Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-Domain Feature Fusion with Random Forest for Bathymetry Inversion from Gravity Anomalies

Yuxuan TAN
Chinese Academy of Surveying and Mapping
Baogui KE
kebaogui@163.com

Chinese Academy of Surveying and Mapping
Shanbo LU
Chinese Academy of Surveying and Mapping

Research Article

Keywords: bathymetry inversion, random forest, multi-domain feature fusion, gravity anomaly, semi-supervised domain adaptation

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11726025/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Multi-Domain Feature Fusion with Random Forest for Bathymetry Inversion from Gravity Anomalies

Yuxuan TAN [1], Baogui KE [1,2], Shanbo LU [1]

1. State Key Laboratory of Spatial Datum, Chinese Academy of Surveying and Mapping, Beijing, China
2. Precise Gravity Measurement Facility, Huazhong University of Science and Technology, Wuhan, China
Corresponding author: Baogui KE, kebaogui@163.com

Abstract Bathymetry is a fundamental basis for marine scientific research and ocean engineering construction, and inversion based on satellite altimetry-derived gravity anomalies is a key technique for large-scale seafloor topography modeling. Conventional methods, however, depend on linear transfer-function assumptions or prior constraints, and are therefore susceptible to noise, continuation depth and scale mismatch in topographically complex regions, which limits both inversion accuracy and stability. To address these problems, this paper proposes a multi-domain feature fusion bathymetry inversion method based on random forest. A 21-dimensional feature vector is derived from gravity anomalies in the spatial, frequency and physical correction domains, and a random forest is used to build a nonlinear mapping from these features to bathymetric elevation, thereby avoiding the error amplification caused by linear assumptions and downward continuation factors. To bridge the domain gap between simulated and measured data, a hierarchical progressive training strategy with a semi-supervised domain adaptation algorithm is designed, in which the model is pre-trained on simulated data and then iteratively fine-tuned on a small representative subset of measured data. The proposed method is compared with the admittance function method, the Wiener filtering method, the wavelet multiscale method and the physical constraint method. In simulated experiments, the inversion accuracy reaches 99.08% (RMSE = 7.51 m) under noise-free conditions and remains 97.91% (RMSE = 16.97 m) at 10 mGal noise, with an accuracy retention rate of 98.826%; over continuation depths of 800–2500 m, the RMSE does not exceed 9.22 m and the retention rate is 98.988%. For measured data in the two study areas, the RMSE of the proposed method is 556.82 m and 393.94 m, which is 46–47% and 33–37% lower than that of conventional methods, and the MAE is reduced by more than 50%, while seamounts, basins and steep slopes are reconstructed more faithfully with a more uniform error distribution. Using only a single gravity anomaly dataset, the method also requires lower training cost than deep learning models such as FC-DNN and ARPENN. The proposed method thus provides a high-accuracy and robust technical approach for bathymetry inversion.

Keywords bathymetry inversion; random forest; multi-domain feature fusion; gravity anomaly; semi-supervised domain adaptation

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 41874022), the National Precise Gravity Measurement Facility (Grant No. PGMF-2024-Z002), and the State Key Laboratory of Spatial Datum (Grant Nos. SKLGIE2024-ZZ-7 and SKLSD2026-KF-06).

Introduction

Bathymetry is a fundamental basis for studying ocean dynamics, plate tectonics and resource distribution, and it plays a crucial role in navigation safety and marine engineering construction. With the ongoing global implementation of the Seabed 2030 initiative, high-resolution global bathymetric mapping has become a central objective in modern hydrographic and geophysical sciences (Mayer et al., 2018). Bathymetry inversion methods based on satellite altimetry gravity have shown outstanding performance over large areas, and altimetry data from CryoSat-2/Jason-1 and other satellites have substantially improved the identification of intermediate- to long-wavelength tectonic structures (Sandwell et al., 2014). Together with existing global sounding products, these data constitute an important basis for bathymetry modeling (Smith & Sandwell, 1997). Traditional bathymetry inversion techniques, such as the gravity admittance function method, rely on linear transfer function assumptions. Consequently, their performance degrades substantially in regions with severe topographic relief (Watts, 2001), where spectral estimation is affected by spectral leakage and noise (Crosby, 2007). Although methods such as frequency domain enhancement and gradient fusion have improved the analytical ability of short bands to a certain extent, noise amplification and instability of inversion solutions are still the key factors restricting the inversion accuracy (Wan et al., 2023).

In recent years, machine learning and deep learning methods have provided new paradigms for modeling complex, nonlinear gravity–bathymetry relationships. The capabilities of machine learning in Earth science have been widely demonstrated (Karpatne et al., 2019). Physics-constrained deep learning methods, such as physics-

informed neural networks (PINNs), improve the physical consistency of analysis by incorporating governing equations (Raissi et al., 2019) and have achieved high accuracy in gravity field modeling (Martin & Schaub, 2025). Furthermore, noise modeling, noise injection, and prior-constraint strategies have effectively improved model stability when applied to observed field data (Liu et al., 2024; Zhang et al., 2021). Recent advances include terrain-unit-partitioned back-propagation (BP) neural network prediction frameworks (Fan et al., 2024), altimetry-gravity fusion schemes (Ke et al., 2018), and multi-source observation joint inversion strategies (Fan, 2024). Machine learning algorithms such as random forest offer good stability and interpretability and have been applied to the prediction of various Earth and ocean parameters (Breiman, 2001; Liaw & Wiener, 2002), especially in high-dimensional heterogeneous feature spaces (Zhang et al., 2023). In addition, multiscale wavelet decomposition can separate gravity anomalies into different scale bands, providing a basis for mitigating scale-mismatch effects in gravity–topography inversion (Lin et al., 2022).

Bathymetry inversion has also seen advances in intelligent methods. Studies have quantitatively compared machine learning and traditional approaches in terms of model selection, applicable conditions, and inversion stability, thereby providing a basis for model construction (Kim & Nakata, 2018), including performance comparison (Moran et al., 2022), global-scale neural network prediction (Harper & Sandwell, 2024), and modeling frameworks that integrate machine learning with satellite altimetry (Kim et al., 2024). In particular, feature extraction from gravity field signals can reveal short-wavelength bathymetric features (Annan & Wan, 2024), and frequency-domain fusion methods can improve the inversion accuracy of high-frequency details (Bu et al., 2025). In addition, uncertainty analyses of machine-learning bathymetry inversion have quantified the spatial distribution of prediction uncertainty and its controlling factors, providing a basis for applications in complex marine environments (Zhou et al., 2026). The existing machine learning and deep learning methods can capture the complex nonlinear mapping between gravity and bathymetry and show good adaptability in noise modeling and feature fusion. However, their prediction performance highly depends on the representativeness of training samples and the physical characteristics of feature input, and most methods focus on a single feature domain. The collaborative inversion or modeling of multi-source heterogeneous features therefore requires further study.

To mitigate the impacts of noise, continuation depth, scale mismatch, and feature quality on gravity-topography inversion, and to overcome the difficulty of single-domain machine learning methods in fully capturing complex mapping relationships, a multi-domain feature fusion bathymetry inversion method based on random forest is proposed. This method integrates multi-domain feature vectors and establishes a nonlinear mapping between gravity anomalies and bathymetry using a random forest algorithm, which avoids the problem of increasing error caused by linear hypothesis and downward continuation factor, and improves the stability of the inversion model against noise and continuation depth. By comparing with the admittance function method, the Wiener filtering method, the wavelet multiscale method and the physical constraint method, the inversion performance under different noise and different continuation depth conditions is evaluated.

1 Random Forest Model for Bathymetry Inversion

Random Forest is an ensemble learning algorithm based on the idea of Bagging. It achieves stable modeling of complex nonlinear relationships by constructing multiple independent decision trees in parallel and synthesizing their prediction results. It retains consistent accuracy under high-noise and imbalanced-sample conditions, lowers model variance, enhances the stability and generalization of bathymetry inversion, and eliminates the linear assumptions and prior constraints required in conventional geophysical inversion.

