remote sensing

MDPI

Article

Comparative Evaluation of Tree-Based Ensemble Learning for Correcting Residuals in SWOT Altimetry-Derived Marine Gravity Anomalies

Li Chen [1,2], Jianbo Wang [1,2,*], Chengcheng Zhu [3] and Dechao An [4]

[1] The School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang 222005, China; 2024220202@jou.edu.cn
[2] The Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510300, China
[3] School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China; zhuchengcheng22@sdjzu.edu.cn
[4] School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China; andch@mail2.sysu.edu.cn
* Correspondence: 2017000015@jou.edu.cn

Highlights

What are the main findings?

- Tree-based ensemble correction improves the SWOT-derived gravity field under different validation strategies, but boosting methods are clearly superior when extrapolation is required.
- The improvement is strongly structured rather than uniform.

What are the implications of the main findings?

- Out-of-fold cross-validation scores can be misleading when selecting models for spatial extrapolation.
- Machine learning refinement should be positioned as a short-to-medium wavelength enhancement.

Abstract

The quality of the SWOT-derived marine gravity field and its subsequent refinement are investigated in the Western Pacific (120–140°E, 0–30°N). Inverting 41 cycles of the SWOT L3_LR_SSH v2.0.1 product yields a gravity field with RMSEs of 2.476 and 4.321 mGal against independent JAMSTEC and NCEI shipborne gravity, respectively, and the accuracy gain converges as the number of SWOT altimetry repeat cycles increases, showing diminishing marginal returns. To further refine this field, three tree-based ensemble learning methods—Random Forest (RF), histogram-based gradient boosting (HGB) and LightGBM (LGBM)—are trained on 67 multi-source features with the SWOT-minus-shipborne residual as the target, and validated under two spatially isolated strategies: a 2° × 2° spatial block hold-out with a 10 km buffer and a cruise-wise hold-out over seven complete cruises, in which training samples within 10 km of any test point are removed. All three methods reduce the error. On the block hold-out, LGBM performs best (RMSE 2.398 mGal, a 3.6% improvement over the 2.488 mGal baseline), followed by HGB (2.402 mGal, 3.5%) and RF (2.422 mGal, 2.7%); on the cruise hold-out the gains are markedly larger, with HGB best (RMSE 1.703 mGal, a 8.6% improvement over the 1.863 mGal baseline), then LGBM (1.714 mGal, 8.0%) and RF (1.794 mGal, 3.7%). A consistent reversal is found between the two assessment levels: RF leads on the in-training out-of-fold RMSE but is weakest on

Check for updates

Academic Editors: Kaoru Ichikawa and Yuanjin Pan

Received: 13 August 2026
Revised: 10 October 2026
Accepted: 22 October 2026
Published: 26 October 2026

Copyright: © 2026 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.

unseen cruises, indicating that RF is competitive within the training distribution yet weaker than boosting methods when extrapolating to unsurveyed areas. Power spectral density analysis shows that the improvements concentrate at wavelengths of 8–45 km, whereas wavenumber below 0.0221 km^{-1} are slightly degraded.

Keywords: SWOT altimetry; marine gravity field inversion; gravity field refinement; tree-based ensemble learning; shipborne gravity measurements; spatial cross-validation

1. Introduction

The marine gravity field is a fundamental component of the Earth's gravity field. Its precise determination is of great significance in geodesy, marine geology and geophysics, and provides essential support for plate tectonic analysis, marine resource exploration and national defence [1–3]. Shipborne and airborne gravimetry offer high accuracy but are costly, time-consuming and limited in spatial coverage, and therefore cannot meet the demands of large-scale applications. Since the 1970s, satellite altimetry, with its global coverage and high spatiotemporal resolution, has become the primary means of deriving marine gravity anomalies [4–6].

Marine gravity recovery from satellite altimetry is well established. Conventional radar altimeters sample SSH along discrete 1-D tracks, from which along-track geoid gradients are used to derive deflection of the vertical (DOV) components and invert gravity anomalies, yielding global models such as DTU17, SIO V32.1, NSOAS22 and SDUST2021GRA [1]. CryoSat-2 improves coastal and high-latitude coverage, while ICESat-2 provides cross-track gradients; in SDUST2022GRA, ICESat-2 contributes 4.3% of the low-to-mid-latitude improvement [1]. Although such data have high along-track resolution, large gaps remain between adjacent tracks; the resulting anisotropy limits both the spatial resolution and the consistency of the recovered gravity field [7,8]. The Surface Water and Ocean Topography (SWOT) mission, launched on 16 December 2022, provides an entirely new observation mode for marine gravity field research [9–11].

The core payload of SWOT is the Ka-band Radar Interferometer (KaRIn), which simultaneously observes two 50 km wide swaths on either side of the nadir track and thus directly acquires high-resolution two-dimensional SSH imagery [12]. This enables, for the first time, a near-synoptic two-dimensional view of SSH without interpolation, and permits the direct computation of sea surface slopes in both the along-track and cross-track directions. By averaging about ten repeat cycles (7.2 months), an accuracy comparable to the best marine gravity maps can be achieved [13]. SWOT data have consequently been rapidly adopted for marine gravity recovery: existing studies have examined multi-cycle stacking, the combination of gradients in different directions, denoising products, and the influence of cross-track systematic errors, water depth and cycle number on the accuracy of the recovered gravity field, in regions including the Philippine Sea, the South China Sea, the Gulf of Mexico and the western Pacific [13–18]. A resolution of about 8 km, and better than 4 km in shallow waters, has been reported [18], considerably surpassing that of conventional altimetry.

Nevertheless, gravity anomalies derived from SWOT still contain spatially heterogeneous errors. Because gravity recovery relies on geoid gradient computations, systematic biases and reference-model errors inherent in the altimetry product propagate directly into the results, and the residuals are typically larger in shallow waters and in areas of complex topography. Shipborne gravity data can effectively compensate for the accuracy deficiencies of satellite altimetry inversion in shallow waters and regions with complex

topography. Machine learning methods, by capitalizing on their ability to exploit the strong nonlinear relationship between satellite altimetry and shipborne gravity, have significantly enhanced inversion accuracy [19–21]. For multiple versions of SWOT data (e.g., SWOT_02, SWOT_03), Xu et al. [22] combined deep learning to perform bathymetry and gravity inversion in the Mariana Trench and the central Pacific seamount region, demonstrating the stability of deep learning in complex regions; Wu et al. [23] constructed a gravity field in the South China Sea by fusing multi-source data based on BPNN, improving the accuracy in shallow waters. Deep learning models are frequently criticised for their "black-box" nature, limited interpretability and weak extrapolation beyond the training distribution.

Tree-based ensemble learning methods provide an interpretable and data-efficient alternative. Random forest (RF) aggregates decorrelated CART trees grown on bootstrap samples (bagging, mainly variance reduction) [24]; it is robust to noise and outliers, requires no assumption about the data distribution, and provides intrinsic measures of feature importance [24]. Boosting methods, by contrast, build trees sequentially, each new tree fitting the negative gradient of the loss function (boosting, mainly bias reduction); histogram-based gradient boosting (HGB) and LightGBM (LGBM) are two widely used implementations [25]. The difference in how these three methods aggregate base learners may affect not only accuracy but also generalisation to locations absent from the training data. To date, however, no systematic comparison of these three methods has been reported for the correction of residuals in SWOT-derived marine gravity anomalies, and it remains unclear which method should be preferred when the corrected field is to be applied beyond the surveyed tracks.

The SWOT Low-Rate Level-3 SSH product v2.0.1 has been released, comprising 21-day science-phase data from Cycle 1 to Cycle 41 (July 2023 to November 2025). Compared with the L2 product, it incorporates several key improvements, including an AI-based (U-Net) denoising algorithm that substantially suppresses instrumental random noise. Such denoising, however, does not remove systematic biases or reference-model errors, which therefore remain in the derived gravity anomalies. In this study, taking the western Pacific (120–140°E, 0–30°N) as the study area, we present a comparative evaluation of three tree-based ensemble learning methods—RF, HGB and LGBM—for correcting the residuals between the SWOT-derived gravity anomalies and independent shipborne gravity measurements. All three methods are trained on an identical set of 67 multi-source geophysical features with the SWOT-minus-shipborne residual as the target, and are assessed under two spatially isolated validation strategies: a 2° × 2°spatial block hold-out with a 10 km buffer, and a cruise-wise hold-out over entire surveys. Particular attention is paid to whether the relative ranking of the three methods is consistent between internal cross-validation and genuinely independent spatial or cruise-wise tests. The remainder of this paper is organized as follows: Section 2 describes the study area and data; Section 3 describes the methods; Section 4 presents the results and analysis; Section 5 presents the discussion; and Section 6 concludes the paper.

2. Study Area and Data

2.1. Study Area

The western Pacific Ocean (120–140°E, 0–30°N) is selected as the study area. This region is situated between two island arcs, where the complex topography and geomorphology pose significant challenges for both shipborne bathymetry and satellite altimetry inversion of seafloor topography [6]. Figure 1 displays the seafloor topography and the distribution of shipborne gravity tracks within the study area; the red lines denote NCEI data, and the green lines denote JAMSTEC data. The shipborne gravity data were obtained from the National Centers for Environmental Information (NCEI; https://maps.ngdc.noaa.

gov/viewers/geophysics/, accessed on 5 May 2025) and the Japan Agency for Marine-Earth Science and Technology (JAMSTEC; https://www.godac.jamstec.go.jp/, accessed on 5 May 2025), while the seafloor topography data are based on the ETOPO 2022 model [26].

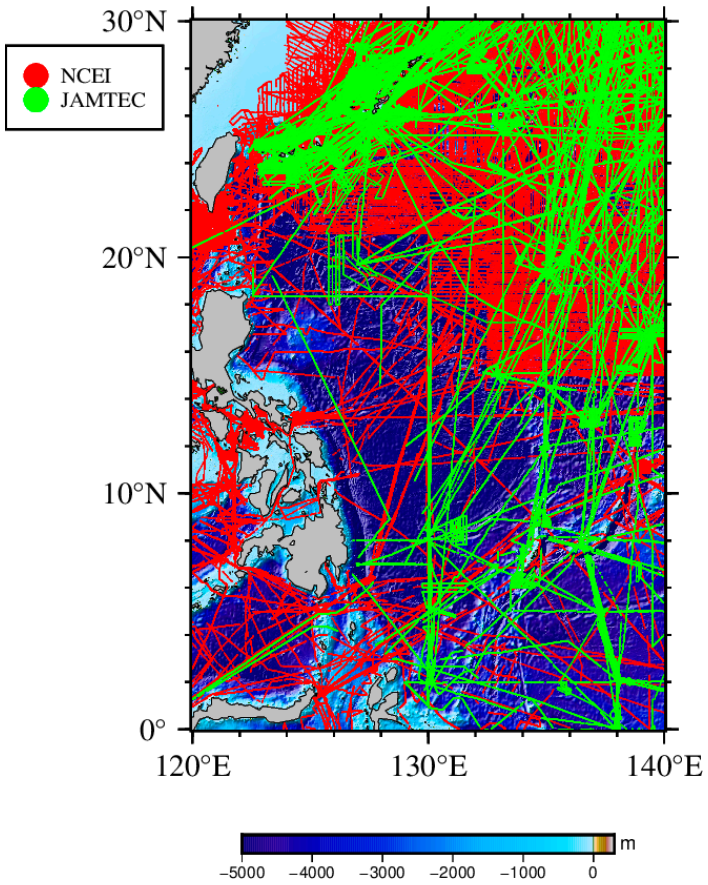


Figure 1. Study area sketch map. The shipborne gravity tracks are obtained from NCEI and JAMSTEC. The seafloor topography data are based on the ETOPO 2022 model.

2.2. SWOT Data

This study utilizes Version 2.0.1 of the SWOT Low-Rate Level-3 SSH product (L3_LR_SSH v2.0.1). The dataset, spanning Cycles 1 to 41 (July 2023 to November 2025), is produced and distributed by the AVISO+ (Archiving, Validation, and Interpretation of Satellite Oceanographic Data) and DUACS (Data Unification and Altimeter Combination System) teams at CNES [9]. The data are accessible via: https://www.aviso.altimetry.fr/en/missions/current-missions/swot, accessed on 1 May 2026. The grav_SWOT_04 model (https://topex.ucsd.edu/pub/global_grav_1min_SWOT/, accessed on 1 May 2026) is used to evaluate the gravity anomalies derived from the SWOT L3_LR_SSH product. Within the study area, the SWOT L3 Version 2.0.1 product contains 41 cycles and a total of 1634 tracks. Both the along-track and cross-track resolutions are 2 km. Standard geophysical corrections have been applied in the L3 Expert product, while the parameters dac, ocean_tide, internal_tide, mss, and mdt are provided separately for verification.

