonstrates the value of neural operators for high-frequency interpolation in coastal settings. In contrast, our approach targets multi-decadal reconstructions at an offshore site (Dokdo)

by integrating tide-gauge data, satellite altimetry, and machine-learning imputation, with uncertainties explicitly quantified using Monte Carlo and moving-block bootstrap methods. These approaches are complementary: neural operators excel at resolving short-term continuity in dense observational networks, while our hybrid framework is designed to provide robust long-term reconstructions in data-sparse offshore environments.

Taken together, these studies provide a broader framework for situating our Dokdo reconstruction within the long-term trajectory of East Sea variability, linking the 30-year satellite era with centennial and millennial records. Future extensions of our methodology will therefore focus on assimilating longer tide-gauge records (>100 years where available), explicitly testing the added value of PDO/IOD indices and other reanalysis predictors, and benchmarking against paleoclimate reconstructions to strengthen projections for marginal seas.

5. Conclusions

This study developed and applied a hybrid reconstruction framework to generate a continuous, high-quality monthly sea level timeseries at Dokdo spanning 1993 to 2023. Leveraging in situ records from 13 nearby PSMSL tide gauge stations and an ensemble of advanced ML imputation techniques, we addressed the long-standing challenge of data sparsity in this strategically important region of the East Sea. By systematically filling data gaps and capturing spatial relationships among stations, the framework provides a robust and interpretable solution for reconstructing sea level variability in observationally limited coastal environments.

Several key insights emerged from this analysis. First, the original PSMSL station records, particularly those in northern locations, contained substantial missing segments, limiting their standalone use in regional sea level trend analysis. To overcome this, we implemented eight ML models spanning tree-based (e.g., Random Forest, XGBoost), linear (e.g., Ridge, Bayesian Ridge), and neural network (e.g., MLP) approaches. These models successfully reconstructed realistic timeseries that preserved key features such as seasonal cycles and interannual variability. Second, ensemble averaging across these models effectively reduced individual model biases and smoothed high-frequency noise, producing stable and reliable inputs for spatial prediction. Using an inverse-distance weighting (IDW), we estimated sea level at Dokdo with high fidelity. These estimates showed strong agreement with neighboring stations, particularly Ulleung and Saigo, underscoring the spatial coherence of sea level in this region. Third, inverse barometric effect (IBE) correction using ERA5 mean sea level pressure further improved agreement with CMS satellite altimetry. Despite this improvement, some discrepancies remained, especially in high-frequency components, likely due to satellite limitations near the coast and unresolved local atmospheric and bathymetric influences.

Complementary to recent neural-operator approaches that achieve high accuracy in short-gap interpolation of tide-gauge data (e.g., [62]), our framework provides robust long-term reconstructions in data-sparse offshore settings, underscoring the value of integrating multiple methodologies across different temporal and spatial scales. Our Dokdo reconstruction for 1993–2023 is also broadly consistent with centennial-scale tide-gauge reconstructions and paleo-sea-level evidence, which show an acceleration in regional and global sea-level rise since the early 20th century. By linking the satellite era with longer records extending over the past 100 to 1000 years, this study provides a bridge between short-term variability and long-term trends, strengthening the basis for future projections in the East Sea and other marginal seas.

Overall, this study demonstrates that a hybrid methodology, combining ML-based gap filling, ensemble modeling, spatial interpolation, and physical correction, can effec-

tively reconstruct sea level in remote island and marginal sea settings. The workflow, illustrated in Figure 10, is transparent, modular, and scalable, and can be readily applied to other regions with sparse observational coverage. Looking ahead, the framework can be extended to integrate satellite altimetry and physically relevant predictors from ocean reanalyses, such as steric height, wind stress, and sea surface temperature, which may further enhance reconstruction skill and physical realism, particularly for earlier time periods. In addition, extending the framework to include temporal forecasting using models like Prophet, Seasonal Auto Regressive Integrated Moving Average, and Long Short-Term Memory network may enable future scenario analysis, detection of seasonal anomalies, and preparation for extreme events. These developments can contribute to the advancement of digital twin systems for coastal monitoring and long-term adaptation planning in the context of accelerating sea-level rise.