1.1 Multi-domain feature extraction of bathymetry

Multi-scale and multi-dimensional information related to topographic relief is extracted from the gravity anomaly field in three domains (spatial, frequency and physical correction), and a 21-dimensional input feature vector X_i is constructed.

$$X_i = [lat, lon, l_{mean}, l_{std}, l_{grad}, l_{cure}, grav, k, F_{amp}, F_{pha},$$
$$B_{anom}, I_{comp}, V_{grad}, D_{north}, D_{east}, D_{mag}, D_{dir}, G_{mag}, G_{dir}, Lap, G_{dirch}]$$

The feature vector combines the characteristics of spatial domain, frequency domain, physical correction domain and derived gradient, which can fully describe the nonlinear mapping relationship between the gravity field and bathymetry. Each feature vector is derived from gravity anomaly data. The relevant instructions are shown in Table 1.

Table 1 Description of the 21-dimensional multi-domain feature vector for bathymetry inversion

Domain Dimension	Feature Symbol	Feature Name	Geophysical & Mathematical Description
Spatial Domain Features	lat	Latitude	Geographic latitude of observation point, establishing spatial positional reference.
	lon	Longitude	Geographic longitude of observation point, establishing spatial positional reference.
	l_{mean}	Local Mean Gravity Anomaly	Mean gravity anomaly within a 3×3 sliding window; reflects regional gravity baseline and macro-topographic trend.
	l_{std}	Local Standard Deviation	Standard deviation of gravity anomalies in sliding window; measures local gravity field variability and topographic roughness.
	l_{grad}	Local Horizontal Gradient	Spatial variation rate of gravity anomalies in sliding window; correlates with terrain slope and steepness.
	l_{cure}	Local Curvature	Degree of curvature variation in gravity field within sliding window; characterizes terrain inflection and ridge/trough structure.
	$grav$	Observed Gravity Anomaly	Surface marine gravity anomaly value; core observation for bathymetric inversion.
	k	Wavenumber	Spatial wavenumber corresponding to gravity field frequencies.
Frequency Domain Features	F_{amp}	Fourier Amplitude	Spectral amplitude derived from 2D Fast Fourier Transform (FFT); quantifies gravity signal energy across multiscale spatial frequencies.
	F_{pha}	Fourier Phase	Spectral phase derived from 2D FFT; governs spatial positioning and structural morphology of topographic features.
Physical Correction Domain	B_{anom}	Bouguer gravity anomaly	Gravity anomaly corrected for terrain mass and seawater layer density; highlights deep crustal structural signals.
	I_{comp}	Isostatic response component	Isostatic equilibrium compensation derived from the Airy isostatic model; characterizes crustal compensation status.
Derived Gradient Domain	V_{grad}	Vertical Gravity Gradient	Vertical derivative of gravity anomaly ($\partial g/\partial z$); sharpens identification of vertical topographic variations, seamounts and basins.
	D_{north}	Vertical deflection north component	North–south component of vertical deflection; reflects N–S terrain slope and structural tilt.
	D_{east}	Vertical deflection east component	East–west component of vertical deflection; reflects E–W terrain slope and structural tilt.
	D_{mag}	Vertical deflection magnitude	Total magnitude of the vertical deflection vector; comprehensively quantifies overall terrain inclination intensity.
	D_{dir}	Vertical deflection azimuth	Spatial direction of the vertical deflection vector; indicates principal terrain slope and tectonic strike.
	G_{mag}	Horizontal Gradient Magnitude	Magnitude of horizontal gravity gradient; pinpoints zones of rapid gravitational change.
	G_{dir}	Horizontal Gradient Direction	Azimuth of horizontal gravity gradient; indicates orientation of topographic escarpments and structural faults.
	Lap	Second Spatial Derivative	Laplacian operator of gravity anomaly ($\nabla^2 g$); highlights extreme gravity anomalies and abrupt topographic boundaries.
	G_{dirch}	Gradient Direction Rate of Change	Spatial change rate of horizontal gradient direction; characterizes structural curvature and inflection complexity.

In the above features, l_{mean}, l_{std}, l_{grad} and l_{cure} are calculated by applying a 3×3 sliding window to the gravity anomaly grid data; F_{amp} and F_{pha} are obtained by two-dimensional fast Fourier transform. B_{anom} is obtained by deducting the gravity effect caused by terrain mass (Nettleton, 1939). I_{comp} is calculated based on the Airy equilibrium model (Airy, 1855); D_{north} and D_{east} are calculated from the gravity anomaly through the inverse Vening-Meinesz formula (Vening Meinesz, 1928); V_{grad}(Haxby et al., 1983), G_{mag}, G_{dir} (Cordell & Grauch, 1985) and *Lap* are calculated by numerical differentiation of gravity anomaly grid. The calculation of each derived feature is based on the gravity anomaly data itself, and no additional observation data is introduced.

1.2 Random forest construction and inversion process

The feature vector listed in Table 1 serves as the model input, and the complex nonlinear mapping from gravity anomaly to bathymetric elevation is realized through the ensemble learning mechanism. The inversion workflow comprises algorithm selection, parameter optimization, iterative fine-tuning and the final inversion.

1.2.1 Model architecture design

By constructing multiple decision trees and integrating their prediction results, the random forest effectively reduces the risk of overfitting that may occur in a single tree and improves the generalization ability of the model. The input of the model is a 21-dimensional feature vector X_i, and the output is the predicted terrain elevation value H_i.

The model training process adopts a hierarchical progressive strategy: firstly, the basic model is established based on the simulated data to learn the general correlation between the gravity field and the terrain; then, through the semi-supervised domain adaptation method, the model is optimized and adjusted by using a small amount of measured data to adapt to the feature distribution of the actual observation data. This two-stage training strategy effectively solves the problem of domain difference between simulated data and measured data.

Aiming at the distribution difference between the simulated data and the measured data, a progressive semi-supervised domain adaptation algorithm is designed. Through iterative training, the model smoothly transitions from the simulated data domain to the measured data domain. The specific implementation steps are as follows:

(1) Representative sample selection: According to the principle of clustering sampling, about 15% of the samples are selected from the measured data as the adaptive training set. Specifically, K-means clustering is used to spatially cluster the gravity anomalies and bathymetric data. In the neighborhood of each cluster center, areas with dramatic bathymetric gradients and diverse bathymetric and geomorphological features are preferentially selected to ensure that the adapted samples can cover the main terrain features of the study area. In addition, the algorithm automatically identifies the regions with large prediction errors of the basic model (samples with errors exceeding the 75th percentile), and gives them double weights in subsequent iterations to strengthen the adaptation to difficult samples.

(2) Construction of mixed target value: At the t-th iteration (t = 1, 2, 3, ..., T_{max}), the mixed target value $y^{(t)}$ is constructed:

$$y^{(t)} = \alpha_t \cdot f^{(t-1)}(X) + (1 - \alpha_t) \cdot y^{(t-1)} \quad (1)$$

where $f^{(t-1)}(X)$ is the predicted value of the input feature X by the model at the (t−1)th iteration, $y^{(t-1)}$ is the target value constructed at the (t−1)th iteration, and $y^{(0)}$ is initialized to the measured true value. The dynamic weight coefficient α_t adopts the piecewise design strategy:

$$\alpha_t = \begin{cases} 0.3 + 0.6 \cdot \frac{t}{T_{half}}, & t \le T_{half} \\ 0.9 + 0.1 \cdot \frac{t - T_{half}}{T_{\max} - T_{half}}, & t > T_{half} \end{cases} \quad (2)$$

where T_{half} = $[T_{max}/2]$. The strategy achieves a rapid transition from relying on simulated data prediction to relying on measured truth values in the first half, and fine adjustment in the second half. The first iteration (t = 1) uses a special initialization: $y^{(1)} = 0.9 \cdot f^{(0)}(X) + 0.1 \cdot y^{(0)}$, that is, 90% of the weight is assigned to the basic model prediction, and 10% is assigned to the measured truth value to ensure initial stability.