2.3. Mean Dynamic Topography Model

The Mean Dynamic Topography model MDT_CNES_CLS22_global was jointly developed by CNES and CLS, available at https://www.aviso.altimetry.fr/en/data/products/auxiliary-products/mdt/mdt-global-cnes-cls.html, accessed on 1 May 2026. It estimates the 1993–2012 mean SSH relative to the geoid on a global 0.125° × 0.125° grid, covering open oceans except the Black Sea. The key upgrade targets the Arctic: improved MSS

coverage eliminates artifacts seen in CNES-CLS18 and yields a more physically consistent circulation representation, e.g., the Norwegian Atlantic Front Current along the Mohn Ridge. Globally, CNES-CLS22 stays close to CNES-CLS18, with only marginal gains—RMS differences in geostrophic velocities drop by ~0.2–0.5% and variance reduction vs. profile-derived heights is ~0.5% [27].

2.4. Vertical Deflection and Earth Gravity Field Models

The reference gravity anomaly and DOV datasets used in this study are derived from the SIO V32.1 global model, released by the Scripps Institution of Oceanography (SIO) in August 2022. These datasets are distributed in a global $1' \times 1'$ grid format (https://topex.ucsd.edu/pub/global_grav_1min/, accessed on 1 May 2026).

Built upon multi-altimeter data (including Jason-1/2, Envisat, SARAL/AltiKa, and CryoSat-2), the SIO V32.1 model achieves a global accuracy of 1–2 mGal [28]. Key advancements in this version include updated waveform retracking and a reduced smoothing filter wavelength for Significant Wave Height (SWH) from 45 km to 30 km. Furthermore, optimized sea-state bias corrections were applied across all altimeters to mitigate SWH noise within the 30 km band. In this research, the SIO V32.1 gravity anomaly (SIOv32.1 GRA) and DOV (SIOv32.1 DOV) grids serve as the baseline for marine gravity field inversion in the study area.

2.5. Shipborne Gravity Data

Shipborne gravity data, acquired through direct measurement, offer higher accuracy compared to satellite-derived gravity anomalies and thus serve as an independent benchmark for accuracy assessment. The data utilized in this study were obtained from NCEI and JAMSTEC, with their spatial distribution in the study area illustrated in Figure 1.

Due to the extensive temporal span of the NCEI dataset (from the 1960s to the early 2000s), significant discrepancies exist among different cruises in terms of positioning accuracy, measurement conditions, and reference datums. To mitigate long-wavelength errors, we applied a quadratic polynomial fitting method [29] and referenced the DTU17 gravity anomaly (DTU17GRA) model [30] for datum unification. The DTU17GRA model was selected because it does not contain shipborne gravity data and can therefore serve as an independent quantity for adjusting shipborne gravity measurements. The DTU17GRA model was obtained from the Technical University of Denmark (DTU Space) via https://ftp.space.dtu.dk/pub, accessed on 1 May 2026. The specific procedure is as follows.

First, the DTU17 grid was interpolated to the shipborne observation locations using a separable Newton forward-difference interpolation scheme to obtain the residuals $\left(l_i = g_i^{model} - g_i^{ship}\right)$. Shipborne points with $\left(|g_i^{ship}| > 500\right)$ mGal or $(|l_i| > 80)$ mGal were discarded. A robust MAD-based outlier rejection was then performed, in which $(\sigma = 1.4826 \times \mathrm{MAD})$ with a lower bound of 2 mGal, and points satisfying $(|l_i - median(l)| > 3\sigma)$ were removed.

Second, for each cruise, a temporal drift adjustment was carried out using the model [1]

$$r_i = d_0 + d_1 t_i + d_2 t_i^2 \tag{1}$$

where r_i is the estimated value of l_i, and t_i is the centered time relative to the cruise start (in days). The shipborne gravity was corrected as

$$g_i{}' = g_i + r_i \tag{2}$$

Here, VRMS denotes the root-mean-square of the post-fit residuals and is defined as

$$VRMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(g_i' - g_i^{model}\right)^2} \quad (3)$$

where N is the number of retained points in the cruise, g_i' is the corrected shipborne gravity, and g_i^{model} is the DTU17 model value at the same location.

The order of the polynomial was adaptively determined: a constant model was used by default; a linear term was added if it reduced the VRMS by more than 5%; and a quadratic term was further included only if its endpoint contribution, $|d_2|(T/2)^2$, exceeded 1 mGal, where $(T = t_{max} - t_{min})$, and the quadratic model further reduced the VRMS by more than 5%. Cruises with unrealistic coefficients, i.e., $|d_0| > 100$ mGal, $|d_1| > 50$ mGal/day, or $\left(|d_2|T^2 > 100 \text{ mGal}\right)$, were rejected. Finally, the corrected shipborne gravity and the post-fit residuals were output for subsequent gridding.

By contrast, the high-precision JAMSTEC data were acquired entirely after 2000. Nevertheless, to eliminate systematic biases between cruises, the same processing protocol was applied to the JAMSTEC dataset. During processing, the extracted DTU17 reference gravity field was extended 2° outward in all four cardinal directions. Within the study area, the adjusted shipborne gravity dataset comprises 158 JAMSTEC cruises and 197 NCEI cruises.

The accuracies of the gravity anomalies at the crossover points for the JAMSTEC and NCEI shipborne gravity data, before and after adjustment, are summarized in Table 1. As shown, the RMSE of the crossover discrepancies decreases for both datasets after adjustment; in particular, the crossover discrepancy RMSE of the adjusted JAMSTEC shipborne gravity is 2.941 mGal. This crossover discrepancy RMSE characterizes the internal consistency of the shipborne gravity data.

Table 1. Statistics of crossover discrepancy accuracies for the JAMSTEC and NCEI shipborne gravity data before and after adjustment (mGal).

Agency	Crossover Adjustment	MAX	MIN	MEAN	STD	RMSE
JAMSTEC	Before adjustment	49.685	−70.994	−0.620	5.333	5.369
	After adjustment	28.654	−34.570	0.107	2.939	2.941
NCEI	Before adjustment	99.059	−85.349	0.319	10.889	10.893
	After adjustment	41.909	−47.833	−0.060	6.064	6.064

The RMSE of crossover discrepancies for the NCEI shipborne gravity data dropped from 10.893 mGal before adjustment to 6.064 mGal after adjustment. This dataset served as an independent test set to verify the improvement in accuracy achieved by different methods for SWOT gravity field residual learning.

3. Marine Gravity Anomaly Inversion and Residual Correction Based on Tree-Based Ensemble Learning

Figure 2 illustrates the workflow for marine gravity field recovery from SWOT altimetry and residual correction based on tree-based ensemble learning methods. The methodology comprises three core modules: (1) computation of the residual DOV; (2) marine gravity anomaly inversion via the inverse Vening Meinesz (IVM) formula; and (3) gravity field refinement through residual correction using tree-based ensemble learning methods.

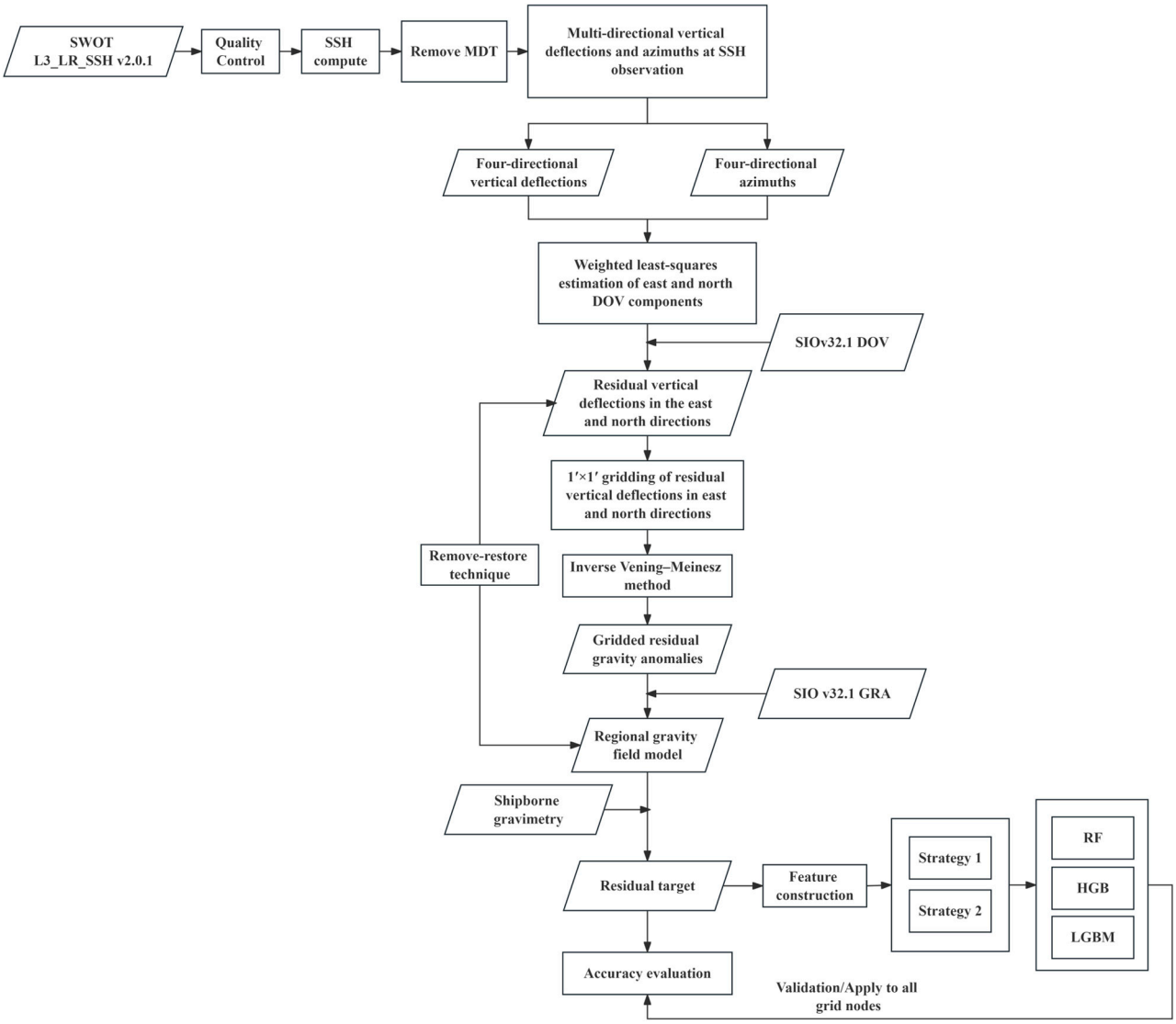


Figure 2. The workflow for marine gravity field recovery from SWOT altimetry and residual learning based on tree-based ensemble learning methods.

3.1. Computation of the Residual Deflection of the Vertical

The preprocessing of altimetry data for marine gravity anomaly inversion includes systematic error correction, removal of multi-source errors, and the extraction of stable along-track sea surface slopes [31,32]. For SWOT two-dimensional altimetry data, the SSH and auxiliary variables are first read from the NetCDF file, including the observation point coordinates (λ,ϕ), the Mean Sea Surface (MSS), the Sea Surface Height Anomaly (SSHA), the ocean_tide correction, and the data quality flag(quality_flag). In this study, the unfiltered SSHA field (ssha_unfiltered) is used for the SSHA. The instantaneous SSH is calculated as:

$$SSH(\lambda,\phi) = MSS(\lambda,\phi) + SSHA(\lambda,\phi) \quad (4)$$

SSH observations are affected by geophysical signals and systematic measurement errors, complicating the separation of signals from errors [33]. In this study, only observations with quality_flag = 0 and $|\text{ocean_tide}| \leq 5$ m were retained.

