University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

Improving ALOS Digital Elevation Model Using ICESat-2 Data and Random Forest

(Selected Paper in the 8th ISPRS Geospatial Conference 2025, University of Tehran, Iran)

Fatemeh Imanpour[1,4], Ramin Saadi Esfangareh[2], and Mohammad Reza Saradjian Maralan[3]✉

1. School of Surveying and Geospatial Engineering, Collage of Engineering, University of Tehran, Tehran, Iran.
   E-mail: Imanpour.fatemeh@ut.ac.ir
2. School of Surveying and Geospatial Engineering, Collage of Engineering, University of Tehran, Tehran, Iran.
   E-mail: ramin.saadi@ut.ac.ir
3. Corresponding author, School of Surveying and Geospatial Engineering, Collage of Engineering, University of Tehran, Tehran, Iran.
   E-mail: sarajian@ut.ac.ir
4. Image processing & Remote Sensing, National Cartographic Center (NCC), Tehran, Iran.

Article Info

Article type:
Research Article

Article history:
Received 2026-01-19
Received in revised form 2026-02-05
Accepted 2026-02-06
Available online 2026-05-12

Keywords:
Digital Elevation Model (DEM)
ICESat-2
ALOS
ATL08
Random Forest

ABSTRACT

Digital Elevation Models (DEMs) are fundamental analytical tools in environmental studies, urban development, natural change monitoring, and spatial analysis. These models are typically generated using remote sensing technologies, satellite imagery, and LiDAR measurements. ICESat-2, a state-of-the-art satellite developed by NASA, employs advanced LiDAR sensors and a unique orbital design to provide high-resolution elevation data. In this study, an integrated approach is presented to enhance the accuracy of ALOS-derived DEMs by incorporating ICESat-2 elevation data (ATL08 product) and the Random Forest machine learning algorithm. The research was conducted over Avajiq County in West Azerbaijan Province, using data from 2018 to 2023. ICESat-2 data were first used as a reference dataset. By extracting topographic features and applying the Random Forest model, the original DEM errors were corrected. Results indicate a significant improvement in elevation accuracy, with an RMSE of 0.0319, R^2 of 0.9727, and MAE of 0.0198. A comparison with elevation data from the National Cartographic Center* of Iran further validated the performance of the proposed method, where RMSE, R^2, and MAE were reported as 2.35, 0.98, and 1.05, respectively.

Cite this article: Imanpour, F., Saadi Esfangareh, R., & Saradjian Maralan, M. R. (2025). Improving ALOS Digital Elevation Model Using ICESat-2 Data and Random Forest. *Earth Observation and Geomatics Engineering*, Volume 9, Issue 1, Pages 137-143. http//doi.org/ 10.22059/eoge.2026.409807.1207

Publisher: University of Tehran.
DOI: http//doi.org/ 10.22059/eoge.2026.409807.1207