(3) Iterative fine-tuning: The model is iteratively fine-tuned with the mixed target value as the supervisory signal. After each iteration, the RMSE of the model is calculated on the selected representative samples. Iteration stops when the maximum number of iterations T_{max} = 7 is reached, or when the RMSE does not decrease for three consecutive iterations. Finally, the model with the smallest RMSE on the validation set is selected as the adaptation result, so as to avoid the model oscillation caused by the large difference between the simulated data and the measured data, and improve the stability and effectiveness of the model. In the adaptation process, the parameters of the random forest model are set as follows: the number of decision trees T is 150, the maximum depth is 20, the minimum leaf node sample number is 1, the feature subset ratio is 0.6, and the sampling ratio of each tree is 0.7.

1.2.2 Inversion process design

(1) Data preprocessing. The quality control of input gravity anomaly data is carried out, including outlier detection and processing, missing value interpolation, grid unification and other operations to improve the consistency and reliability of data.

(2) Feature extraction and standardization. The 21-dimensional feature vector is extracted and standardized:

$$X_{\text{norm}} = \frac{X - \mu}{\sigma} \quad (3)$$

where X, μ and σ are the feature values, means and standard deviations of the 21-dimensional feature vector, respectively. Standardization eliminates the dimensional differences between features and improves the stability of model training.

(3) Model prediction and bathymetry reconstruction. The standardized features are input into the random forest model described in Section 1.2.1 to obtain a preliminary bathymetric prediction. The final prediction of the random forest is the average of the predicted values of all decision trees:

$$\hat{h}(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (4)$$

where T is the number of decision trees, and $h_t(x)$ is the predicted value of the t-th tree. All prediction results are then integrated and gridded to construct a complete bathymetry grid.

1.3 Evaluation of inversion accuracy and algorithm performance

In order to comprehensively and objectively evaluate the performance of different inversion methods, the following accuracy and stability evaluation metrics are used in this paper:

(1) Accuracy evaluation metrics

Root mean square error (RMSE): measures the overall deviation between predicted and true values.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}\left(h^{(i)} - \hat{h}^{(i)}\right)^2} \tag{5}$$

Mean absolute error (MAE): insensitive to outliers, reflecting the mean absolute deviation.

$$\mathrm{MAE} = \frac{1}{N_p}\sum_{i=1}^{N_p}\left|h^{(i)} - \hat{h}^{(i)}\right| \tag{6}$$

Coefficient of determination (R^2): the ability of the model to explain the variance of the target variable.

$$R^2 = 1 - \frac{\sum_i (h^{(i)} - \hat{h}^{(i)})^2}{\sum_i (h^{(i)} - \bar{h})^2} \tag{7}$$

where $h^{(i)}, \hat{h}^{(i)}$ and $\bar{h}$ are the true bathymetric elevation, the model-predicted elevation and the mean true bathymetric elevation, respectively.

(2) Stability evaluation metrics

Accuracy change percentage: the difference between the initial and final accuracy divided by the initial accuracy, giving the relative change percentage. It represents the anti-interference ability of each inversion method.

$$\delta_{rel} = \frac{\left|\mathrm{Accuracy}_{initial} - \mathrm{Accuracy}_{final}\right|}{\mathrm{Accuracy}_{initial}} \times 100\% \tag{8}$$

Accuracy change rate: the absolute difference between the initial and final accuracy divided by the change in interference intensity, giving the decrease in accuracy per unit interference intensity for each inversion method. It represents the sensitivity of each method to changes in interference intensity; the larger the value, the faster the performance decays as interference increases.

$$r_{decline} = \left|\frac{\mathrm{Accuracy}_{initial} - \mathrm{Accuracy}_{final}}{\Delta\mathrm{Noise}/\Delta\mathrm{H}}\right| \tag{9}$$

Accuracy retention rate: the final accuracy divided by the initial accuracy, expressed as a percentage. It represents the inversion capability that each method retains after disturbance.

$$R_{retention} = \frac{\mathrm{Accuracy}_{final}}{\mathrm{Accuracy}_{initial}} \times 100\% \tag{10}$$

where $\mathrm{Accuracy}_{initial}$ is the initial inversion accuracy, $\mathrm{Accuracy}_{final}$ is the final inversion accuracy, ΔNoise and ΔH represent the noise level and the variation in continuation depth, respectively.

2 Analysis of Factors Influencing Bathymetry Inversion

2.1 Simulated Data

To investigate the effects of gravity anomaly errors on bathymetry inversion and to analyze the sensitivity of random forest inversion to continuation depth, experiments were conducted using simulated bathymetry and gravity anomaly data. The simulated bathymetry served as the target field for the inversion. The study area spans longitude 120°–123° and latitude 15°–18°. The peripheral region is characterized by gentle slopes, with elevations ranging from 0 to 200 m. The central region, in contrast, exhibits strong topographic variability, with elevations rising sharply from 200 m to nearly 900 m and forming a peak-shaped bathymetric high. The overall relief and slope variations are centrally symmetric (Fig. 1).

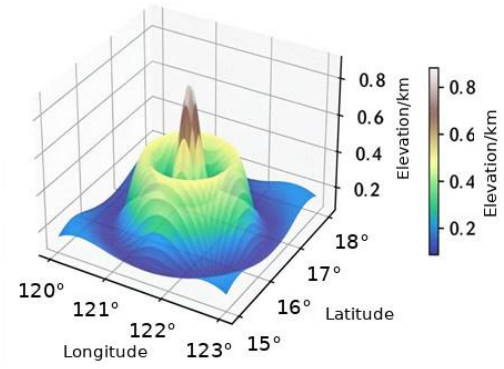


Fig. 1 Simulated bathymetry

The simulated gravity anomaly data were generated by forward modeling based on the bathymetry model. Because satellite altimetry-derived gravity anomalies reflect the combined gravity effect of geological bodies at different depths between the sea surface and the seafloor, upward continuation to different heights simulates the effect of satellite observation height. The 800 m height corresponds to typical observation conditions in intermediate-water-depth areas, whereas 2500 m represents an abyssal-plain setting; together, these heights span the main seafloor types from continental slope to abyssal plain. Therefore, gravity anomaly data were calculated at seven heights from 800 to 2500 m to test the effect of continuation depth on random forest bathymetry inversion. Random noise was added to the data during the analysis.

When constructing the random forest model, the number of decision trees (T) is a key parameter affecting both accuracy and computational efficiency. Table 2 compares the inversion results for different T values. T = 150 achieved the best balance between inversion accuracy and computational cost. Too small a value of T increases model variance, whereas too large a value causes computational redundancy with limited accuracy gain. Therefore, T = 150 was used in subsequent experiments.

Table 2 Inversion accuracy for different numbers of decision trees

T	RMSE(m)	MAE(m)	R^2
70	3.242	1.295	0.999500
100	3.239	1.338	0.999501
120	3.235	1.307	0.999502
150	3.226	1.317	0.999505
180	3.229	1.289	0.999504

2.2 Effects of Gravity Anomaly Errors on Bathymetry Inversion Results

To evaluate the robustness of the random forest model to gravity anomaly errors, the admittance function method (McKenzie & Bowin, 1976), the Wiener filtering method (Wiener, 1949), the wavelet multiscale method (Little et al., 1993), the physical constraint method (Parker, 1973) and the random forest model were used for comparative inversion. Random noise with amplitudes up to 10 mGal was added to the simulated gravity anomaly data. The accuracy metrics of each method are shown in Fig. 2, and the robustness evaluation results are given in Table 3.