After completing the SSH quality control, the Mean Dynamic Topography (MDT) must be subtracted to obtain the instantaneous geoid at the sea surface locations covered by the satellite tracks. In this study, this is achieved by removing the MDT_CNES_CLS22_global

model. Since the MDT model contains missing values at a few grid points, the nearest-neighbor interpolation method is adopted to fill these voids. After void filling, the MDT is interpolated to the observation points, and the instantaneous geoid can then be calculated as:

$$SSH(\lambda,\phi) = MSS(\lambda,\phi) + SSHA(\lambda,\phi) \tag{5}$$

where SSH(λ,ϕ) contains the combined effects of both dynamic and gravitational components, while MDT(λ,ϕ)represents the mean dynamic component. Subtracting the latter from the former yields N(λ,ϕ), which reflects the sea surface morphology governed by the gravity field, i.e., the geoid.

The geoid gradient reflects the rate of change in the equipotential surface elevation at different locations, that is, the degree of tilt in various horizontal directions [34]. To comprehensively characterize the local morphology of the geoid and improve the inversion accuracy of residual vertical deflections and gravity anomalies, gradients are computed in four directions in this study. Specifically, using the central difference method [35], the geoid gradients and their corresponding azimuths are calculated from the geoid at the sea surface locations covered by the tracks for the along-track, cross-track, main diagonal, and secondary diagonal directions, as illustrated in Figure 3. The grid spacing is taken as isotropic (2 km × 2 km) in the SWOT swath grid. The geoid gradients in these four directions, denoted a_s, ε_a, ε_{c1}, ε_{c2} and ε_{c3}, can be expressed as:

$$\begin{aligned}
\varepsilon_a &= \frac{\partial N}{\partial s_a}\Big|_{(i,j)} = \frac{N_{i+1,j} - N_{i-1,j}}{2d} \\
\varepsilon_{c1} &= \frac{\partial N}{\partial s_{c1}}\Big|_{(i,j)} = \frac{N_{i,j+1} - N_{i,j-1}}{2d} \\
\varepsilon_{c2} &= \frac{\partial N}{\partial s_{c2}}\Big|_{(i,j)} = \frac{N_{i+1,j+1} - N_{i-1,j-1}}{2\sqrt{2}d} \\
\varepsilon_{c3} &= \frac{\partial N}{\partial s_{c3}}\Big|_{(i,j)} = \frac{N_{i-1,j+1} - N_{i+1,j-1}}{2\sqrt{2}d}
\end{aligned} \tag{6}$$

where i and j denote the pixel indices in the along-track and cross-track directions, respectively; the subscripts a, c1, c2, and c3 correspond to the along-track, cross1-track, cross2-track, and cross3-track directions, respectively; N is the geoid height; e is the unit vector; and d = 2 km [35] is the grid spacing of the SWOT L3_LR_SSH product.

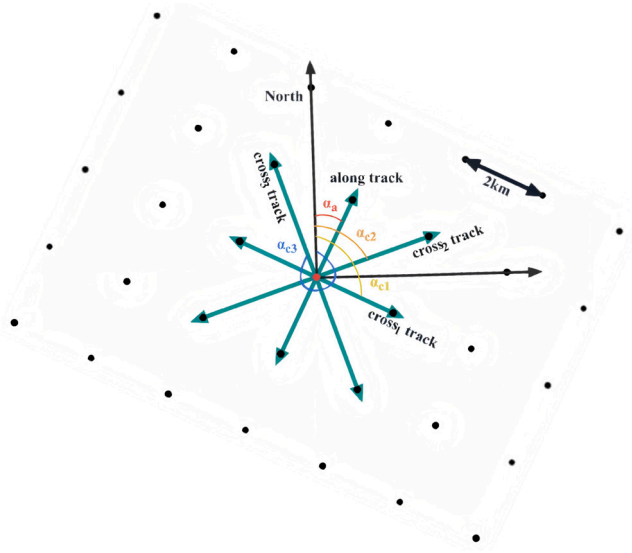


Figure 3. Schematic diagram of satellite track points and the central difference method.

For each differencing pair, the azimuth of the corresponding baseline is computed from the coordinates of its two end nodes. Given the starting point (λ_1,ϕ_1) and the ending

point (λ_2,ϕ_2) of the observation, the great-circle spherical azimuth α (measured clockwise from north) directed from the starting point to the ending point is:

$$\alpha = arctan^{2(sin(\lambda_2-\lambda_1)\cdot cos\phi_2, cos\phi_1\cdot sin\phi_2 - sin\phi_1\cdot cos\phi_2\cdot cos(\lambda_2-\lambda_1))} \quad (7)$$

Let η and ξ denote the east and north components of the DOV, respectively, which are opposite in sign to the geoid gradients. The formula for calculating the geoid gradient ε at a given azimuth α from ξ and η is:

$$\begin{bmatrix} \varepsilon_a \\ \varepsilon_{c1} \\ \varepsilon_{c2} \\ \varepsilon_{c3} \end{bmatrix} = - \begin{bmatrix} cos\alpha_a & sin\alpha_a \\ cos\alpha_{c1} & sin\alpha_{c1} \\ cos\alpha_{c2} & sin\alpha_{c2} \\ cos\alpha_{c3} & sin\alpha_{c3} \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} \quad (8)$$

where α_a, α_{c1}, α_{c2}, and α_{c3} denote the azimuths of the along-track, cross$_1$-track, cross$_2$-track (main diagonal), and cross$_3$-track (secondary diagonal) directions, respectively. According to Equation (8), a system of linear equations can be established as:

$$Ax = b \quad (9)$$

where $A = \begin{bmatrix} cos\alpha_a & sin\alpha_a \\ cos\alpha_{c1} & sin\alpha_{c1} \\ cos\alpha_{c2} & sin\alpha_{c2} \\ cos\alpha_{c3} & sin\alpha_{c3} \end{bmatrix}$ and $b = - \begin{bmatrix} \varepsilon_a \\ \varepsilon_{c1} \\ \varepsilon_{c2} \\ \varepsilon_{c3} \end{bmatrix}$. The deflection components are obtained by weighted least squares:

$$\hat{x} = (A^T PA)^{-1} A^T Pb \; with \; P = diag(1,1,2,2) \quad (10)$$

Because the four directional observations are exactly consistent (the two diagonal estimates are linear combinations of the other two), the solution is independent of the particular choice of P. In this study, bilinear interpolation is employed to map the east and north components of the vertical deflection from the SIOv32.1 DOV model to the locations corresponding to the vertical deflection derived from the SWOT data, thereby obtaining the model values for the east and north components at these positions. Finally, the residual vertical deflection is obtained by subtracting the SIOv32.1 DOV model values from the vertical deflection derived from the SWOT data. For different cycles and tracks, the east west residual vertical deflection and the north south residual vertical deflection are gridded separately.

3.2. Inversion of Marine Gravity Anomalies Using the Inverse Vening Meinesz Method

After obtaining the vertical deflection field in a regular grid format, this study adopts the IVM method to invert the marine gravity anomalies. This method originates from the Vening Meinesz formula, the fundamental concept of which is to exploit the integral relationship between vertical deflection and gravity anomalies, thereby enabling the transformation from geometric quantities to physical quantities [35,36]. Under the spherical approximation, an integral relationship exists between the vertical deflection components and the gravity anomaly Δg. The IVM method recovers the corresponding gravity anomaly distribution by performing spatial integration over the vertical deflection field, and its basic formulation can be expressed as [35,36]:

$$\Delta g(p) = \frac{\gamma_0}{4\pi} \iint_{\sigma} H'(\xi_q cos\alpha_{qp} + \eta_q sin\alpha_{qp}) d\sigma_q \quad (11)$$

where p is the computation point, q is the moving point, $\Delta g(p)$ is the gravity anomaly at the computation point, γ_0 is the normal gravity, α_{qp} is the azimuth from q to p, ξ_q and η_q are the north and east components of the vertical deflection at point q, respectively, and the kernel function H' is defined as:

$$H' = -\frac{cos\frac{\psi_{pq}}{2}}{2sin^2\frac{\psi_{pq}}{2}} + \frac{cos\frac{\psi_{pq}}{2}(3 + 2sin\frac{\psi_{pq}}{2})}{2sin\frac{\psi_{pq}}{2}(1 + sin\frac{\psi_{pq}}{2})} \tag{12}$$

where ψ_{pq} denotes the spherical angular distance between points p and q. The IVM method follows the 1-D FFT algorithm of Hwang [36], with a truncation radius of 110 km. Integration is performed by direct summation in latitude and FFT convolution in longitude. The kernel is set to zero beyond the truncation radius, and 100% zero padding is applied in longitude to avoid cyclic convolution. For an odd number of longitudinal grid points, the last column is obtained by linear extrapolation. The input north and east components are checked for grid consistency before inversion. The innermost zone is treated separately using a local 13 × 13 window (span = 6): the gradients are quadratically interpolated, their local derivatives at the central node are evaluated, and the resulting effect is added to the 1-D FFT result; at grid boundaries the window is truncated to the available data.

3.3. Refinement of the Marine Gravity Field Using Tree-Based Ensemble Learning Methods

Let the marine gravity anomaly field derived from satellite altimetry be denoted by $G_{ref}(x)$, defined on a regular grid with $x = (\lambda, \varphi)$. Because this field still contains residual errors with respect to the shipborne observations, a correction field is estimated using the shipborne data and subsequently added back to the reference field. Following the remove—restore strategy, the correction rather than the gravity field itself is modelled, so that the long-wavelength information carried by the reference model is preserved and only the residual part has to be learned. Three tree-based ensemble learning methods are employed for this purpose. The procedure consists of six steps, described below.

(1) Definition of the learning target

For a set of N shipborne control points, let ${g_i}^{ship}$ denote the shipborne gravity anomaly and ${g_i}^{ref} = G_{ref}(x_i)$ the reference model value sampled at the same location. The learning target is defined as the difference between them:

$$y_i = {g_i}^{ref} - {g_i}^{ship}, \quad i = 1, \ldots, N. \tag{13}$$

The objective is therefore to determine a regression function f that maps a feature vector x_i, which characterises the local geophysical environment of the observation, to the correction y_i. Once f has been estimated, the refined gravity field is obtained by adding the predicted correction to the reference field.

(2) Feature construction

All input grids are first resampled onto a common master grid by bilinear interpolation, so that every grid node is associated with one value of each input quantity. For each observation the feature vector is then obtained by sampling the corresponding grids:

$$x_i = [F_1(\varphi_i, \lambda_i), \ldots, F_P(\varphi_i, \lambda_i)]^T \tag{14}$$

where P is the number of features. The features are built from the available gravity-field models, the DOV and vertical gravity gradient (VGG) models, and the bathymetry model, together with their spatial derivatives and their multi-scale transforms. They fall into the following groups: (i) the gravity-field quantities themselves and the differences between the individual models; (ii) the north and east components of the DOV, and the VGG;

(iii) the bathymetry and its derivatives, including the slope, the curvature, the roughness and the distance to the coast; (iv) the gravity effect of the seafloor relief computed from the bathymetry using the Parker formula; (v) the multi-scale representations obtained by Gaussian smoothing; (vi) the band-pass components of the residual fields within prescribed wavelength bands; (vii) the local texture of the residual fields; and (viii) the horizontal gradient magnitude and the Laplacian of the gravity fields.

The window statistics used in groups (iii) and (vii) are the local mean and the local standard deviation within a $w \times w$ window:

$$\overline{F}_w = \frac{1}{w^2} \sum_{i,j \in W} F(i,j) \tag{15}$$

$$\sigma_w = \sqrt{\frac{1}{w^2} \sum_{i,j \in W} F^2(i,j) - \overline{F}_w^2} \tag{16}$$

and the horizontal derivatives are approximated by first-order central differences:

$$\left. \frac{\partial F}{\partial x} \right|_{i,j} \approx \frac{F_{i+1,j} - F_{i-1,j}}{2\Delta x}, \tag{17}$$

$$\left. \frac{\partial F}{\partial y} \right|_{i,j} \approx \frac{F_{i,j+1} - F_{i,j-1}}{2\Delta y}, \tag{18}$$

$$\nabla^2 F = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}. \tag{19}$$

Consensus statistics of the gravity-field model ensemble are computed to summarize local model agreement and indicate the reliability of the reference field at each grid node.

(3) Sample construction

Each shipborne observation is geographically collocated with the master grid, and its feature vector (Equation (14)) is extracted from the gridded features via bilinear interpolation. Observations with a missing target or with any missing feature are discarded. To avoid information leakage between the training samples and the independent test samples, every training sample located within a prescribed radius r_{buf} of any test sample is removed prior to training. Only the training set is reduced by this operation, so that the test set remains intact while the training and test samples are spatially separated at the scale of r_{buf}.