Author Contributions: Conceptualization, M.H.; methodology, M.H.; software, M.H.; validation, M.H.; formal analysis, M.H.; investigation, M.H.; resources, M.H. and H.S.L.; data curation, M.H.; visualization, M.H.; writing—original draft preparation, M.H.; writing—review and editing, M.H. and H.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Trade, Industry, and Energy (MOTIE) of Korea under the "Regional Innovation Cluster Development Program (PN92750, P0025418)", supervised by the Korea Institute for Advancement of Technology (KIAT). It was also supported by the KIOST In-house Project (PEA0331), "Research of Prediction Techniques to Reduce Damage from Marine Disasters". Additionally, this study was supported by the project "Sustainable Research and Development of Dokdo (PG54801)" of the Ministry of Oceans and Fisheries, Republic of Korea. This research was a part of the project titled 'Global Industry-Leading Research and Training Program for Innovative Human Resources in Marine Leisure Tourism (PM65040, RS-2025-02317535)', funded by the Ministry of Oceans and Fisheries, Republic of Korea.

Data Availability Statement: The tide-gauge records used in this study are publicly available from the Permanent Service for Mean Sea Level (PSMSL) at https://psmsl.org, accessed on 28 May 2025. Satellite altimetry products were obtained from the Copernicus Marine Service (CMS) at https://marine.copernicus.eu, accessed on 12 July 2025. All processed reconstruction datasets generated in this study, including the Dokdo sea-level time series (1993–2023) and associated performance metrics, are available from the corresponding author upon reasonable request. All data used in this study were collected by the authors.

Acknowledgments: Tide-gauge sea level data provided by the Permanent Service for Mean Sea Level (PSMSL; https://psmsl.org/data/obtaining, accessed on 28 May 2025), Satellite altimetry data (SSALTO/DUACS altimeter products) produced and distributed by the CMS (https://www.marine.copernicus.eu, accessed on 12 July 2025), and MSL atmospheric pressure data from the ECMWF Reanalysis v5 (http://www.ecmwf.int, accessed on 12 July 2025) were used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CMS	Copernicus Marine Service
IBE	inverted barometric effect
IDW	inverse distance weighting
ML	machine learning
PSMSL	Permanent Service for Mean Sea Level
SLA	sea level anomaly