As gravity anomaly noise increased from 0 to 10 mGal, the accuracy metrics of the inversion methods exhibited different trends. The random forest model had the highest accuracy (99.08%) under noise-free conditions, but its accuracy decreased as noise increased, reaching 97.91% at 10 mGal. The corresponding RMSE increased from 7.51 to 16.97 m, while R^2 remained above 0.986. The wavelet multiscale method

achieved 98.45% accuracy at 0 mGal and performed well at low noise levels, but it was sensitive to increasing noise; at 10 mGal, its accuracy decreased to 98.08%, and RMSE increased from 12.64 to 15.63 m. The accuracy trends of the Wiener filtering method and the physical constraint method were similar. At 10 mGal noise, their accuracies were approximately 97.65% and 97.11%, respectively, and RMSE increased to approximately 19.09 and 23.53 m. The accuracy metrics of the admittance function method fluctuated least across noise levels, remaining within 93.06–93.07%, indicating robustness to noise; however, its absolute accuracy was lower than that of the other methods (Fig. 2).

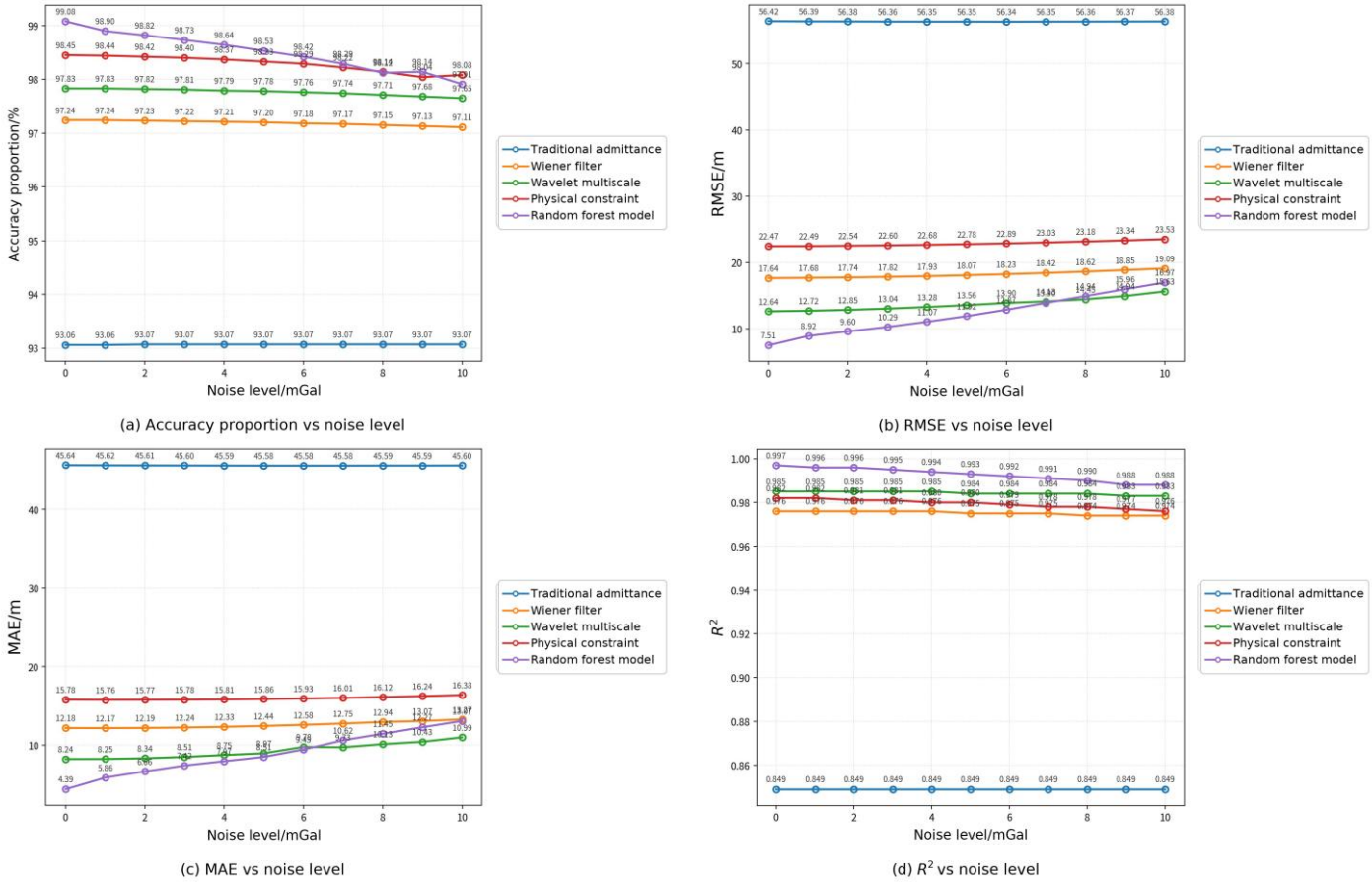


Fig. 2 Influence of different noise levels on inversion accuracy indicators

In terms of percentage change in accuracy, the admittance function method was least affected by noise (0.005%), followed by the physical constraint method (0.134%). The Wiener filtering method, the wavelet multiscale method and the random forest model showed progressively larger changes (0.182%, 0.373%, and 1.174%, respectively). The accuracy retention rate was 99.995% for the admittance function method and 99.866% for the physical constraint method, indicating good noise stability. The random forest model had a relatively low retention rate of 98.826% (Table 3).

Table 3 Robustness evaluation of various inversion methods under different noise levels

Method	δ_{rel} (%)	$r_{decline}$ (%/mGal)	$R_{retention}$ (%)
Admittance function	0.005	0.0004	99.995
Wiener filtering	0.182	0.0178	99.818
Wavelet multiscale	0.373	0.0367	98.627
Physical constraint	0.134	0.0130	99.866
Random forest	1.174	0.1163	98.826

The lower accuracy retention rate of the random forest model (98.826%) than those of the admittance function method (99.995%) and the physical constraint method (99.866%) does not indicate insufficient inversion stability; rather, it reflects differences in how the methods respond to noise. The admittance function and physical constraint methods are based on linear physical models, which smooth small noise; therefore, their accuracy fluctuates only slightly, but their absolute accuracy is lower (about 93% and 97% under noisy conditions). The random forest model achieved the highest initial accuracy (99.08%) under noise-free conditions. Its accuracy decrease was mainly due to the model's high sensitivity to noise components in the data. Some decision trees may misinterpret noise as a valid signal, causing prediction bias. Nevertheless, the random forest model still maintained an absolute accuracy of 97.91% at 10 mGal noise, much higher than the 93.06% of the admittance function method.

2.3 Sensitivity of the Random Forest Model to Continuation Depth

To analyze the performance of the random forest model for bathymetry inversion at different continuation depths, accuracy and stability metrics were used to compare inversion results at seven observation heights from 800 to 2500 m; RMSE and R^2 were calculated. The results are shown in Fig. 3 and Table 4.

The random forest model showed better overall accuracy and stability than the other methods at all depths: accuracy remained above 98.87%, RMSE did not exceed 9.22 m, and R^2 was higher than 0.995. The accuracies of the wavelet multiscale and Wiener filtering methods at different depths ranged from 96.99% to 98.50%, with RMSE of about 12–24 m; their overall performance was stable but slightly inferior to that of the random forest model. The physical constraint method achieved 97.24% accuracy at 800 m, but its accuracy decreased to 96.50% at 2500 m, with an RMSE of 28.48 m. The accuracy metrics of the admittance function method changed little with continuation depth, remaining at 92.69–93.09%, with RMSE of about 56–59 m; its overall accuracy was low and showed no clear trend (Fig. 3).