(4) Spatial block cross-validation

Because marine gravity observations are strongly spatially autocorrelated, a random partition of the samples would yield optimistically biased error estimates. The study area is therefore divided into spatial blocks of prescribed size s:

$$i_{blk} = \lfloor \varphi / s \rfloor, \quad j_{blk} = \lfloor \lambda / s \rfloor, \tag{20}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Each block is assigned to one of K folds. To balance the sample size among the folds, the blocks are sorted in descending order of their sample count and assigned greedily to the fold that is currently the least loaded. When one fold serves for validation, all training samples whose distance to the centre of any validation block is smaller than the sum of the block circumradius and a prescribed buffer are excluded, so that a spatial buffer separates the training and the validation samples. The out-of-fold (OOF) prediction of a sample, obtained from the model trained without its own fold, is

$$\hat{g}_i^{OOF} = f^{(-k(i))}(x_i), \tag{21}$$

where $k(i)$ denotes the fold that contains the i-th sample. The out-of-fold predictions give an approximately unbiased estimate of the generalisation error and are used to assess the candidate models.

(5) Tree-based ensemble regression models

The correction is estimated by tree-based ensemble learning methods, in which a number of regression trees are combined into a single predictor. Three such methods are employed and are trained independently on the same features and the same samples; they differ only in the way in which the individual trees are combined.

Random Forest: the trees are grown independently on bootstrap resamples of the training data and the prediction is their average:

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x), \tag{22}$$

where T_b denotes the b-th regression tree and B the number of trees. Averaging over decorrelated trees reduces the variance of the individual trees while leaving their bias essentially unchanged.

Gradient-boosted trees (LightGBM and HGB): the trees are built sequentially, each new tree fitting the negative gradient of the loss function of the current model, so that the model is updated additively as

$$F_m(x) = F_{m-1}(x) + \nu\, h_m(x), \tag{23}$$

where ν is the learning rate, which controls the contribution of each tree, and h_m the m-th tree. In contrast to bagging, boosting reduces the bias by fitting the residuals sequentially.

(6) Full-grid correction

The trained model is applied to every node of the master grid, and the refined gravity anomaly is obtained by subtracting the predicted correction to the reference field:

$$G_{corr}(\varphi, \lambda) = G_{ref}(\varphi, \lambda) - \hat{y}(\varphi, \lambda). \tag{24}$$

Prediction is restricted to ocean nodes, identified by a positive water depth. Nodes at which a feature is missing or the prediction is not finite are filled with the reference field, so that the refined field remains consistent with the reference model outside the valid prediction domain.

4. Results and Analysis

4.1. Accuracy of SWOT Inversion Results with Different Cycle Numbers

To investigate the impact of temporal coverage on inversion accuracy, eight experiments were designed using cumulative SWOT cycles (Groups 1–8: Cycles 1–5, 1–10, 1–15, 1–20, 1–25, 1–30, 1–35, and 1–41, respectively). The calculated residual DOV components were gridded at a $1' \times 1'$ resolution using the blockmean and surface modules in GMT [35].

Accuracy assessment for gravity anomaly was conducted within research region (Figure 1). The accuracy of the final SWOT-derived gravity anomalies (relative to shipborne gravity) across the study area is presented in Table 2.

The accuracy (RMSE) of the inverted gravity anomalies in Table 2 improves progressively with increasing SWOT cycle counts, regardless of whether assessed against NCEI or JAMSTEC shipborne gravity data. Overall, the rate of accuracy improvement follows a "rapid initial gain followed by gradual stabilization" pattern, characterized by significant early enhancements and subsequent plateauing.

Table 2. Accuracy of gravity anomalies for different numbers of cycles (mGal). Note: 569,360 points were used for NCEI, and 5,260,481points were used for JAMSTEC.

Group & Number of Cycles	Institute	MAX	MIN	MEAN	STD	RMSE
Group 1: 5 cycles	JAMSTEC	32.569	−34.744	0.029	2.831	2.831
	NCEI	67.839	−44.444	0.101	4.578	4.579
Group 2: 10 cycles	JAMSTEC	29.967	−34.703	−0.011	2.617	2.617
	NCEI	67.535	−46.434	0.058	4.418	4.419
Group 3: 15 cycles	JAMSTEC	30.057	−34.810	−0.028	2.551	2.552
	NCEI	66.378	−44.760	0.075	4.370	4.370
Group 4: 20 cycles	JAMSTEC	27.792	−34.738	−0.028	2.510	2.510
	NCEI	66.275	−44.525	0.082	4.346	4.346
Group 5: 25 cycles	JAMSTEC	27.947	−34.925	−0.027	2.501	2.501
	NCEI	65.571	−45.478	0.073	4.335	4.336
Group 6: 30 cycles	JAMSTEC	28.254	−34.811	−0.032	2.494	2.494
	NCEI	65.196	−45.461	0.073	4.328	4.328
Group 7: 35 cycles	JAMSTEC	28.325	−34.827	−0.029	2.483	2.483
	NCEI	65.407	−45.383	0.072	4.325	4.326
Group 8: 41 cycles	JAMSTEC	28.278	−34.691	−0.021	2.476	2.476
	NCEI	66.846	−45.714	0.072	4.321	4.321

In detail, the NCEI-derived RMSE values are systematically higher than those from JAMSTEC. This discrepancy primarily stems from the broader temporal span of NCEI data acquisition, which introduces larger inter-cruise systematic discrepancies compared to the more homogeneous JAMSTEC dataset. For the models inverted from the data of each group, the accuracies evaluated using both JAMSTEC and NCEI shipborne gravity data show monotonic improvement. However, the magnitude of this improvement decreases progressively as the number of altimetry cycles in each group increases. A clear saturation effect is also evident: the accuracy gain from Group 7 to Group 8 is marginal, measuring less than 0.01 mGal.

Given the superior accuracy and consistency of the JAMSTEC dataset, the subsequent residual gravity correction experiments will utilize the Group 8 inversion results—which yielded the highest JAMSTEC-assessed accuracy—as the baseline. We also evaluated the RMSE of the difference between the Group 8 inversion results and the grav_SWOT_04 model, which is 2.065 mGal. Figure 4 compares the spatial distributions of the gridded residual gravity anomalies for Groups 1, 5, and 8. These residual gravity anomalies were computed, using the remove–restore method, from the corresponding residual deflections of the vertical for Groups 1, 5, and 8.

4.2. Refinement of the SWOT-Derived Marine Gravity Field Using Tree-Based Ensemble Learning Methods

Since the JAMSTEC shipborne gravity data are predominantly collected after 2000, systematic discrepancies between different survey lines are relatively minor. In this study, these data are employed to derive residuals relative to the SWOT-inverted gravity field, which are then used to perform corrections on the SWOT gravity model. Two complementary test-set selection strategies are adopted in this study.

Strategy One: spatial block hold-out with a buffer zone. To mitigate the influence of spatial autocorrelation on accuracy assessment, the study area is partitioned into $2^\circ \times 2^\circ$ spatial grid blocks. Using the "block" as the minimal allocation unit, the blocks are sorted in descending order of sample count and assigned to five groups via greedy bin packing so that the groups are balanced in size; each group is paired with a corresponding

training set formed by merging the remaining four groups; in this study the fifth group is used as the test set. All samples from the other four groups that lie within 10 km of any test point are removed, establishing a 10 km buffer between the training and test samples to suppress information leakage arising from spatial adjacency. To improve computational efficiency, the JAMSTEC shipborne gravity data within the study area (120–140°E, 0–30°N) are first thinned at a regular 10:1 interval, yielding 526,048 points, upon which the partitioning described above is performed. The resulting partition is shown in Figure 5: the test set contains 108,954 points (20.7%), and the other four groups contain 417,094 points in total; after removing 16,853 points by the buffer, the training set comprises 400,241 points (76.1%).

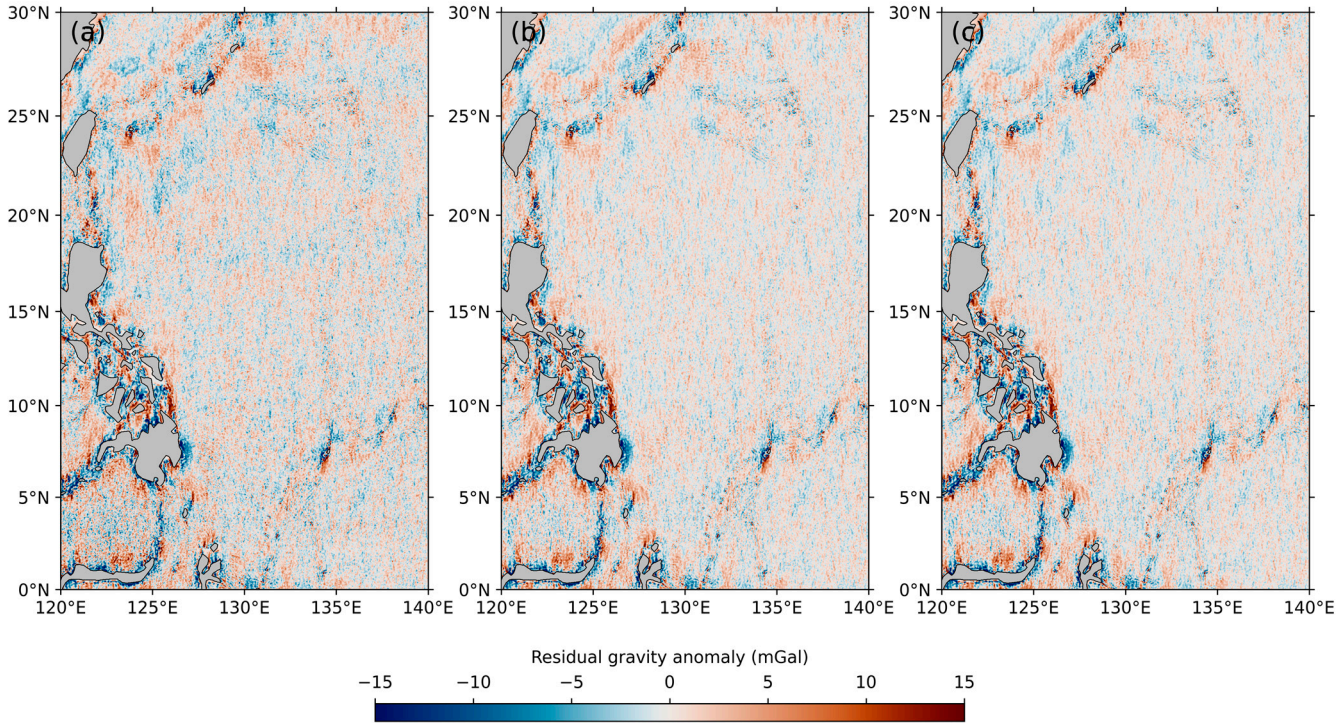


Figure 4. Spatial distributions of the SWOT residual gravity anomalies for different accumulated cycles: (**a**) Group 1 (Cycles 1–5); (**b**) Group 5 (Cycles 1–25); (**c**) Group 8 (Cycles 1–41).

Strategy Two: cruise-wise hold-out. To evaluate the accuracy when the model is extrapolated to cruises that are entirely unseen during training, a cruise-wise hold-out scheme is adopted. Because measurement points within the same cruise share the vessel, instrument, cruise-specific bias and adjustment corrections, and therefore do not constitute independent samples, the "cruise" rather than the "point" is taken as the partitioning unit. All shipborne gravity data are first aggregated into 158 cruise units. The study area is then divided into seven sub-regions according to track density and geographic location, and within each sub-region one representative cruise with relatively high spatial isolation is selected as the test set. This yields 7 test cruises containing 423,986 points (8.1%), while the remaining 151 cruises with 4,836,495 points constitute the training set. No points are discarded during the partitioning stage, so the test set remains complete. To improve training efficiency, the training set is thinned at a regular 10:1 interval to 483,650 points; subsequently, during the training stage, a 10 km buffer is imposed and 106,120 training samples (21.9%) located within 10 km of any test point are removed. The number of points actually used for training is 377,530, whereas the test set retains all 423,986 points. The resulting partition is shown in Figure 6.