References

1. Passeri, D.L.; Hagen, S.C.; Medeiros, S.C.; Bilskie, M.V.; Alizad, K.; Wang, D. The dynamic effects of sea level rise on low-gradient coastal landscapes: A review. *Earth's Futur.* **2015**, *3*, 159–181. [CrossRef]
2. Baker, R.; McGowan, S. Periodicities in mean sea-level fluctuations and climate change proxies: Lessons from the modelling for coastal management. *Ocean Coast. Manag.* **2014**, *98*, 187–201. [CrossRef]
3. Royston, S.; Bingham, R.J.; Bamber, J.L. Attributing decadal climate variability in coastal sea-level trends. *Ocean Sci.* **2022**, *18*, 1093–1107. [CrossRef]
4. Kim, Y.-B.; Chang, K.-I.; Park, J.-H.; Park, J.J. Variability of the Dokdo Abyssal Current observed in the Ulleung Interplain Gap of the East/Japan Sea. *Acta Oceanol. Sin.* **2013**, *32*, 12–23. [CrossRef]
5. Lee, D.E.; Kim, J.; Heo, Y.; Kang, H.; Lee, E.Y. Climate Change Implications Found in Winter Extreme Sea Level Height Records around Korea. *J. Mar. Sci. Eng.* **2021**, *9*, 377. [CrossRef]
6. Roy, P.; Pal, S.C.; Chakrabortty, R.; Chowdhuri, I.; Saha, A.; Shit, M. Effects of climate change and sea-level rise on coastal habitat: Vulnerability assessment, adaptation strategies and policy recommendations. *J. Environ. Manag.* **2023**, *330*, 117187. [CrossRef]
7. Lawrence, J.; Bell, R.; Blackett, P.; Stephens, S.; Allan, S. National guidance for adapting to coastal hazards and sea-level rise: Anticipating change, when and how to change pathway. *Environ. Sci. Policy* **2018**, *82*, 100–107. [CrossRef]
8. Kim, K.; Kim, Y.B.; Park, J.J.; Nam, S.; Park, K.-A.; Chang, K.-I. Long-term and real-time monitoring system of the East/Japan sea. *Ocean Sci. J.* **2005**, *40*, 25–44. [CrossRef]
9. Chang, K.-I.; Zhang, C.-I.; Park, C.; Kang, D.-J.; Ju, S.-J.; Lee, S.-H.; Wimbush, M. *Oceanography of the East Sea (Japan Sea)*; Springer: Cham, Switzerland, 2016. [CrossRef]
10. Holgate, S.J.; Matthews, A.; Woodworth, P.L.; Rickards, L.J.; Tamisiea, M.E.; Bradshaw, E.; Foden, P.R.; Gordon, K.M.; Jevrejeva, S.; Pugh, J. New Data Systems and Products at the Permanent Service for Mean Sea Level. *J. Coast. Res.* **2012**, *29*, 493. [CrossRef]
11. Woodworth, P.; Player, R. The permanent service for mean sea level: An update to the 21stCentury. *J. Coast. Res.* **2003**, *19*, 287–295.
12. Kadow, C.; Hall, D.M.; Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **2020**, *13*, 408–413. [CrossRef]
13. Zhao, T.; Wang, S.; Ouyang, C.; Chen, M.; Liu, C.; Zhang, J.; Yu, L.; Wang, F.; Xie, Y.; Li, J.; et al. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *Innovation* **2024**, *5*, 100691. [CrossRef]
14. Xu, T.; Zhang, Y.; Zhang, C.; Abodoukayimu, A.; Ma, D. Deep learning tool: Reconstruction of long missing climate data based on spatio-temporal multilayer perceptron. *Theor. Appl. Clim.* **2024**, *155*, 5835–5847. [CrossRef]
15. Appel, M. Efficient Data-Driven Gap Filling of Satellite Image Time Series Using Deep Neural Networks with Partial Convolutions. *Artif. Intell. Earth Syst.* **2024**, *3*, 220055. [CrossRef]
16. Bahari, N.A.A.B.S.; Ahmed, A.N.; Chong, K.L.; Lai, V.; Huang, Y.F.; Koo, C.H.; Ng, J.L.; El-Shafie, A. Predicting Sea Level Rise Using Artificial Intelligence: A Review. *Arch. Comput. Methods Eng.* **2023**, *30*, 4045–4062. [CrossRef]
17. Lu, G.Y.; Wong, D.W. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. [CrossRef]
18. Rougier, J. Ensemble Averaging and Mean Squared Error. *J. Clim.* **2016**, *29*, 8865–8870. [CrossRef]
19. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef]
20. Niazkar, M.; Menapace, A.; Brentan, B.; Piraei, R.; Jimenez, D.; Dhawan, P.; Righetti, M. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environ. Model. Softw.* **2024**, *174*, 105971. [CrossRef]
21. Wang, D.-N.; Li, L.; Zhao, D. Corporate finance risk prediction based on LightGBM. *Inf. Sci.* **2022**, *602*, 259–268. [CrossRef]
22. McDonald, G.C. Ridge regression. *WIREs Comput. Stat.* **2009**, *1*, 93–100. [CrossRef]
23. Shi, Q.; Abdel-Aty, M.; Lee, J. A Bayesian ridge regression analysis of congestion's impact on urban expressway safety. *Accid. Anal. Prev.* **2016**, *88*, 124–137. [CrossRef]
24. Zhang, S. Challenges in KNN Classification. *IEEE Trans. Knowl. Data Eng.* **2021**, *34*, 4663–4675. [CrossRef]
25. Mohammadi, M.R.; Sadrossadat, S.A.; Mortazavi, M.G.; Nouri, B. A brief review over neural network modeling techniques. In Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 21–22 September 2017; IEEE: New York, NY, USA, 2017; pp. 54–57.
26. Cazenave, A.; Henry, O.; Munier, S.; Delcroix, T.; Gordon, A.L.; Meyssignac, B.; Llovel, W.; Palanisamy, H.; Becker, M. Estimating ENSO Influence on the Global Mean Sea Level, 1993–2010. *Mar. Geod.* **2012**, *35*, 82–97. [CrossRef]
27. Zuo, J.-C.; He, Q.-Q.; Chen, C.-l.; Chen, M.-X.; Xu, Q. Sea level variability in East China Sea and its response to ENSO. *Water Sci. Eng.* **2012**, *5*, 164–174. [CrossRef]
28. Liu, X.; Liu, Y.; Guo, L.; Rong, Z.; Gu, Y.; Liu, Y. Interannual changes of sea level in the two regions of East China Sea and different responses to ENSO. *Glob. Planet. Chang.* **2010**, *72*, 215–226. [CrossRef]
29. Rong, Z.; Liu, Y.; Zong, H.; Cheng, Y. Interannual sea level variability in the South China Sea and its response to ENSO. *Glob. Planet. Chang.* **2007**, *55*, 257–272. [CrossRef]