1. Introduction

1.1. General Instructions

Digital Elevation Models (DEMs) represent the Earth's surface in 2.5 dimensions and are essential tools in a wide range of disciplines such as disaster management, hydrology, forestry, environmental monitoring, and infrastructure development (Li et al., 2022). These models are typically generated from a variety of remote sensing technologies, each offering different levels of spatial resolution, coverage, and accuracy. Common techniques for DEM generation include photogrammetry based on stereo aerial or satellite images, which is cost-effective and widely used in mapping projects (Baltsavias, 1999). Another prevalent method is Light Detection and Ranging (LiDAR), especially airborne LiDAR, which provides highly accurate and dense elevation data but is limited by high operational costs and narrow coverage (AHN, 2023). Interferometric Synthetic Aperture Radar (InSAR) offers large-scale elevation mapping with the ability to operate in cloudy and night-time conditions; however, its accuracy is often affected by vegetation, water content, and temporal decorrelation (Bürgmann et al., 2000). The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2), launched by NASA in 2018, is equipped with the Advanced Topographic Laser Altimeter System (ATLAS), a photon-counting LiDAR system designed to capture high-resolution elevation data. This mission was developed to continue and enhance the capabilities of its predecessor, ICESat-1, which operated using a single laser beam with a footprint of approximately 65 meters and spacing of 172 meters between observations. In contrast, ICESat-2 transmits six beams arranged in three pairs—each consisting of a strong and a weak beam—with cross-track spacing of about 3 kilometers and intra-pair spacing of around 90 meters. The resulting footprint diameter is approximately 17 meters, with a fine along-track resolution of about 0.7 meters (Li et al., 2023; Pancholi et al., 2023). The satellite's orbit, inclined at 92 degrees, allows it to cover latitudes from 88° South to 88° North, ensuring near-global observations (Neumann et al., 2019). ICESat-2 data are widely used for measuring the elevation of both the Earth's surface and vegetation layers. Among its data products, ATL08 provides height measurements classified into ground and canopy returns. These data have been validated across various terrains, demonstrating Root Mean Square Error (RMSE) values between 0.5 and 2 meters depending on surface complexity (Neuenschwander et al., 2021). In recent years, machine learning algorithms have been increasingly utilized to enhance the quality and resolution of Digital Elevation Models (DEMs) derived from satellite data. Among them, Random Forest Regression (RFR) has demonstrated notable performance due to its ability to model complex nonlinear relationships and mitigate overfitting. Nevertheless, standard RFR models do not account for the spatial dependencies inherent in elevation data, often resulting in prediction errors. To address this, spatial extensions such as Random Forest for spatial prediction (RFsp) have been proposed, incorporating spatial coordinates and distances as predictor variables to retain spatial structure within the model (Hengl et al., 2018). When applied to ICESat-2 ATL08 elevation data, spatially-informed RFR methods have shown improved accuracy in estimating ground surface levels. RFsp has been found to produce results more comparable to geostatistical interpolation methods such as Inverse Distance Weighting (IDW) and Ordinary Kriging (OK), while also achieving lower Root Mean Square Error (RMSE) (da Silva Júnior et al., 2019). These findings suggest that combining ICESat-2's high-resolution elevation observations with machine learning frameworks like RFsp can significantly enhance DEM generation, particularly in complex terrain environments.

Building upon these advancements, this research focuses on enhancing the vertical accuracy of the ALOS Digital Elevation Model using ICESat-2 ATL08 elevation data in combination with the Random Forest Regression (RFR) method. Through the integration of spatial variables and remote sensing features, the study develops a refined DEM with improved performance across various terrains. The process follows a data fusion perspective, where precise point-based ICESat-2 measurements are synergistically combined with raster-based ALOS data. This integration addresses individual limitations of each dataset and supports the generation of high-resolution elevation models suitable for advanced geospatial analyses such as terrain modeling and land surface change detection.

1.2. Data sets and case study

ICESat-2, launched by NASA in 2018, was developed to acquire highly precise elevation measurements across Earth's polar regions, land surfaces, sea ice, and forest canopies. Operating in a near-polar orbit with a 92° inclination, the satellite ensures near-global coverage between latitudes 88°S and 88°N. Its core instrument, the Advanced Topographic Laser Altimeter System (ATLAS), continues and improves upon the legacy of ICESat-1 by using a multi-beam photon-counting LiDAR system, enabling finer spatial sampling and broader coverage. ATLAS emits six laser beams organized into three pairs, each comprising a strong and a weak beam. The beams are spaced 90 meters apart within each pair and 3.3 kilometers between pairs, resulting in a swath width of 6.6 kilometers (see Fig. 1). The system operates at a green wavelength (532 nm) and achieves a footprint diameter of approximately 11–17 meters, with a dense along-track sampling interval of 0.7 meters. Strong beams have pulse energies around 120 μJ, while weak beams deliver about 30 μJ. This high-resolution configuration, along with ATL08's capability to classify returns into ground, canopy, and noise, provides a robust basis for enhancing DEM accuracy using machine learning approaches (Abdalati et al., 2010; Neuenschwander et al., 2021; Pronk et al., 2024).

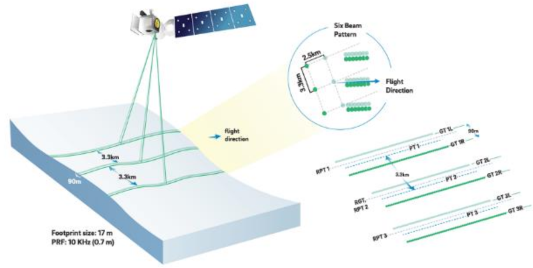

Figure 1. ICESat-2 mission beam pattern (Smith et al., 2019)

In this study, ATL08 (a key Level-3 product of ICESat-2) was utilized to extract elevation profiles over the Avajiq region in northwestern Iran. Data were retrieved using Google Earth Engine API via Google Colab, allowing for efficient spatial and temporal filtering. The extracted dataset includes geographic coordinates (latitude and longitude), interpolated terrain elevation (h_te_interp), best-estimated elevation (h_te_best), and the corresponding elevation uncertainty (h_te_uncertainty). These parameters serve as the basis for model training, spatial prediction, and accuracy assessment in subsequent stages.