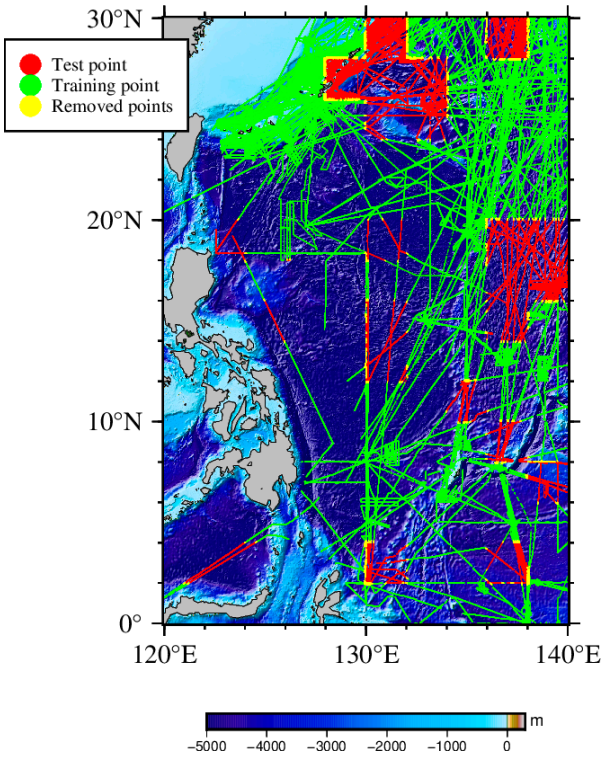


Figure 5. Spatial-block hold-out partition under Strategy 1. The study area is divided into $2^\circ \times 2^\circ$ blocks assigned to five groups; the fifth group forms the test set and the remaining four form the training set, with a 10 km spatial buffer radius enforced between the training and test samples.

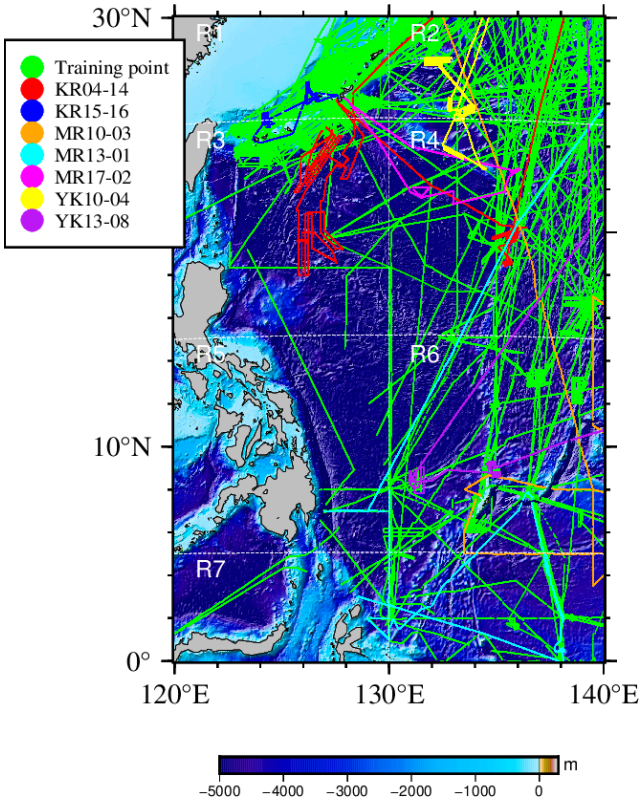


Figure 6. Cruise-wise hold-out partition under Strategy 2. Training points are plotted in green, and the remaining colors represent the test points of the seven held-out cruises, each selected from a different sub-region. The test set is kept complete, while training samples within 10 km of any test point are removed during training.

A total of 67 features were engineered for Tree-based ensemble learning methods, categorized into nine groups, shown in Table 3. The hyperparameter settings of the three methods are given in Table 4. The main grid uses the $1' \times 1'$ SIO V32.1 grid; XGM2019e [37], and SDUST2022GRA [1] are both publicly available and widely used reference models, which facilitates cross-comparison. More importantly, conventional satellite altimetry and SWOT are complementary in wavelength coverage—this is precisely the motivation for constructing a multi-model feature set in this study and using inter-model discrepancies as predictors. Using only the latest single product would instead lose this information.

Table 3. Feature groups and the feature list.

Group	Feature Family	Number of Features	Description
1	Gravity-field model quantities and their differences	16	XGM; SDUST; SWOT-derived gravity anomaly; SWOT-derived residual gravity anomaly field relative to SIO; SIO–XGM; SDUST–XGM; SWOT–XGM; SWOT–SDUST; std(SIO, XGM, SWOT); Mean of the four members (SIO, XGM, SDUST, SWOT); std(SIO, XGM, SDUST, SWOT); Range of the four members (max–min); Deviation of SIO from the four-member median; Deviation of XGM from the median; Deviation of SDUST from the median; Deviation of SWOT from the median
2	DOV and VGG	7	North component of the SIO DOV; East component of the SIO DOV; Curl of the DOV field; North component of the SWOT DOV residual; East component of the SWOT DOV residual; VGG (SIO); Difference between VGG(SIO) and the VGG converted from the DOV
3	Basic seafloor topography	7	Water depth; Slope magnitude; East–west slope; North–south slope; North–south curvature; Local standard deviation of water depth in a 9×9 window; Distance to coast
4	Parker topographic correction	3	Parker topographic gravity correction; High-pass component of the Parker correction (after subtracting the 0.5° smoothed field); Parker correction smoothed at 1.0° (Gaussian)
5	Topographic multi-scale components and high-pass terms	5	Roughness of water depth in a 5×5 window; Roughness of water depth in a 15×15 window; Roughness of water depth in a 31×31 window; High-pass water depth (after subtracting the 0.5° smoothed field); Gradient magnitude of the slope field
6	Multi-scale smoothing	9	Water depth smoothed at 1.25° (Gaussian); Slope smoothed at 1.25°; SIO–XGM residual smoothed at 0.5°; SIO–XGM residual smoothed at 1.25°; SIO–XGM residual smoothed at 2.0°; SIO–XGM residual smoothed at 3.0°; SWOT-derived residual relative to SIO smoothed at 0.5°; SWOT-derived residual relative to SIO smoothed at 1.25°; Parker correction smoothed at 1.25°
7	Band-pass decomposition	8	SIO–XGM residual, 4–12 km band-pass; SIO–XGM residual, 12–30 km; SDUST–XGM residual, 4–12 km band-pass; SDUST–XGM residual, 12–30 km; SDUST–XGM residual, 30–80 km; SWOT-derived residual relative to SIO, 4–12 km band-pass; SWOT-derived residual relative to SIO, 12–30 km; SWOT-derived residual relative to SIO, 30–80 km
8	Local texture	6	Local standard deviation of the SIO–XGM residual, 5×5 window; Local standard deviation of the SIO–XGM residual, 15×15; Local standard deviation of the SDUST–XGM residual, 5×5 window; Local standard deviation of the SDUST–XGM residual, 15×15; Local standard deviation of the SWOT-derived residual relative to SIO, 5×5 window; Local standard deviation of the SWOT-derived residual relative to SIO, 15×15

Table 3. *Cont.*

Group	Feature Family	Number of Features	Description
9	Horizontal gradient magnitude and Laplacian	6	Gradient magnitude of the SIO–XGM residual; Gradient magnitude of the SWOT-derived residual relative to SIO; Gradient magnitude of the SDUST–XGM residual; Gradient magnitude of the SIO gravity anomaly; Laplacian of the SIO gravity anomaly; Laplacian of the SWOT gravity anomaly

Note: XGM = gravity anomaly from XGM2019e; SDUST = gravity anomaly from SDUST2022GRA; SIO = gravity anomaly from SIO V32.1; SWOT = SWOT-derived gravity anomaly used in this paper; SIO–XGM, SDUST–XGM, SWOT–XGM and SWOT–SDUST denote the corresponding residual fields.

Table 4. Hyperparameters of the three tree-based ensemble methods.

Hyperparameter	RF	HGB	LGBM
Number of base learners/iterations	n_estimators = 800	max_iter = 3000	n_estimators = 3000
Learning rate	—	learning_rate = 0.008	learning_rate = 0.008
Tree complexity	max_depth = 15	max_leaf_nodes = 15	num_leaves = 15
Minimum samples per leaf	min_samples_leaf = 25	min_samples_leaf = 150	min_child_samples = 150
Minimum samples per split	min_samples_split = 50	—	—
Feature subsampling	max_features = 0.3	—	colsample_bytree = 0.7
Sample subsampling	bootstrap = True	—	subsample = 0.8, subsample_freq = 1
L2 regularisation	—	l2_regularization = 20.0	reg_lambda = 10.0
L1 regularisation	—	—	reg_alpha = 0.0
Early stopping	—	Not explicitly set (sklearn default: auto)	—
Random seed	random_state = 42	random_state = 42	random_state = 42
Parallel jobs	n_jobs = 8	—	n_jobs = 8

4.3. Refined SWOT-Derived Marine Gravity Field via Tree-Based Ensemble Learning Methods

(1) Results of Strategy 1

A two-level, spatially isolated accuracy assessment was adopted. The first level is internal cross-validation within the training set: the training set was partitioned into 5 folds according to 2° × 2° spatial blocks, with each fold held out in turn as the validation fold, and a 0.5° buffer was imposed around the validation fold—samples used for training had to lie beyond the circumscribed circle radius of that fold (1.414°) plus an additional 0.5° extension, i.e., farther than 1.914° (approximately 213 km) from the center of the held-out block. Out-of-fold predictions were thus obtained for every training sample, from which the OOF-RMSE was computed. This metric characterizes the stable accuracy of the model within the region covered by the training data. The second level is evaluation on an independent test set: the entire 5th spatial block group was held out, and during training all training samples within 10 km of any test point were removed, while the test set itself was kept intact. Tree-based ensemble learning methods were evaluated in terms of the accuracy improvement of the corrected gravity anomaly field, as shown in Table 5.

The results show that, under the same independent test set (N = 108,954), the same 67 features, and the same random seed, all three tree-based models reduced the error of the SWOT gravity field relative to shipborne measurements of JAMSTEC, yet the magnitude of improvement differed systematically: LGBM performed best (RMSE 2.398 mGal, a 3.6% improvement over SWOT), followed by HGB (2.402 mGal, 3.5%), while RF was weakest (2.422 mGal, 2.7%). The RMSE on the independent test set ranged from 2.398 to 2.422 mGal. The in-training-set OOF-RMSE of the three methods (2.389–2.394 mGal) differed from the test-set results (2.398–2.422 mGal) by less than 2%. OOF-RMSE reflects the model's

correction capability over regions already covered by existing data as well as its training stability, whereas the independent test set reflects the model's generalization accuracy over entirely unseen spatial units. The closeness of the two indicates that model accuracy does not degrade appreciably under more stringent isolation conditions, and that no evident overfitting occurs.

Table 5. Accuracy of the SWOT-derived gravity field before and after Tree-Based Ensemble Learning Methods correction under strategy 1 (mGal). Note: Validation utilized 569,360 NCEI measurements. The JAMSTEC data were split into training (400,241points) and testing (108,954 points) subsets.

	Institute	MAX	MIN	MEAN	STD	RMSE	Relative Improvement	OOF-RMSE
Before Correction	NCEI	66.846	−45.714	0.072	4.321	4.321	—	—
	JAMSTEC (test)	21.479	−34.691	−0.243	2.476	2.488	—	—
	JAMSTEC (train)	26.834	−30.482	0.033	2.471	2.471	—	—
Corrected by RF	NCEI	53.616	−44.984	−0.046	4.228	4.228	2.2%	—
	JAMSTEC (test)	21.174	−34.536	−0.087	2.420	2.422	2.7%	—
	JAMSTEC (train)	21.925	−32.513	0.005	2.060	2.060	16.6%	2.389
Corrected by HGB	NCEI	55.499	−46.826	−0.040	4.199	4.199	2.8%	—
	JAMSTEC (test)	20.849	−34.220	0.030	2.401	2.402	3.5%	—
	JAMSTEC (train)	23.967	−33.446	0.001	2.047	2.047	17.2%	2.394
Corrected by LGBM	NCEI	55.943	−46.781	−0.048	4.192	4.193	3.0%	—
	JAMSTEC (test)	20.836	−34.327	0.027	2.398	2.398	3.6%	—
	JAMSTEC (train)	23.612	−34.154	0.003	2.046	2.046	17.2%	2.390

Based on the evaluation using independent NCEI shipborne gravity data, the accuracy improvements of different methods from before to after correction range from 2.2% to 3.0%. The LGBM method shows the highest improvement at 3.0%, while the RF method shows the lowest at 2.2%.