30. Gordon, A.L.; Giulivi, C.F. Pacific decadal oscillation and sea level in the Japan/East sea. *Deep. Sea Res. Part I Oceanogr. Res. Pap.* **2004**, *51*, 653–663. [CrossRef]
31. Moon, J.; Song, Y.T.; Lee, H. PDO and ENSO modulations intensified decadal sea level variability in the tropical Pacific. *J. Geophys. Res. Ocean* **2015**, *120*, 8229–8237. [CrossRef]
32. Han, G.; Huang, W. Pacific Decadal Oscillation and Sea Level Variability in the Bohai, Yellow, and East China Seas. *J. Phys. Oceanogr.* **2008**, *38*, 2772–2783. [CrossRef]
33. Zhang, L.; Han, W.; Li, Y.; Lovenduski, N.S. Variability of Sea Level and Upper-Ocean Heat Content in the Indian Ocean: Effects of Subtropical Indian Ocean Dipole and ENSO. *J. Clim.* **2019**, *32*, 7227–7245. [CrossRef]
34. Mohan, S.; Vethamony, P. Interannual and long-term sea level variability in the eastern Indian Ocean and South China Sea. *Clim. Dyn.* **2017**, *50*, 3195–3217. [CrossRef]
35. Duan, J.; Li, Y.; Zhang, L.; Wang, F. Impacts of the Indian Ocean Dipole on Sea Level and Gyre Circulation of the Western Tropical Pacific Ocean. *J. Clim.* **2020**, *33*, 4207–4228. [CrossRef]
36. Wang, S.; Shum, C.; Bevis, M.; He, X.; Zhang, Y.; Ding, Y.; Zhang, C.; Montillet, J.-P. Sea level reconstruction reveals improved separations of regional climate and trend patterns over the last seven decades. *Earth Syst. Sci. Data Discuss.* **2025**, *2025*, 1–36. [CrossRef]
37. Chang, C.W.J.; Hsu, H.H.; Wu, C.R.; Sheu, W.J. Interannual mode of sea level in the South China Sea and the roles of El Niño and El Niño Modoki. *Geophys. Res. Lett.* **2008**, *35*, 82–97. [CrossRef]
38. Hamlington, B.D.; Leben, R.R.; Strassburg, M.W.; Nerem, R.S.; Kim, K. Contribution of the Pacific Decadal Oscillation to global mean sea level trends. *Geophys. Res. Lett.* **2013**, *40*, 5171–5175. [CrossRef]
39. Han, W.; Meehl, G.A.; Stammer, D.; Hu, A.; Hamlington, B.; Kenigson, J.; Palanisamy, H.; Thompson, P. Spatial Patterns of Sea Level Variability Associated with Natural Internal Climate Modes. In *Integrative Study of the Mean Sea Level and Its Components*; Springer: Cham, Switzerland, 2017; pp. 221–254. [CrossRef]
40. Lee, K.; Nam, S.; Cho, Y.-K.; Jeong, K.-Y.; Byun, D.-S. Determination of Long-Term (1993–2019) Sea Level Rise Trends Around the Korean Peninsula Using Ocean Tide-Corrected, Multi-Mission Satellite Altimetry Data. *Front. Mar. Sci.* **2022**, *9*, 810549. [CrossRef]
41. Ashe, E.L.; Khan, N.S.; Toth, L.T.; Dutton, A.; Kopp, R.E. A statistical framework for integrating nonparametric proxy distributions into geological reconstructions of relative sea level. *Adv. Stat. Clim. Meteorol. Oceanogr.* **2022**, *8*, 1–29. [CrossRef]
42. Wang, X.H.; Cho, Y.-K.; Guo, X.; Wu, C.-R.; Zhou, J. The status of coastal oceanography in heavily impacted Yellow and East China Sea: Past trends, progress, and possible futures. *Estuar. Coast. Shelf Sci.* **2015**, *163*, 235–243. [CrossRef]