The study area examined in this research is Avajiq County, located in West Azerbaijan Province, Iran, as shown in Fig. 2. As mentioned in the introduction, the primary objective of this study is to enhance the Digital Elevation Model using satellite-based ICESat-2 data. The satellite's ground track over the target area is illustrated in Fig. 3. The figure displays the ICESat-2 orbital path as it passes over a section of the Avajiq region in West Azerbaijan Province. The blue dots represent elevation data points collected by the ATL08 product, arranged in a linear and angled pattern. This configuration highlights the satellite's precise orbital design, with parallel tracks and consistent spacing, which ensures systematic and comprehensive coverage for topographic and geospatial analysis. The Digital Elevation Model provided by the NCC of Iran was used as a reference for accuracy assessment. The spatial coverage of both datasets is illustrated in Fig. 4.

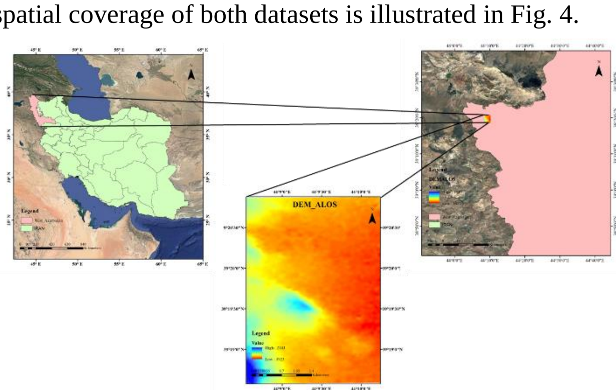

Figure 2. Study area

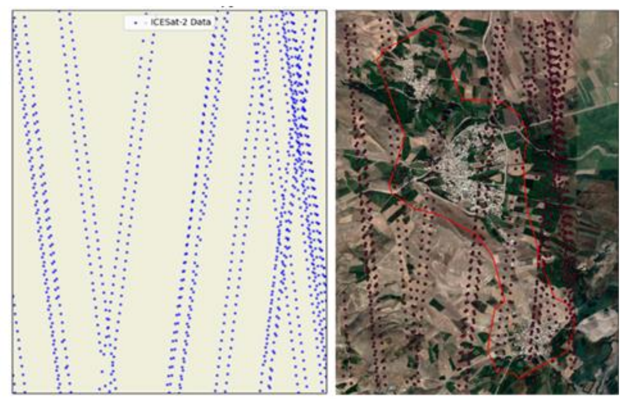

Figure 3. ICESat-2 data tracks over Avajiq County.

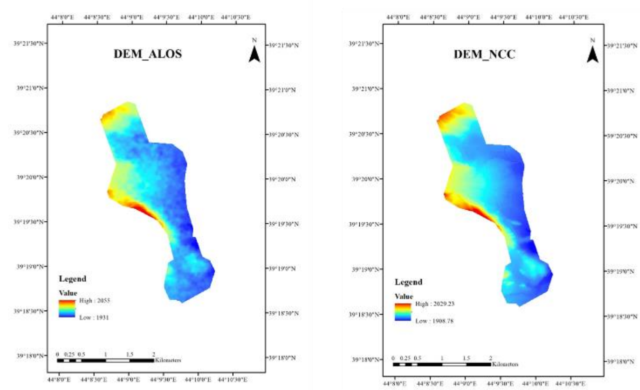

Figure 4. ALOS and NCC Digital Elevation Models.

2. Methodology

In this study, the Random Forest for Spatial Predictions (RFsp) method was employed to enhance the vertical accuracy of the ALOS Digital Elevation Model using ICESat-2 ATL08 elevation data. RFsp extends the traditional Random Forest algorithm by integrating spatial components—specifically geographic coordinates (latitude and longitude)—into the regression model, allowing for spatial interpolation and prediction without reliance on auxiliary covariates. Figure 5 presents a schematic overview of the research workflow, including data acquisition, preprocessing, model training, prediction, and validation.