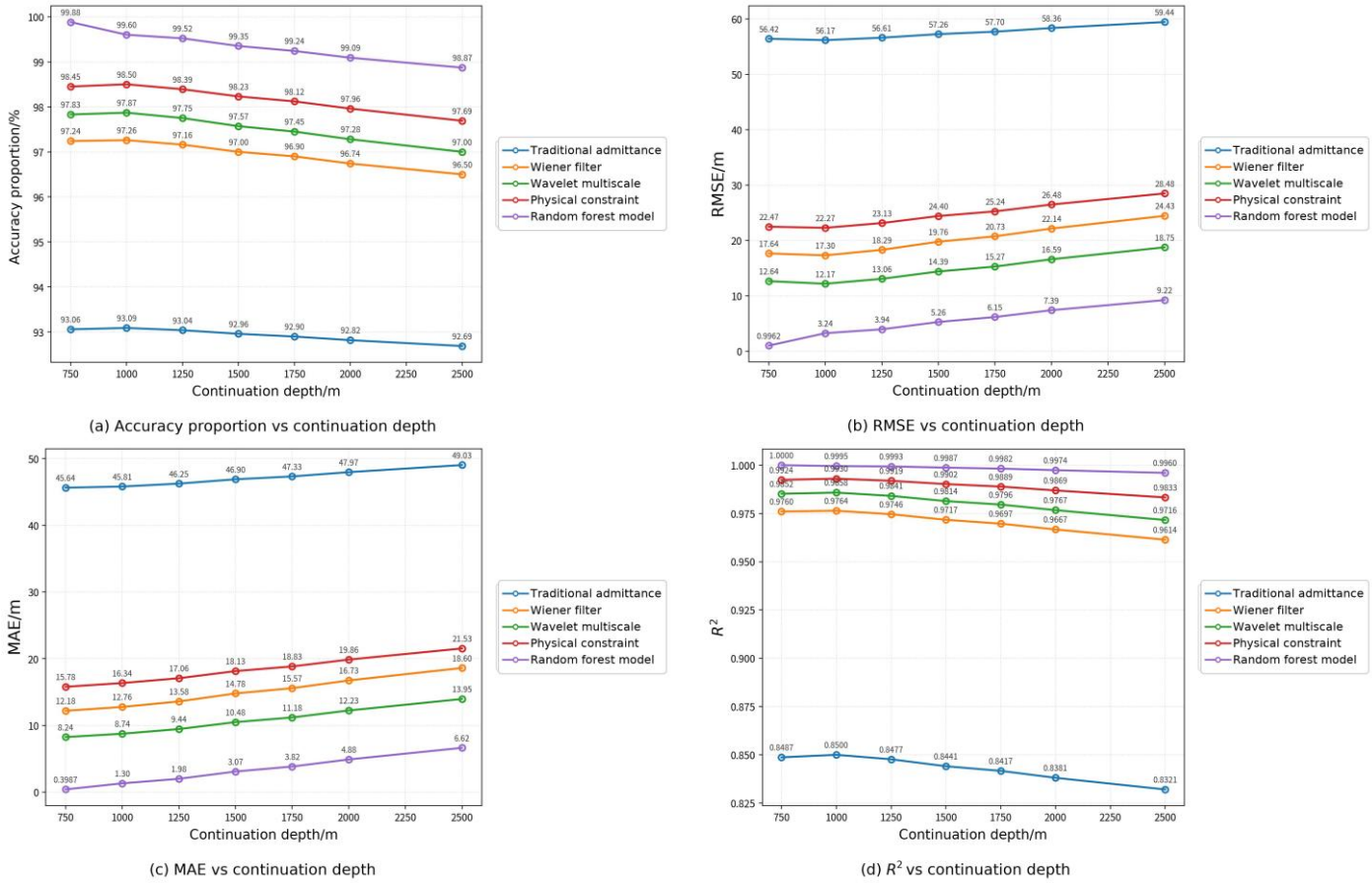


Fig. 3 Effects of continuation depth on inversion accuracy

For the random forest model, the accuracy changed by 1.012% over the 800–2500 m range, with a decrease rate of 0.595%/km, indicating a weak dependence on continuation depth. The physical constraint method showed a percentage change of 0.761% and a rate of 0.435%/km, and the wavelet multiscale method showed 0.772% and 0.447%/km, respectively; both had similar depth adaptability, and their changes were smaller than those of the Wiener filtering method (0.853%, 0.491%/km). The admittance function method had a percentage change of 0.399%, but its overall accuracy was low, limiting its practical response to depth (Table 4).

Table 4 Accuracy analysis of various inversion methods under different continuation depths

Method	δ_{rel} (%)	$r_{decline}$ (%/km)	$R_{retention}$ (%)
Admittance function	0.399	0.218	99.601
Wiener filtering	0.853	0.491	99.147
Wavelet multiscale	0.763	0.442	98.237
Physical constraint	0.761	0.435	99.239
Random forest	1.012	0.595	98.988

Similarly, in the continuation depth sensitivity analysis experiment, the accuracy retention rate of the random forest model (98.988%) was lower than that of the physical constraint method (99.239%) and the Wiener filtering method (99.147%), but its absolute accuracy at 2500 m depth (98.87%, RMSE = 9.22 m) was still significantly better than that of the physical constraint method (96.50%, RMSE = 28.48 m). This shows that the accuracy of the random forest model decreases slightly on the basis of very high initial accuracy, and its absolute accuracy remains optimal at all depths.

2.4 Performance Advantages of the Random Forest Model

The admittance function method relies on a linear assumption and struggles to capture the complex relationship between bathymetry and gravity anomalies. The Wiener filtering method requires prior knowledge of noise statistics. The wavelet multiscale method is susceptible to noise interference and error amplification. The prior model used in the physical constraint method becomes increasingly inconsistent with reality with depth. In contrast, by integrating multiple decision trees, the random forest model can adaptively learn nonlinear relationships in the data without prescribing a functional relationship.

3 Bathymetry Inversion Using Measured Data: Results and Discussion

To evaluate the effectiveness and applicability of the inversion methods in real ocean environments, experiments were conducted using two measured datasets corresponding to the study areas of Zhao et al. (2025) and Li et al. (2023), respectively. Bathymetric data were obtained from the GEBCO 2025 global bathymetric model (GEBCO Compilation Group, 2025), and gravity anomaly data were obtained from the SIO V33.1 global gravity anomaly model (Sandwell et al., 2021). After cropping the data to each study area, inversion comparisons were performed, and the inversion accuracy and bathymetric reconstruction capability of the different methods were systematically analyzed.

3.1 Inversion results for the study area of Li et al. (2023)

The data in Table 5 show that the inversion results differed significantly among the methods in this area. The RMSEs of the admittance function method, the Wiener filtering method, the wavelet multiscale method and the physical constraint method ranged from 1024.22 to 1053.30 m, and their MAEs ranged from 916.30 to 965.52 m, with little difference among them. This indicates that in this region, relying solely on linear assumptions, statistical noise assumptions, frequency-domain scale processing, or prior constraints is insufficient to fully characterize the complex gravity–topography relationship.

In contrast, the random forest model reduced the RMSE to 556.82 m and the MAE to 456.22 m. Compared with the conventional methods, it reduced the RMSE by approximately 46–47% and the MAE by approximately 50–53%, indicating that the random forest model not only has an advantage in overall inversion but also exhibits stronger error suppression in terms of mean absolute error. Furthermore, the RMSE/maximum depth ratio of the random forest model decreased from 10.44–10.74% for the conventional methods to 5.68%, approximately halving the error proportion relative to topographic relief. Its STD was 369.01 m, lower than the 407.19–457.63 m of the conventional methods, indicating smaller error dispersion and more stable inversion results.

Table 5 Accuracy metrics of bathymetry inversion based on measured data

Method	RMSE(m)	MAE(m)	STD(m)	MAXdepth(m)	RMSE/MAXdepth(%)
Admittance function	1029.46	945.50	407.19	9811.17	10.49
Wiener filtering	1024.22	916.30	457.63	9811.17	10.44
Wavelet multiscale	1053.30	965.52	420.96	9811.17	10.74
Physical constraint	1037.55	941.51	435.98	9811.17	10.58
Random forest	556.82	456.22	369.01	9811.17	5.68

Fig. 4 shows the original bathymetry of the region and the three-dimensional inversion results obtained by each method. Seamounts, knolls, basins, and steep slopes are interleaved in this area, with large topographic relief. The admittance function method and the physical constraint method produced overly smooth results overall and reproduced only the general outlines of seamounts, with substantial loss of detail on seamount flanks,

knolls, and slope-break zones. The Wiener filtering method reproduced the general bathymetry, but local topographic distortion occurred in the slope zone. The wavelet multiscale method captured some high-frequency information, but surface noise was prominent, making it easy to confound bathymetric signals with observation noise. The random forest results agreed best with the original bathymetry, better reproducing seamount peaks, steep flanks, and basin slope breaks. Errors were sparsely distributed only along the edges of extremely steep seamounts, and the overall error distribution was uniform.