43. Strassburg, M.W.; Hamlington, B.D.; Leben, R.R.; Manurung, P.; Gaol, J.L.; Nababan, B.; Vignudelli, S.; Kim, K.-Y. Sea level trends in Southeast Asian seas. *Clim. Past* **2015**, *11*, 743–750. [CrossRef]
44. PSMSL. Tide Gauge Data. 2020. Available online: http://www.psmsl.org/data/obtaining/ (accessed on 5 May 2025).
45. Copernicus Marine Data Store. 2025. Available online: https://data.marine.copernicus.eu/products?pk_vid=9227025c302a27b417605815849bbd3f (accessed on 5 May 2025).
46. Copernicus.eu. Global Ocean Gridded L 4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing. 2025. Available online: https://data.marine.copernicus.eu/product/SEALEVEL_GLO_PHY_L4_MY_008_047/services (accessed on 5 May 2025).
47. Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46*, 175–185. [CrossRef]
48. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
49. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
50. MacKay, D.J.C. Bayesian Interpolation. *Neural Comput.* **1992**, *4*, 415–447. [CrossRef]
51. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]
52. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
53. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
54. Gan, M.; Pan, S.; Chen, Y.; Cheng, C.; Pan, H.; Zhu, X. Application of the Machine Learning LightGBM Model to the Prediction of the Water Levels of the Lower Columbia River. *J. Mar. Sci. Eng.* **2021**, *9*, 496. [CrossRef]
55. Pfeffer, J.; Allemand, P. The key role of vertical land motions in coastal sea level variations: A global synthesis of multisatellite altimetry, tide gauge data and GPS measurements. *Earth Planet. Sci. Lett.* **2016**, *439*, 39–47. [CrossRef]
56. Wöppelmann, G.; Marcos, M. Vertical land motion as a key to understanding sea level change and variability. *Rev. Geophys.* **2015**, *54*, 64–92. [CrossRef]
57. Oelsmann, J.; Marcos, M.; Passaro, M.; Sanchez, L.; Dettmering, D.; Dangendorf, S.; Seitz, F. Regional variations in relative sea-level changes influenced by nonlinear vertical land motion. *Nat. Geosci.* **2024**, *17*, 137–144. [CrossRef]

58. Govorcin, M.; Bekaert, D.P.S.; Hamlington, B.D.; Sangha, S.S.; Sweet, W. Variable vertical land motion and its impacts on sea level rise projections. *Sci. Adv.* **2025**, *11*, 5. [CrossRef] [PubMed]
59. He, X.; Huang, J.; Montillet, J.-P.; Wang, S.; Kermarrec, G.; Shum, C.K.; Hu, S.; Wang, F. A Noise Reduction Approach for Improve North American Regional Sea Level Change from Satellite and In Situ Observations. *Surv. Geophys.* **2025**, *19*, 1–32. [CrossRef]
60. Church, J.A.; White, N.J. Sea-Level Rise from the Late 19th to the Early 21st Century. *Surv. Geophys.* **2011**, *32*, 585–602. [CrossRef]
61. Trusenkova, O.O. Long-Term Variation in the Level of the Sea of Japan Based on Satellite Altimetry Measurements. *Izv. Atmos. Ocean. Phys.* **2018**, *54*, 1023–1030. [CrossRef]
62. Lee, E.-J.; Kim, K.; Park, J.-H. Reconstruction of long-term sea-level data gaps of tide gauge records using a neural network operator. *Front. Mar. Sci.* **2022**, *9*, 1–15. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.