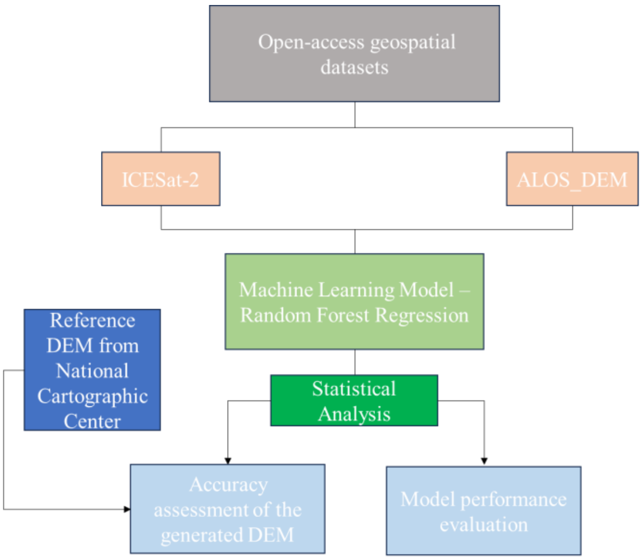

Figure 5. Methodology flowchart

2.1. Data Preprocessing

ICESat-2 ATL08 elevation points were initially filtered to remove outliers and low-confidence returns. The refined dataset served as the reference elevation values. Concurrently, ALOS DEM values were extracted at the ICESat-2 sample locations using bilinear interpolation to ensure pixel-level consistency across spatial layers.

2.1.1. Vertical Datum Harmonization

Three elevation datasets with distinct vertical reference systems were used in this study: ICESat-2 laser altimetry data, the ALOS DEM, and a high-resolution DEM from the National Cartographic Center (NCC). While the ICESat-2 and NCC DEM datasets are referenced to the WGS84 ellipsoidal height system, the ALOS DEM adopts a global orthometric height system based on EGM96 geoid model. To enable vertical harmonization, all elevations were converted to orthometric heights using a high-resolution geoid model developed for Iran.

This geoid model was generated using an RBF interpolation approach combining terrestrial gravity measurements and satellite graviometry. The geoid undulations were computed on a 2.5′ × 2.5′ grid and refined using a six-parameter polynomial fit to national GNSS/levelling benchmarks. Orthometric height for each point was then calculated using the standard transformation $H=h-N$, where H is orthometric height, h is ellipsoidal height, and N is geoid undulation.

2.2. Covariate Preparation

Unlike traditional RF models that incorporate diverse terrain or environmental predictors, the RFsp model in this study utilized only spatial coordinates (latitude and longitude) and ALOS DEM elevation values as input predictors. This minimalist design aimed to isolate the spatial patterns in elevation discrepancy and leverage the high vertical precision of ICESat-2, despite its sparse spatial coverage. The exclusion of ancillary covariates simplified the model and reduced potential overfitting, allowing a clearer assessment of how spatial location and topography alone influence the elevation discrepancies between ALOS DEM and ICESat-2 observations.

2.3. Model Implementation

The RFsp model was trained using the filtered ICESat-2 elevation values as target variables and their corresponding ALOS DEM values and spatial coordinates as inputs. This regression setup allowed the model to learn local elevation discrepancies across the landscape. The model was implemented with default Random Forest parameters (e.g., number of trees, node size), and hyperparameters were optimized via cross-validation.

2.4. Spatial Prediction and Output Generation

Once trained, the RFsp model was used to predict elevation corrections at each pixel of the ALOS DEM across the target region. These predictions were then applied to the original ALOS DEM, producing an enhanced DEM that integrates the precision of ICESat-2 with the full coverage of ALOS. The resulting surface represents a spatially continuous elevation model that benefits from localized ICESat-2 corrections.

2.5. Model Validation

To assess the performance of the enhanced DEM, a portion of the ICESat-2 ground points withheld from training was used as validation data. Statistical metrics including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were calculated by comparing predicted elevations to reference ICESat-2 values. Table 1 provides the equations for the accuracy metrics.

Table 1. Statistical metric used between predicted DEM and NCC DEM.