Figures 7 and 8 present two key results: the spatial distributions of the predicted gravity residuals obtained by different methods, and the residual distributions between the corrected models derived from different methods and the independent JAMSTEC shipborne gravity data. As shown in the figures, the residual fields obtained by all methods exhibit good spatial continuity. The residuals from the different methods are all approximately normally distributed, as illustrated by the histograms, with only small differences among the methods.

(2) Results of Strategy 2

The modeling part of Strategy 2 is completely identical to that of Strategy 1, with the only difference being the data split. Tree-based ensemble learning methods were evaluated in terms of the accuracy improvement of the corrected gravity anomaly field, as shown in Table 6. As can be seen from Table 6, under the track-based split, all three tree models reduced the error of the SWOT gravity field relative to shipborne gravity over seven complete test cruises (N = 423,986), and the improvement was markedly larger than that under Strategy 1: HGB performed best (RMSE 1.703 mGal, improvement 8.6%), followed by LGBM (RMSE 1.714 mGal, 8.0%), while RF was the weakest (RMSE 1.794 mGal, 3.7%). In terms of OOF-RMSE within the training set, RF was the best (RMSE 2.488 mGal, improvement 1.3%), whereas HGB was the weakest (RMSE 2.497 mGal, improvement 1.0%); on the independent test set, however, the results were exactly reversed. This is consistent with the pattern observed in Strategy 1, where RF also led on OOF and LGBM

led on the test set, indicating that this is not coincidental: RF is superior in cross-validation within the training distribution but is weaker than boosting when extrapolating to entire unseen cruises.

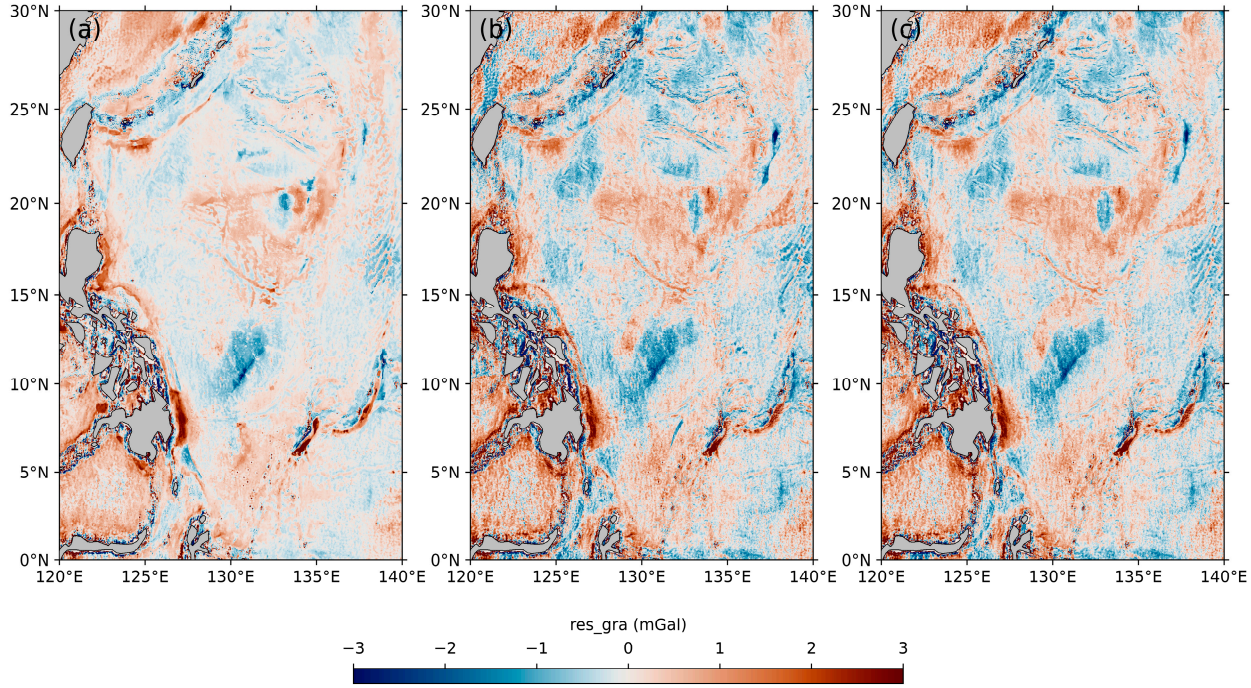


Figure 7. For Strategy 1, spatial distribution of predicted gravity residuals: (**a**) RF; (**b**) HGB; (**c**) LGBM.

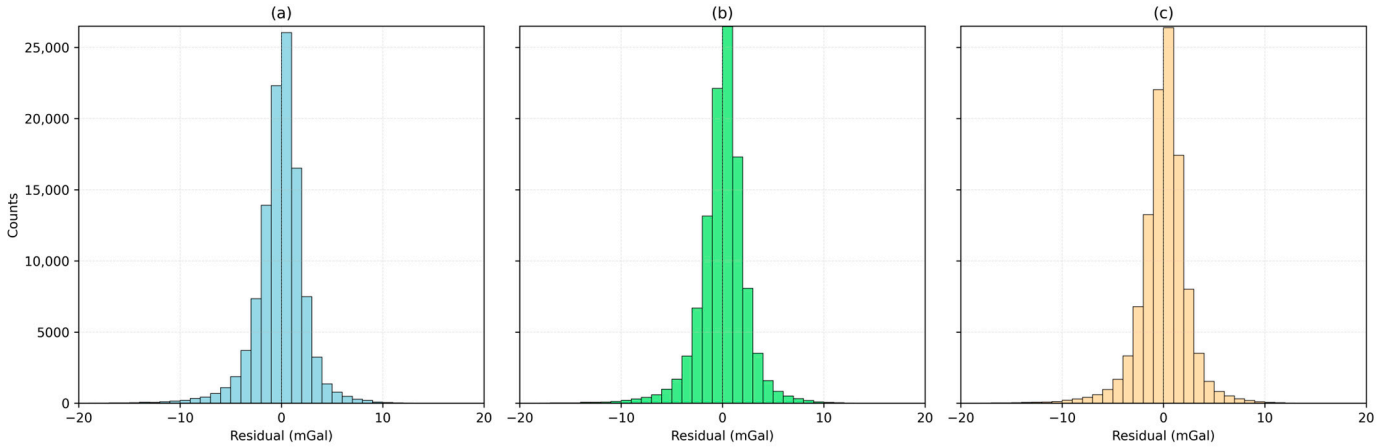


Figure 8. For Strategy 1, histograms of the differences between the gravity fields refined by different methods and the JAMSTEC shipborne gravity data at test set points: (**a**) RF; (**b**) HGB; (**c**) LGBM.

In Strategy 2, the training set was thinned at a 10:1 ratio, whereas the test set comprised complete, unthinned cruises, resulting in a mismatch in data density between the two. In Strategy 1, both sets were thinned and therefore being density-matched. This difference in test set density means that the resulting RMSE values, 1.863 mGal for Strategy 2 and 2.488 mGal for Strategy 1, are derived under different sampling conditions. Consequently, they represent distinct evaluation baselines rather than a direct performance comparison.

Figures 9 and 10 show the spatial distributions of predicted gravity residuals from different methods under Strategy 2, together with the residual distributions of the corrected models relative to the independent JAMSTEC shipborne gravity data. The conclusions are consistent with those of Strategy 1.

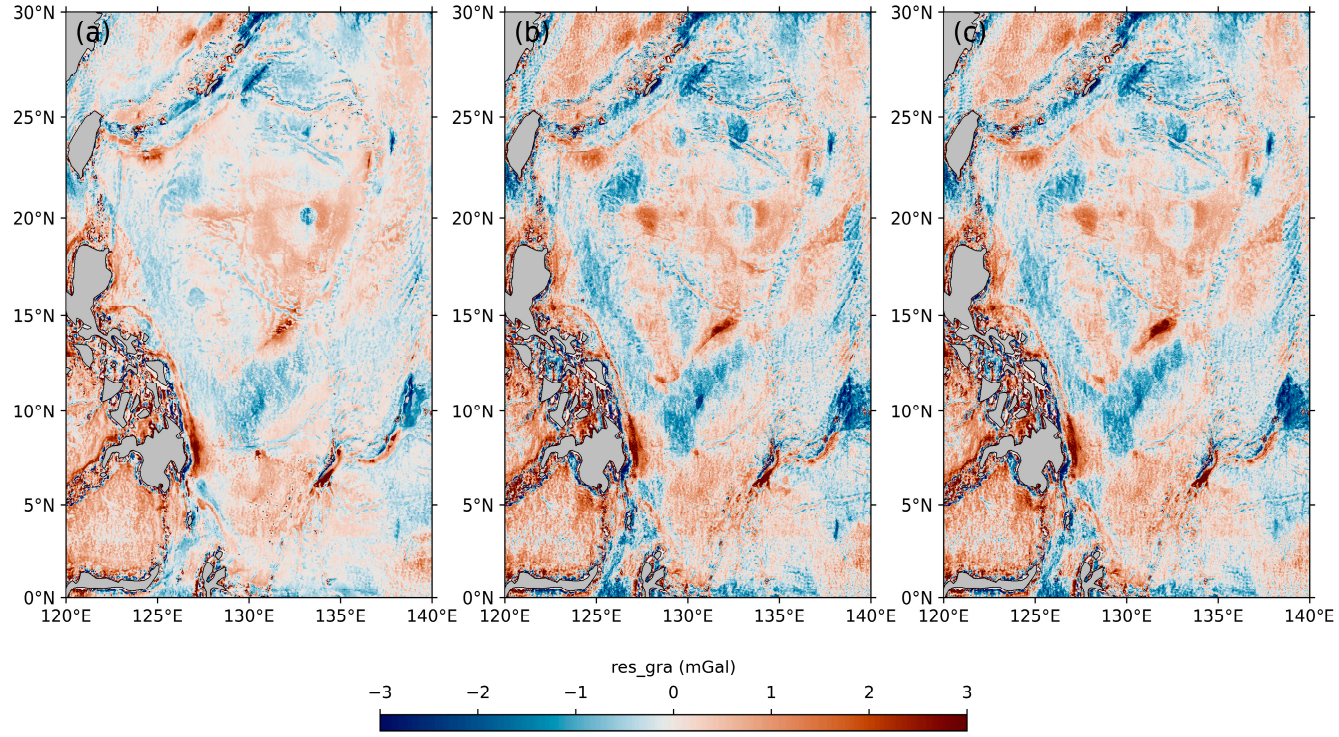


Figure 9. For Strategy 2, spatial distribution of predicted gravity residuals. (**a**) RF; (**b**) HGB; (**c**) LGBM.

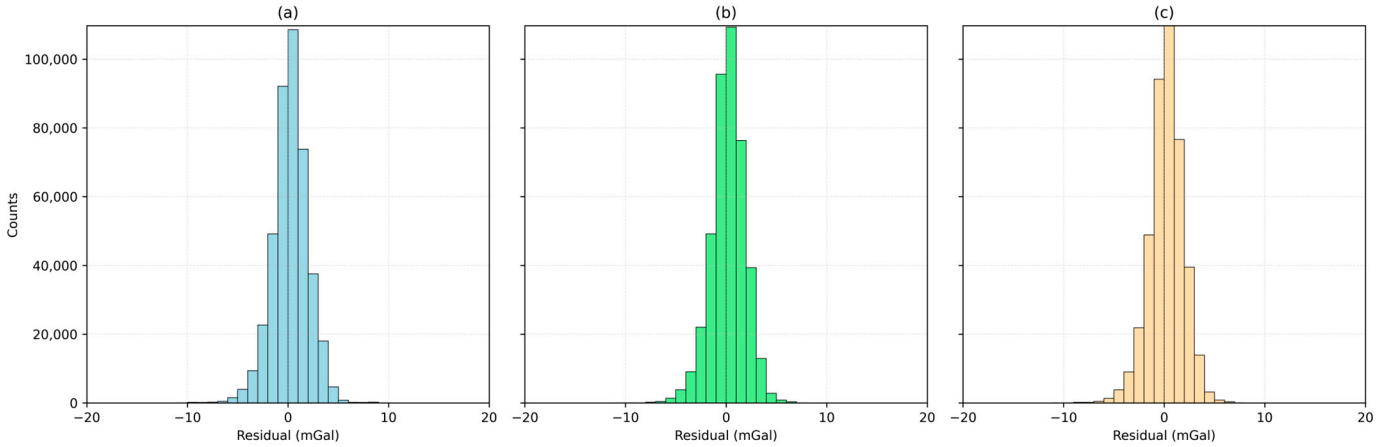


Figure 10. For Strategy 2, histograms of the differences between the gravity fields refined by different methods and the JAMSTEC shipborne gravity data at test set points: (**a**) RF; (**b**) HGB; (**c**) LGBM.