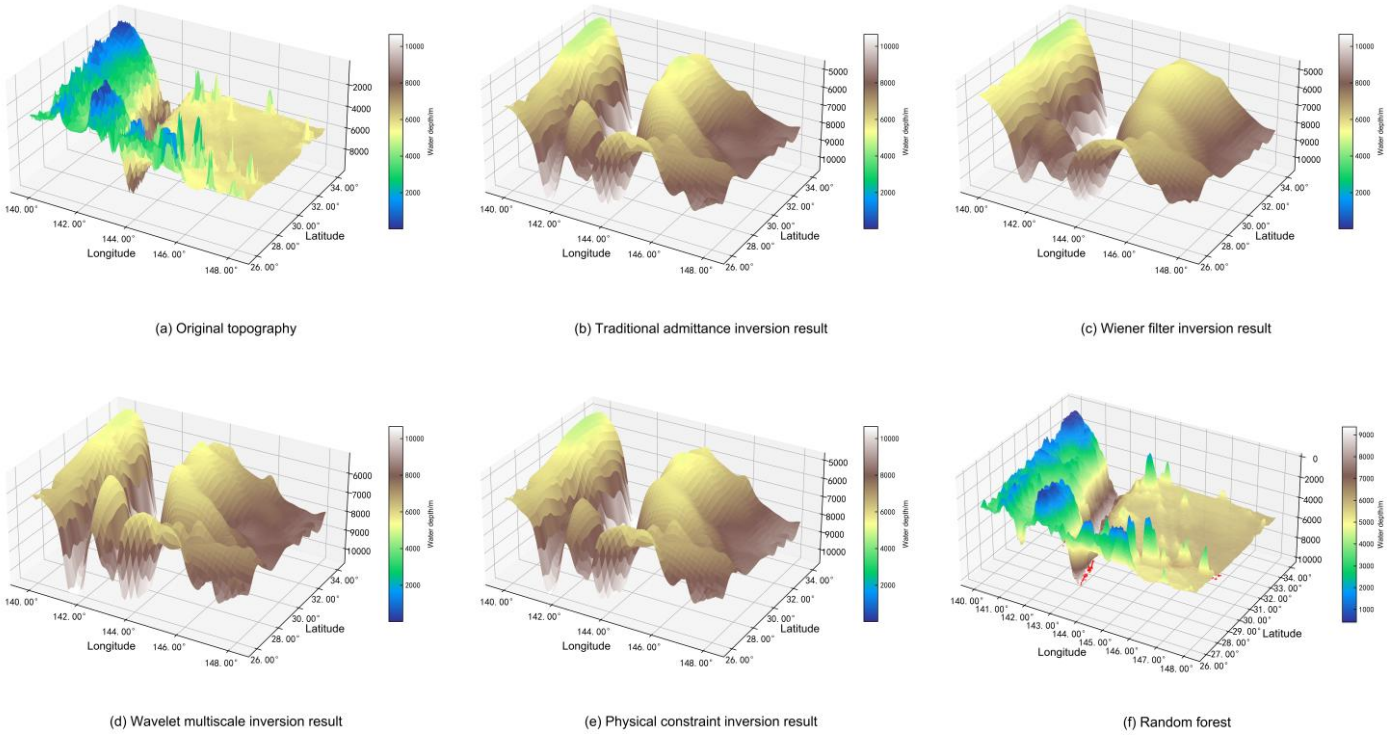


Fig. 4 Measured bathymetry and inversion error maps for each method

The fully connected deep neural network (FC-DNN) used by Li et al. (2023) integrates multiple gravity field elements, including gravity anomalies, vertical gravity gradient anomalies, and deflections of the vertical, and achieved a standard deviation of 49.20 m in this region, with a correlation coefficient of 0.9987. Unlike the random forest model in this study, FC-DNN uses multi-source gravity field data, whereas the proposed model uses only a single gravity anomaly dataset as input. Although the training samples and study areas are broadly similar, the data-splitting strategies differ. Therefore, the main contribution of this study is that the proposed model, using only a single gravity anomaly dataset, achieves better inversion accuracy than conventional methods while requiring lower training cost than FC-DNN.

3.2 Inversion results for the study area of Zhao et al. (2025)

Table 6 shows that the RMSEs of the conventional methods in this region ranged from 589.68 to 623.75 m, and their MAEs ranged from 457.50 to 472.33 m. The overall accuracy was higher than that in the Li study area, indicating that gravity anomalies are more strongly correlated with bathymetry in this region and that the conventional methods perform better. Nevertheless, the random forest model reduced the RMSE to 393.94 m and the MAE to 221.11 m. Compared with the conventional methods, the RMSE decreased by approximately 33–37% and the MAE by approximately 52–53%.

Table 6 Accuracy metrics of bathymetry inversion based on measured data

Method	RMSE(m)	MAE(m)	STD(m)	MAXdepth(m)	RMSE/MAXdepth(%)
Admittance function	591.26	457.96	433.29	7771.70	7.61
Wiener filtering	589.68	457.50	430.89	7771.70	7.59
Wavelet multiscale	623.75	472.33	491.55	7771.70	8.03
Physical constraint	597.70	459.75	447.28	7771.70	7.69
Random forest	393.94	221.11	347.15	7771.70	5.07

The reduction in MAE for the random forest model was greater than that in RMSE, indicating that it better suppresses small- and medium-magnitude errors for most samples, whereas the conventional methods may contain localized large errors that inflate the RMSE. In addition, the RMSE/maximum depth ratio of the random forest model decreased from 7.59–8.03% for the conventional methods to 5.07%, and the STD decreased from 430.89–491.55 m to 347.15 m, indicating more concentrated errors and more stable bathymetric reconstruction.

Fig. 5 shows the original bathymetry of the study area and the three-dimensional inversion results obtained by each method. The area includes a flat abyssal plain in the west and a steep slope with a strong gradient in the east. The admittance function method, the Wiener filtering method and the physical constraint method reproduced the macroscale bathymetric trend of the basin, but the steep slopes were markedly smoothed, reflecting the insufficient response of linear models to high-frequency variations in strongly gradient terrain. The wavelet multiscale method partially captured slope variations, but large errors remained on the steep slopes and in the basin, where high-frequency signals were easily aliased with noise. The random forest model, using 21-dimensional multi-domain feature fusion, accurately reconstructed the bathymetry of the western abyssal basin without spurious undulations and well reproduced the slope gradients, step-like relief, and land–sea boundary in the east, demonstrating its comprehensive inversion capability across various landforms.