Metric	Mathematical Definition
R	$R = \frac{E\left[\left(DEM_{NCC} - E(DEM_{NCC})\right)\left(DEM_{pre} - E(DEM_{pre})\right)\right]}{\sigma_{NCC}\sigma_{pre}}$
RMSE	$RMSE = \sqrt{E\left[\left(DEM_{NCC} - DEM_{pre}\right)^2\right]}$
MAE	$MAE = E\left(\left\|DEM_{NCC} - DEM_{pre}\right\|\right)$

For training and testing the Random Forest model, a total of 64,512 test samples and 250 training samples were used, corresponding to a training-to-test ratio of approximately 0.38%. Details of the dataset distribution are presented in Table .2.

Table 2. Distribution of training and test data used in the modelling process

Number of Training Samples	Number of Test Samples	Training-to-Test Ratio
250	64,512	0.38%

3. Result

To evaluate the performance of the proposed method, three different datasets were analyzed over 50 randomly selected sample points. The elevation behavior of the various datasets is illustrated in Fig. 6.

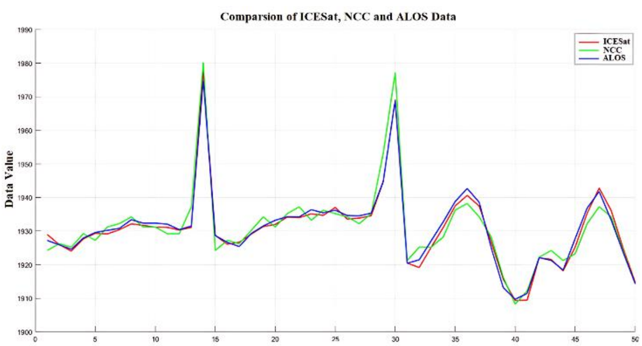

Figure 6. Elevation behaviour of three datasets: ALOS, ICESat-2, and the NCC DEM

Subsequently, the accuracy of each dataset was assessed in comparison with a reference DEM provided by the NCC, which has a reported vertical accuracy of 40 centimeters, as shown in Tables 3 and 4. Figures 7 and 8 show the correlation between ICESat-2 and ALOS elevation values with the reference DEM (NCC), respectively. The scatterplots include the regression equations and R^2 values, indicating a strong linear relationship between the datasets and the reference model.

Table 3. Accuracy assessment of ICESat-2 points compared to the reference Digital Elevation Model

R^2	RMSE	MAE
0.97	2.63	1.99

Table 4. Accuracy assessment of ALOS points compared to the reference Digital Elevation Model

R^2	RMSE	MAE
0.96	3.14	2.44

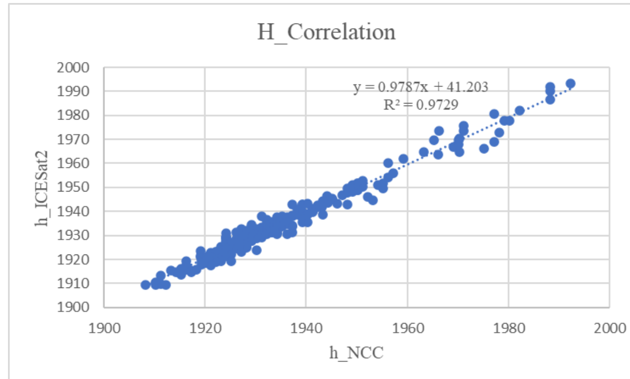

Figure 7. Correlation between ICESat-2 data and the elevation data from the NCC

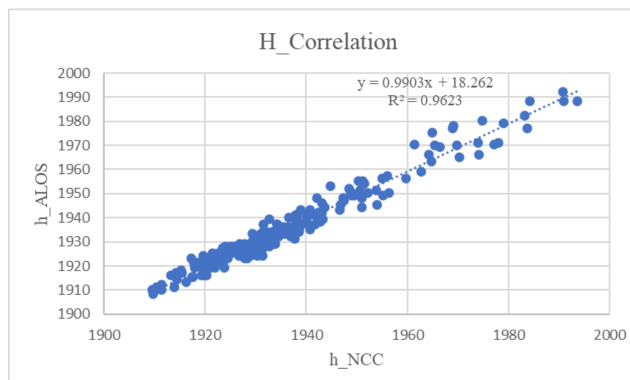

Figure 8. Correlation between ALOS data and the elevation data from the NCC

The modeling accuracy of the Random Forest algorithm and the performance of the proposed DEM, evaluated against reference points, are presented in Tables 5 and 6.