Table 6. Accuracy of the SWOT-derived gravity field before and after Tree-Based Ensemble Learning Methods correction under strategy 2 (mGal). Note: Validation utilized 569,360 NCEI measurements. The JAMSTEC data were split into training (483,650 points) and testing (423,986 points) subsets.

	Institute	MAX	MIN	MEAN	STD	RMSE	Relative Improvement	OOF-RMSE
Before Correction	NCEI	66.846	−45.714	0.072	4.321	4.321	—	—
	JAMSTEC (test)	17.894	−17.881	0.186	1.854	1.863	—	—
	JAMSTEC (train)	27.971	−34.618	−0.037	2.522	2.522	—	—
Corrected by RF	NCEI	50.248	−43.903	0.021	4.249	4.249	1.7%	—
	JAMSTEC (test)	11.010	−15.618	0.253	1.776	1.794	3.7%	—
	JAMSTEC (train)	21.762	−34.186	−0.021	2.179	2.179	13.6%	2.488

Table 6. *Cont.*

	Institute	MAX	MIN	MEAN	STD	RMSE	Relative Improvement	OOF-RMSE
Corrected by HGB	NCEI	57.479	−46.011	0.052	4.204	4.204	2.7%	—
	JAMSTEC (test)	7.719	−13.334	0.218	1.689	1.703	8.6%	—
	JAMSTEC (train)	24.972	−34.112	−0.025	2.198	2.199	12.8%	2.497
Corrected by LGBM	NCEI	56.541	−45.716	0.038	4.208	4.209	2.6%	—
	JAMSTEC (test)	7.927	−13.710	0.234	1.698	1.714	8.0%	—
	JAMSTEC (train)	24.942	−34.193	−0.019	2.194	2.194	13.0%	2.494

4.4. Per-Voyage Analysis of Accuracy Improvement in SWOT-Retrieved Gravity Field via Residual Learning Under Strategy 2

Table 7 shows the accuracy improvements for different cruises under Strategy 2. As can be seen from Table 7, the improvement rates vary greatly among cruises regardless of the method used. In particular, cruise YK13-08 exhibited consistent deterioration in accuracy under all three methods. Cruises KR15-16 and MR17-02 showed consistently high accuracy improvements across the three models (all greater than 15%, with a maximum of 25.2%); these two cruises had lower baseline accuracy and thus greater room for improvement. Cruise YK13-08 had a baseline RMSE of 1.366 mGal, which was already very accurate and left little room for improvement. For some cruises, the baseline mean (MEAN) increased after correction, such as cruise YK10-04, but the overall RMSE still decreased. For cruise MR10-03, which showed the largest degradation in accuracy after RF correction, the RMSE increased by 5.7%. On most cruises where the RF method performed well, boosting outperformed RF by only 1–3%.

Table 7. Accuracy (mGal) of the SWOT-derived gravity field before and after correction by different tree-based ensemble learning methods under Strategy 2, evaluated using shipborne gravity data from seven JAMSTEC test-set cruises.

	Institute	Number of Points	MAX	MIN	MEAN	STD	RMSE	Relative Improvement
Before Correction	KR04-14	29,420	11.775	−16.065	−0.100	1.594	1.597	—
	KR15-16	37,474	11.716	−17.881	−0.151	2.104	2.109	—
	MR10-03	96,495	7.772	−8.437	0.656	1.760	1.878	—
	MR13-01	90,374	7.328	−8.290	0.040	1.953	1.953	—
	MR17-02	36,374	17.894	−8.002	0.083	2.025	2.027	—
	YK10-04	64,266	6.616	−12.321	0.118	2.032	2.035	—
	YK13-08	69,583	7.840	−7.177	0.145	1.358	1.366	—
Corrected by RF	KR04-14	29,420	5.358	−14.568	−0.126	1.490	1.496	6.3%
	KR15-16	37,474	9.290	−15.618	−0.269	1.761	1.781	15.6%
	MR10-03	96,495	8.250	−8.730	0.828	1.804	1.985	−5.7%
	MR13-01	90,374	7.189	−8.337	−0.013	1.836	1.836	6.0%
	MR17-02	36,374	11.010	−7.865	0.105	1.691	1.694	16.4%
	YK10-04	64,266	7.189	−11.363	0.425	1.947	1.993	2.1%
	YK13-08	69,583	7.381	−7.337	0.159	1.383	1.392	−1.9%

Table 7. *Cont.*

	Institute	Number of Points	MAX	MIN	MEAN	STD	RMSE	Relative Improvement
Corrected by HGB	KR04-14	29,420	5.416	−13.334	−0.127	1.438	1.443	9.6%
	KR15-16	37,474	7.510	−13.246	−0.309	1.555	1.585	24.8%
	MR10-03	96,495	7.719	−9.302	0.626	1.618	1.735	7.6%
	MR13-01	90,374	7.145	−8.106	−0.015	1.804	1.804	7.6%
	MR17-02	36,374	7.435	−8.077	0.195	1.661	1.672	17.5%
	YK10-04	64,266	7.588	−11.184	0.532	1.909	1.982	2.6%
	YK13-08	69,583	7.071	−7.202	0.107	1.391	1.395	−2.1%
Corrected by LGBM	KR04-14	29,420	5.438	−13.161	−0.137	1.431	1.438	10.0%
	KR15-16	37,474	7.126	−13.710	−0.313	1.545	1.577	25.2%
	MR10-03	96,495	7.927	−9.342	0.696	1.653	1.794	4.5%
	MR13-01	90,374	7.150	−8.223	−0.009	1.798	1.798	7.9%
	MR17-02	36,374	7.676	−7.874	0.204	1.660	1.672	17.5%
	YK10-04	64,266	7.603	−11.196	0.516	1.910	1.979	2.8%
	YK13-08	69,583	6.894	−7.209	0.118	1.389	1.394	−2.0%

Overall, under the track-based split, HGB achieves the highest accuracy and the most robust extrapolation; LGBM is comparable (differing by only 0.011 mGal); RF is notably weaker on independent cruises and should not be adopted as the final method.

5. Discussion

To evaluate the efficacy of the HGB method in refining the SWOT-derived gravity field model from a spatial frequency (wavenumber) domain perspective, this study conducted a power spectral density (PSD) analysis on three spatially extensive, high-precision, and continuous JAMSTEC shipborne gravity surveys (MR10-03, KR04-14, and MR17-02; track locations shown in Figure 6). The results are presented in Figure 11. For Figure 11a, the accuracy improvement is mainly concentrated at wavenumbers < 0.15 km^{-1}, i.e., wavelengths longer than 7 km. For Figure 11c, the accuracy improvement is also mainly concentrated at wavelengths longer than 7 km, with a middle segment (wavenumbers 0.0076–0.0155 km^{-1}) showing a decrease in accuracy. For Figure 11e, the accuracy improvement is mainly concentrated at wavenumbers 0.0221–0.126 km^{-1}. As shown in Figure 11e, when the wavenumber falls below 0.0221 km^{-1} (wavelength > ~45 km), the corrected model degrades the original long-wavelength components, resulting in a decrease in accuracy. Overall, therefore, the PSD analysis reveals that the accuracy improvements of the gravity field model are predominantly confined to spatial wavelengths of 8–45 km, corresponding to short-to-medium wavelengths.

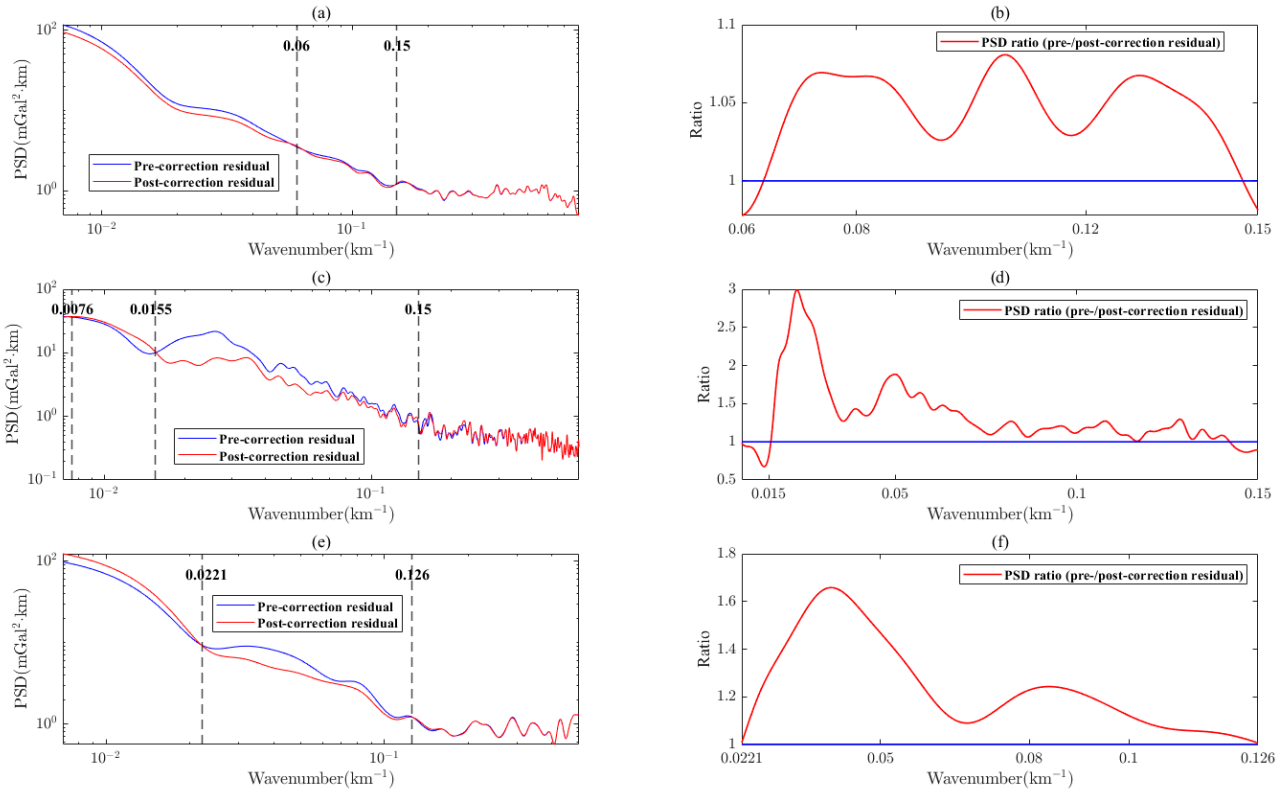


Figure 11. Power spectral density (PSD) of the gravity residuals between the SWOT-derived gravity field, before and after correction by the HGB method, and shipborne gravity measurements: (**a**) PSD of residuals, Cruise MR10-03; (**b**) PSD ratio, Cruise MR10-03; (**c**) PSD of residuals, Cruise KR04-14; (**d**) PSD ratio, Cruise KR04-14; (**e**) PSD of residuals, Cruise MR17-02; (**f**) PSD ratio, Cruise MR17-02.

6. Conclusions

Taking the Western Pacific (120–140°E, 0–30°N) as the study area, this study evaluated three tree-based ensemble learning methods—RF, HGB, and LGBM—for correcting the SWOT-derived gravity field using independent shipborne gravity measurements under two spatially isolated validation strategies. The main conclusions are as follows:

(1) The SWOT L3_LR_SSH v2.0.1 product effectively supports marine gravity field inversion. With 41 observation cycles accumulated, the inverted field achieves an RMSE of 2.476 mGal against the independent JAMSTEC shipborne benchmark and 4.321 mGal against the NCEI benchmark, with accuracy gains showing a converging trend of diminishing marginal returns as the number of SWOT altimetry repeat cycles increases. Using all cycles 1–41, the resulting field has an RMSE of 2.065 mGal relative to the grav_SWOT_04 model.