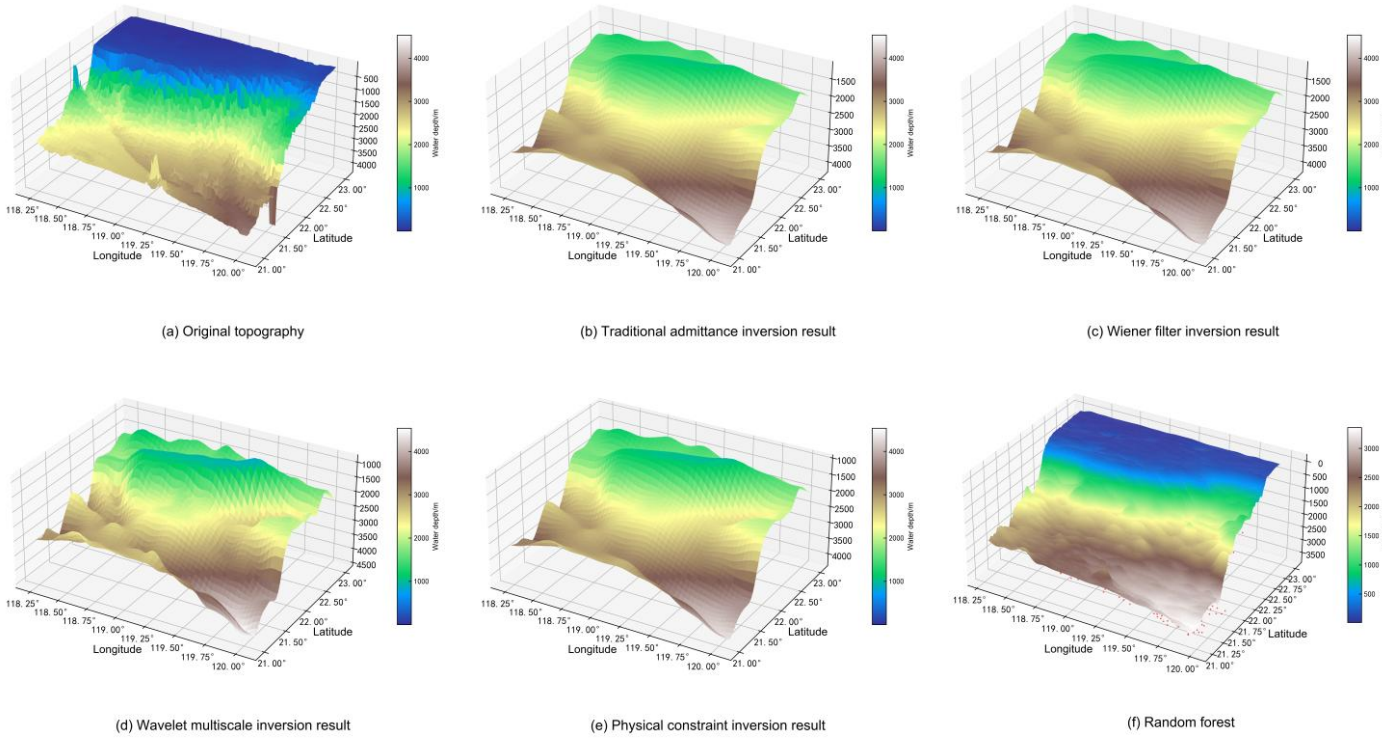


Fig. 5 Measured bathymetry and inversion error maps for each method

The physics-augmented neural network (ARPENN) proposed by Zhao et al. (2025), which combines residual modules and physical constraints, achieved an RMSE of 77.37 m in this region, an improvement of

40.11% over the conventional SAS method. The input features of ARPENN differ from those of the proposed model: ARPENN introduces a gravity–topography physical correlation prior through a physics-constrained loss function, whereas the proposed model is purely data-driven and does not incorporate physical constraints. This comparison is intended to illustrate the effect of embedding physical information on inversion accuracy, rather than to provide a strictly controlled comparison. Even when using only gravity anomaly data, the proposed method achieved an RMSE of 393.94 m, demonstrating the effectiveness of the multi-domain feature fusion strategy.

3.3 Discussion

The comparative analysis of measured data from the two regions reveals several aspects of the random forest model that require improvement. First, as a purely data-driven method, its inversion performance depends on how well the training samples cover the target geological space. If the geological structure of the target area differs systematically from that of the training area, the model's interpolation/extrapolation capability may degrade considerably. Second, the current model does not explicitly incorporate physical coupling constraints between the gravity field and bathymetry. Consequently, under extreme topographic conditions, such as steep seamount flanks or deep-water basin centers, the physical consistency between the inversion results and gravity potential field theory lacks rigorous closure testing. Although these limitations are beyond the scope of this study, they provide a reference for further developing hybrid modeling strategies that integrate prior geological information.

4 Conclusions

Based on analyses of synthetic data and tests on measured data, the multi-domain feature fusion random forest method for bathymetry inversion achieved the following results.

In the synthetic noise experiments, when the gravity anomaly noise was 0 mGal, the random forest model achieved an inversion accuracy of 99.08% and an RMSE of 7.51 m. When the noise increased to 5 mGal, the accuracy was 98.53% and the RMSE was 11.92 m. At 10 mGal, the accuracy remained 97.91%, the RMSE was 16.97 m, and the retention rate of inversion accuracy was 98.826%, indicating stronger noise robustness than conventional methods such as the admittance function method.

At different continuation depths, the random forest model achieved an inversion accuracy of 99.88% and an RMSE of 0.9962 m at a continuation depth of 800 m, 99.35% and 5.26 m at 1500 m, and 98.87% and 9.22 m at 2500 m. Over the full depth range of 800–2500 m, the retention rate of inversion accuracy was 98.988%, indicating strong stability.

The inversion results for the measured data from the Li and Zhao study areas show that the random forest model outperformed the conventional methods in both regions. In the Li study area, the random forest model achieved an RMSE of 556.82 m, an MAE of 456.22 m, an STD of 369.01 m, and an RMSE/MAXdepth ratio of 5.68%; the corresponding values for the conventional methods were 1024.22–1053.30 m, 916.30–965.52 m, and 10.44–10.74%, respectively. In the Zhao study area, the random forest model achieved an RMSE of 393.94 m, an MAE of 221.11 m, an STD of 347.15 m, and an RMSE/MAXdepth ratio of 5.07%; the corresponding values for the conventional methods were 589.68–623.75 m, 457.50–472.33 m, and 7.59–8.03%, respectively. Overall, the random forest model substantially reduced both the RMSE and MAE in both regions, with the MAE

decreasing by more than 50%, indicating more balanced suppression of small and medium-magnitude errors. Figures 4 and 5 also show that the random forest model better reconstructed typical landforms such as seamounts, basins, and steep topographic gradients, with a relatively uniform error distribution.

Compared with existing machine learning methods such as FC-DNN and ARPENN, the random forest model in this study relies only on a single gravity anomaly dataset, requires lower training cost, and produces a more uniform error distribution in topographically complex areas.

The multi-domain feature fusion random forest inversion method proposed in this study is superior to conventional inversion methods such as the admittance function method in terms of accuracy, stability, and practicality, providing a reference for high-accuracy bathymetry inversion. Future work will explore embedding physical constraints (such as the gravity admittance function and isostatic equilibrium model) into the loss function or the decision-making process of the random forest model to develop a physics-guided data-driven inversion framework. Systematic validation across multiple global regions and geological tectonic settings will also be conducted to evaluate the model's generalization ability and applicability in different marine environments.

References

Airy G B. On the computation of the effect of the attraction of mountain-masses, as disturbing the apparent astronomical latitude of stations in geodetic surveys[J]. *Philosophical transactions of the Royal Society of London*, 1855, 145: 101-104. https://doi.org/10.1098/rstl.1855.0003

Annan R F, Wan X. Refined bathymetric prediction based on feature extraction of gravity field signals: BATHY-FE[J]. *Journal of Geophysical Research: Machine Learning and Computation*, 2024, 1(3): e2024JH000205. https://doi.org/10.1029/2024JH000205

Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Bu Xianhai, Tan Xinyue, Zhang Jianxing, et al. Development of a high-precision deep-sea seabed terrain model through frequency domain weighted fusion of multi-source data: A case study in the southern waters of Greenland [J]. *Haiyang Xuebao*, 2025, 47(08): 101-115.

Cordell L, Grauch V J S. Mapping basement magnetization zones from aeromagnetic data in the San Juan Basin, New Mexico[M]//*The Utility of Regional Gravity and Magnetic Anomaly Maps*. Society of Exploration Geophysicists, 1985: 181-197. https://doi.org/10.1190/1.0931830346.ch16

Crosby A G. An assessment of the accuracy of admittance and coherence estimates using synthetic data[J]. *Geophysical Journal International*, 2007,171(1):25-54. https://doi.org/10.1111/j.1365-246X.2007.03520.x

Fan Diao. Research on the theory and method of bathymetry prediction combining satellite altimetry gravity data[J]. *Acta Geodaetica et Cartographica Sinica*, 2024, 53(1): 204-204.