(2) All three methods are capable of learning the residual and refining the SWOT-derived field. Under Strategy 1 (2° × 2° spatial block hold-out with a 10 km buffer; test set N = 108,954), LGBM performed best (RMSE 2.398 mGal, 3.6% improvement), followed by HGB (2.402 mGal, 3.5%), while RF was weakest (2.422 mGal, 2.7%). Under Strategy 2 (cruise-wise hold-out over seven complete cruises, N = 423,986), the improvements were markedly larger: HGB achieved 1.703 mGal (8.6%), LGBM 1.714 mGal (8.0%) and RF 1.794 mGal (3.7%). A systematic reversal is observed between the two assessment levels: RF leads on the in-training-set OOF-RMSE but is weakest on the independent cruises, whereas HGB shows the opposite. Consistent under both strategies, this indicates that RF is competitive within the training distribution but weaker than boosting methods when extrapolating to entirely unseen surveys. The OOF-RMSE and independent-test RMSE are of comparable magnitude, indicating no evident overfitting.

(3) Power spectral density analysis shows that the improvements are concentrated at wavelengths of about 8–45 km, i.e., short-to-medium wavelengths; the correction therefore primarily refines short-to-medium-wavelength errors of the SWOT-derived field. It should be noted that for wavenumbers below 0.0221 km^{-1} the correction slightly degrades the field, which is masked by the dominant short-wavelength gains in the overall RMSE.

Author Contributions: Conceptualization, J.W.; data curation, J.W.; formal analysis, L.C., J.W., D.A. and C.Z.; funding acquisition, J.W. and C.Z.; investigation, L.C.; methodology, J.W.; project administration, J.W.; resources, J.W.; software, L.C., J.W. and C.Z.; validation, L.C., J.W. and C.Z.; visualization, J.W.; writing—original draft, L.C.; writing—review and editing, L.C., J.W., D.A. and C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the foundation of the Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, China (Grant No. MESTA-2024-A004), the Shandong Provincial Natural Science Foundation (Grant No. ZR2023QD032) and the National Natural Science Foundation of China (Grant No. 42304001).

Data Availability Statement: Publicly available datasets from NCEI (https://maps.ngdc.noaa.gov/viewers/geophysics/, accessed on 5 May 2025), JAMSTEC (https://www.godac.jamstec.go.jp/, accessed on 5 May 2025), SWOT (https://www.aviso.altimetry.fr/en/missions/current-missions/swot, accessed on 1 May 2026), MDT_CNES_CLS22_global (CNES/CLS; https://www.aviso.altimetry.fr/en/data/products/auxiliary-products/mdt/mdt-global-cnes-cls.html, accessed on 1 May 2026), XGM2019e model (https://icgem.gfz-potsdam.de/calcgrid, accessed on 1 May 2026), SDUST2022GRA (https://doi.org/10.5281/zenodo.8337387, accessed on 1 May 2026) and the SIO V32.1 global model's $1' \times 1'$ gravity anomaly and deflection of the vertical grids (https://topex.ucsd.edu/pub/global_grav_1min/, accessed on 1 May 2026) were used in this study.

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

References

1. Li, Z.; Guo, J.; Zhu, C.; Liu, X.; Hwang, C.; Lebedev, S.; Chang, X.; Soloviev, A.; Sun, H. The SDUST2022GRA global marine gravity anomalies recovered from radar and laser altimeter data: Contribution of ICESat-2 laser altimetry. *Earth Syst. Sci. Data* **2024**, *16*, 4119–4135. [CrossRef]
2. Liu, G.-D.; Chen, J. Marine geophysical applications in the field of national security. *Prog. Geophys.* **2011**, *26*, 1885–1896. [CrossRef]
3. Sandwell, D.T.; Garcia, E.S.; Soofi, K.; Wessel, P.; Chandler, M.T.; Smith, W.H.F. Toward 1-mGal accuracy in global marine gravity from CryoSat-2, Envisat, and Jason-1. *Lead. Edge* **2013**, *32*, 892–899. [CrossRef]
4. Wan, X.; Hao, R.; Jia, Y.; Wu, X.; Wang, Y.; Feng, L. Global marine gravity anomalies from multi-satellite altimeter data. *Earth Planets Space* **2022**, *74*, 165. [CrossRef]
5. Liu, X.; Xu, Y.; Peng, H.; Li, Z.; Bian, S. Global marine gravity field models at 1-arcminute resolution derived from optimized fusion of ICESat-2 along-track and cross-track measurements. *Sci. Data* **2026**, *13*, 953. [CrossRef] [PubMed]
6. An, D.; Guo, J.; Chang, X.; Wang, Z.; Jia, Y.; Liu, X.; Bondur, V.; Sun, H. High-precision $1' \times 1'$ bathymetric model of Philippine Sea inversed from marine gravity anomalies. *Geosci. Model Dev.* **2024**, *17*, 2039–2052. [CrossRef]
7. Yan, W.; Liu, X.; Li, Z.; Zhu, C.; Sun, Y.; Huang, L.; Guo, J. Performance of marine gravity anomalies extracted from SWOT KaRIn data of science phase in the Gulf of Mexico. *Sci. Rep.* **2025**, *15*, 38385. [CrossRef] [PubMed]
8. Yu, D.; Hwang, J.; Zhu, H.; Luo, J.; Yuan, J. Enhancing marine gravity field precision using SWOT wide-swath altimetry data. *Geomat. Inf. Sci. Wuhan Univ.* **2024**, *51*, 703–713. [CrossRef]
9. CNES. *DUACS Level-3 SWOT KaRIn (L3_LR_SSH) User Handbook*; SALP-MU-P-EA-23629-CLS, Issue 1.3; CLS/CNES: Toulouse, France, 2024.
10. DUACS, A.V.I.S.O. *SWOT Level-3 SSH Expert, Version 1.0.2 [Data Set]*; CNES: Toulouse, France, 2023. [CrossRef]
11. CNES. *DUACS Level-3 SWOT KaRIn (L3_LR_SSH) User Handbook*; CNES: Toulouse, France, 2025. Available online: https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/handbook_duacs_SWOT_L3.pdf (accessed on 19 April 2026).
12. Jin, T.; Zhou, M.; Zhang, H.; Li, J.; Jiang, W.; Zhang, S.; Hu, M. Analysis of vertical deflections determined from one cycle of simulated SWOT wide-swath altimeter data. *J. Geod.* **2022**, *96*, 30. [CrossRef]
13. Sandwell, D.T. Global Marine Gravity, Bathymetry, and Small-Scale Ocean Interactions from SWOT. Available online: https://swot.jpl.nasa.gov/documents/4357/ (accessed on 12 August 2026).

14. Sun, M.; Feng, W.; An, D.; Chen, X.; Yang, M.; Zhong, M. Preliminary results of marine gravity anomaly and bathymetry from SWOT wide-swath altimeter data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 18107–18116. [CrossRef]
15. Tu, Z.; Jiang, T.; Zhao, F. Marine gravity field modeling using SWOT altimetry data in South China Sea. *J. Mar. Sci. Eng.* **2025**, *13*, 827. [CrossRef]
16. Zhu, C.; Guo, J.; Ya, S.; Li, W.; Gao, J.; Qiu, L. Moving geoid gradient method for high-precision and high-resolution gravity recovery from SWOT wide-swath data. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4205613. [CrossRef]
17. Feng, W.; An, D.; Hwang, C.; Sun, M.; Chen, X.; Zhang, Z.; Yang, M.; Zhong, M. SYSU_Topo: A 1-arc-minute global bathymetry from SWOT-derived gravity using the gravity-geological method. *Sci. Data* **2026**, *13*, 386. [CrossRef] [PubMed]
18. Yu, Y.; Sandwell, D.T.; Dibarboure, G. Abyssal marine tectonics from the SWOT mission. *Science* **2024**, *386*, 1251–1256. [CrossRef] [PubMed]
19. Zhu, C.; Guo, J.; Yuan, J.; Jin, X.; Gao, J.; Li, C. Refining altimeter-derived gravity anomaly model from shipborne gravity by multi-layer perceptron neural network: A case in the South China Sea. *Remote Sens.* **2021**, *13*, 607. [CrossRef]
20. Zhu, C.; Yang, L.; Bian, H.; Li, H.; Guo, J.; Liu, N.; Lin, L. Recovering gravity from satellite altimetry data using deep learning network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5911311. [CrossRef]
21. Qiu, L.; Zhu, C.; Guo, J.; Yang, L.; Li, W. Enhanced deep-learning method for marine gravity recovery from altimetry and bathymetry data. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1501805. [CrossRef]
22. Xu, B.; Chen, X.; Yang, M.; Sun, M.; Feng, W.; Zhong, M. Seafloor topography inversion and quality assessment in complex marine regions based on SWOT_02/03 altimetry data. *Adv. Space Res.* **2026**, *77*, 11524–11537. [CrossRef]
23. Wu, G.; Zhang, Y.; Chen, H.; Zou, F.; Xu, C.; Wu, C.; Zhang, F. Enhancing marine gravity anomaly recovery in the South China Sea using back propagation neural network. *Geophys. J. Int.* **2026**, *245*, ggag126. [CrossRef]
24. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
25. Veloso, D.T. Ensembling Decision Trees. Master's Thesis, University of Porto, Porto, Portugal, 2024.
26. MacFerrin, M.; Amante, C.; Carignan, K.; Love, M.; Lim, E. The Earth Topography 2022 (ETOPO 2022) global DEM dataset. *Earth Syst. Sci. Data* **2025**, *17*, 1835–1849. [CrossRef]
27. Jousset, S.; Mulet, S.; Wilkin, J.; Greiner, E.; Dibarboure, G.; Picot, N. New global mean dynamic topography CNES-CLS-22 combining drifters, hydrological profiles and high frequency radar data. In Proceedings of the Presented at the 2022 Ocean Surface Topography Science Team Meeting, Venice, Italy, 31 October–4 November 2022. [CrossRef]
28. Sandwell, D.T.; Harper, H.; Tozer, B.; Smith, W.H.F. Gravity field recovery from geodetic altimeter missions. *Adv. Space Res.* **2021**, *68*, 1059–1072. [CrossRef]
29. Hwang, C.; Parsons, B. Gravity anomalies derived from Seasat, Geosat, ERS-1 and TOPEX/POSEIDON altimetry and ship gravity: A case study over the Reykjanes Ridge. *Geophys. J. Int.* **1995**, *122*, 551–568. [CrossRef]
30. Andersen, O.B.; Knudsen, P. The DTU17 Global Marine Gravity Field: First Validation Results. In *Fiducial Reference Measurements for Altimetry*; Mertikas, S., Pail, R., Eds.; International Association of Geodesy Symposia; Springer: Cham, Switzerland, 2019; Volume 150. [CrossRef]
31. Sandwell, D.T.; Smith, W.H.F. Marine gravity anomalies from Geosat and ERS 1 satellite altimetry. *J. Geophys. Res. Solid Earth* **1997**, *102*, 10039–10054. [CrossRef]
32. Andersen, O.B.; Knudsen, P. Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry. *J. Geophys. Res. Oceans* **1998**, *103*, 8129–8137. [CrossRef]
33. Fu, L.-L.; Ubelmann, C. On the transition from profile altimeter to swath altimeter for observing global ocean surface topography. *J. Atmos. Ocean. Technol.* **2014**, *31*, 560–568. [CrossRef]
34. Heiskanen, W.A.; Moritz, H. *Physical Geodesy*; W. H. Freeman and Company: San Francisco, CA, USA, 1967.
35. Li, Y.; Zhou, J.; Li, J.; Li, N.; Zhu, F.; Sun, H. Global marine gravity from 27 cycles of SWOT satellite data: Accuracy and spatial resolution evaluation. *J. Geod.* **2025**, *99*, 51. [CrossRef]
36. Hwang, C. Inverse Vening Meinesz formula and deflection-geoid formula: Applications to the predictions of gravity and geoid over the South China Sea. *J. Geod.* **1998**, *72*, 304–312. [CrossRef]
37. Zingerle, P.; Pail, R.; Gruber, T.; Oikonomidou, X. The combined global gravity field model XGM2019e. *J. Geod.* **2020**, *94*, 66. [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.