Fan Meiqi, Chen Yilan, Sun Heyuan, et al. Inversion of seafloor topography using BP neural network based on geographical unit partition[J]. *Hydrographic Surveying and Charting*, 2024, 44(1): 16-20.

GEBCO Compilation Group. GEBCO 2025 Grid[DB/OL]. 2025. https://doi.org/10.5285/37c52e96-24ea-67ce-e063-7086abc05f29.

Harper H, Sandwell D T. Global predicted bathymetry using neural networks[J]. *Earth and Space Science*, 2024, 11(3): e2023EA003199. https://doi.org/10.1029/2023EA003199

Haxby W F, Karner G D, LaBrecque J L, et al. Digital images of combined oceanic and continental data sets and their use in tectonic studies[J]. *Eos, Transactions American Geophysical Union*, 1983, 64(52): 995-1004. https://doi.org/10.1029/EO064i052p00995

Karpatne A, Ebert-Uphoff I, Ravela S, et al. Machine learning for the geosciences: Challenges and opportunities[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(8): 1544-1554. https://doi.org/10.1109/TKDE.2018.2861006

Ke Baogui, Zhang Liming, Zhang Chuanyin, et al. Fusion the Altimetric and Shipborne Gravity Data Based on Point Mass Fit Method[J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(7): 924-929.

Kim K B, Kim J, Yun H S. Improved bathymetry estimation using satellite altimetry-derived gravity anomalies and machine learning in the East Sea[J]. *Journal of Marine Science and Engineering*, 2024, 12(9): 1520. https://doi.org/10.3390/jmse12091520

Kim Y, Nakata N. Geophysical inversion versus machine learning in inverse problems[J]. *The Leading Edge*, 2018, 37(12): 894-901. https://doi.org/10.1190/tle37120894.1

Li Q, Zhai Z, Li Q, et al. Improved bathymetry in the South China Sea from multisource gravity field elements using fully connected neural network[J]. *Journal of Marine Science and Engineering*, 2023, 11(7): 1345. https://doi.org/10.3390/jmse11071345

Liaw A, Wiener M. Classification and regression by randomForest[J]. *R News*, 2002, 2(3): 18-22.

Lin Wenqi, Wen Hanjiang, Liu Huanling, et al. Spherical wavelet multiscale signal analysis of regional gravity anomaly[J]. *Science of Surveying and Mapping*, 2022, 47(03): 9-15.

Little S A, Carter P H, Smith D K. Wavelet analysis of a bathymetric profile reveals anomalous crust[J]. *Geophysical research letters*, 1993, 20(18): 1915-1918. https://doi.org/10.1029/93GL01880

Liu W, Wang H, Xi Z, et al. Physics-informed deep learning inversion with application to noisy magnetotelluric measurements[J]. *Remote Sensing*, 2024, 16(1): 62. https://doi.org/10.3390/rs16010062

Martin J, Schaub H. The Physics-Informed Neural Network Gravity Model Generation III[J]. *The Journal of the Astronautical Sciences*, 2025, 72(2): 1-47. https://doi.org/10.1007/s40295-025-00480-z

Mayer L, Jakobsson M, Allen G, et al. The Nippon Foundation—GEBCO seabed 2030 project: The quest to see the world's oceans completely mapped by 2030[J]. *Geosciences*, 2018, 8(2): 63. https://doi.org/10.3390/geosciences8020063

McKenzie D, Bowin C. The relationship between bathymetry and gravity in the Atlantic Ocean[J]. *Journal of Geophysical Research*, 1976, 81(11): 1903-1915. https://doi.org/10.1029/JB081i011p01903

Moran N, Stringer B, Lin B, et al. Machine learning model selection for predicting bathymetry[J]. *Deep Sea Research Part I: Oceanographic Research Papers*, 2022, 185: 103788. https://doi.org/10.1016/j.dsr.2022.103788

Nettleton L L. Determination of density for reduction of gravimeter observations[J]. *Geophysics*, 1939, 4(3): 176-183. https://doi.org/10.1190/1.1437088

Parker R L. The rapid calculation of potential anomalies[J]. *Geophysical Journal International*, 1973, 31(4): 447-455. https://doi.org/10.1111/j.1365-246X.1973.tb06513.x

Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2019, 378: 686-707. https://doi.org/10.1016/j.jcp.2018.10.045

Sandwell D T, Müller R D, Smith W H F, et al. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure[J]. *Science*, 2014, 346(6205): 65-67. https://doi.org/10.1126/science.1258213

Sandwell D T, Harper H, Tozer B, et al. Gravity field recovery from geodetic altimeter missions[J]. *Advances in Space Research*, 2021, 68(2): 1059-1072. https://doi.org/10.1016/j.asr.2019.09.011

Smith W H F, Sandwell D T. Global seafloor topography from satellite altimetry and ship depth soundings[J]. *Science*, 1997, 277(5334): 1956-1962. https://doi.org/10.1126/science.277.5334.1956

Vening Meinesz F A. *A formula expressing the deflection of the plumb-line in the gravity anomalies and some formulae for the gravity-field and the gravity-potential outside the geoid*[M]. Koninklijke Nederlandsche Akademie van Wetenschappen, 1928.

Wan X, Annan R F, Ziggah Y Y. Altimetry-derived gravity gradients using spectral method and their performance in bathymetry inversion using back-propagation neural network[J]. *Journal of Geophysical Research: Solid Earth*, 2023, 128(2): e2022JB025785. https://doi.org/10.1029/2022JB025785

Watts A B. *Isostasy and Flexure of the Lithosphere*[M]. Cambridge University Press, 2001.

Wiener N. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*[M]. The MIT press, 1949. https://doi.org/10.7551/mitpress/2946.001.0001

Zhang J, Li J, Chen X, et al. Robust deep learning seismic inversion with a priori initial model constraint[J]. *Geophysical Journal International*, 2021, 225(3): 2001-2019. https://doi.org/10.1093/gji/ggab074

Zhang Xinjun, Yao Yuhu, Wang Kai, et al. 3D Morphology Inversion of Magnetic Basin Basement Based on Random Forest Regression Algorithm: A Case Study of Tianzhen-Yanggao Depression in Datong Basin[C]//Chinese Geophysical Society. *Proceedings of 2023 Annual Meeting of Chinese Geoscience Union—Special Topic 19 Frontiers and Advances in Environmental and Disaster Geophysical Monitoring Technologies and Equipment, Special Topic 20 Urban Subsurface Medium Imaging and Detection*. Beijing Botong Electronic Publishing Group, 2023: 24-27.

Zhao F, Xu Y, Zheng N, et al. ARPENN: an improved deep convolutional neural network for bathymetry inversion with integrated physical constraints[J]. *Geophysical Journal International*, 2025, 241(2): 891-900. https://doi.org/10.1093/gji/ggaf081

Zhou S, Liao B, Guo R, et al. Uncertainty analysis of bathymetry inversion in the South China Sea: a comparison of deep learning and Bayesian approaches[J]. *Geophysical Journal International*, 2026, 245(2): ggag093. https://doi.org/10.1093/gji/ggag093

Statements and Declarations

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 41874022), the National Precise Gravity Measurement Facility (Grant No. PGMF-2024-Z002), and the State Key Laboratory of Spatial Datum (Grant Nos. SKLGIE2024-ZZ-7 and SKLSD2026-KF-06).

Competing interests: The authors declare that they have no competing interests relevant to the content of this article.

Author contributions: Yuxuan TAN contributed to conceptualization, methodology, software, formal analysis and writing of the original draft. Baogui KE contributed to conceptualization, supervision, funding acquisition and review and editing of the manuscript. Shanbo LU contributed to data curation, validation and review and editing of the manuscript. All authors read and approved the final manuscript.

Data availability: The satellite altimetry-derived gravity anomaly grids used in this study (Sandwell et al., 2014, 2021) and the GEBCO 2025 grid (GEBCO Compilation Group, 2025) are publicly available. The shipborne bathymetric data used for validation were obtained from the study areas of Li et al. (2023) and Zhao et al. (2025).