Table 5. Accuracy assessment of the Random Forest modelling

R^2	RMSE	MAE
0.97	0.03	0.01

Table 6. Accuracy assessment of the proposed DEM compared to the NCC DEM

R^2	RMSE	MAE
0.98	2.35	1.35

The final output of the Random Forest regression model is illustrated in Fig. 9, which presents the spatial distribution of the predicted Digital Elevation Model (DEM_PREDICTED) across the study area.

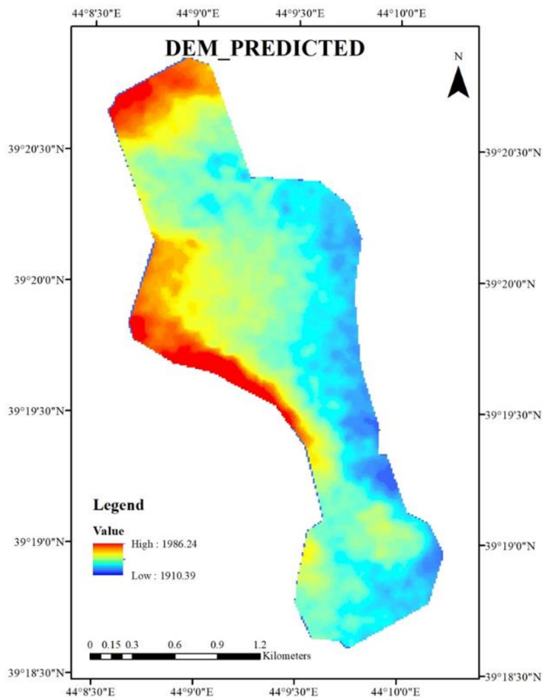

Figure 9. Digital Elevation Model predicted by the Random Forest regression model.

4. Conclusion

In this study, a method based on the Random Forest regression algorithm was proposed and evaluated to improve the accuracy of Digital Elevation Models (DEMs). By integrating high-precision ICESat-2 elevation data as reference points with ALOS DEM data and incorporating geometric and spatial features, a correction model was developed to reduce elevation errors in the original DEM. The results demonstrated a notable improvement in accuracy, as the RMSE of the ALOS DEM was reduced from 3.14 m to 2.35 m after applying the proposed approach.

The Random Forest model demonstrated high flexibility and effectiveness in handling topographic complexity and data heterogeneity, providing a statistically robust solution for DEM enhancement. These findings highlight the potential of machine learning–based data fusion techniques in remote sensing, particularly in regions where large-scale DEM products require accuracy refinement using limited but high-precision reference datasets.

The proposed method effectively leveraged the dense vertical accuracy of ICESat-2 photon-based laser altimetry to correct systematic errors present in the ALOS DEM. As a result, the generated enhanced DEMs are suitable for a wide range of geospatial applications, including topographic mapping, land surface monitoring, hydrological modeling, and natural hazard assessment.

This research demonstrates the applicability of Random Forest regression for DEM correction across varying terrain types and confirms the feasibility of integrating heterogeneous datasets such as ICESat-2 and ALOS through statistical learning. Moreover, the proposed framework is replicable and relies on freely available remote sensing data, making it a practical solution for large-area DEM improvement.

Despite these strengths, several limitations were identified. The accuracy of the model is strongly dependent on the quality and spatial distribution of the training data. In addition, noise in the input datasets and the applied pre-processing techniques can significantly influence model performance, particularly in areas with complex terrain.

Future research should focus on incorporating additional environmental and geospatial variables, such as land cover, climatic, and hydrological data, to further improve model performance. Testing the robustness of the proposed method across diverse geographic regions with different elevation characteristics is also recommended. Furthermore, performance comparisons with advanced algorithms, including XGBoost and deep learning models, should be explored. Extending the methodology to data from other satellite missions such as GEDI, TanDEM-X, and Sentinel-1, as well as improving the pre-processing pipeline to reduce noise and enhance data normalization, would further strengthen the approach.

In summary, this study confirms that integrating ICESat-2 laser altimetry with machine learning techniques can significantly enhance existing DEM products and provides a reliable methodology for generating high-quality elevation datasets to support environmental and geospatial analyses.

Reference

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., ... & Webb, C. (2010). The ICESat-2 laser altimetry mission. Proceedings of the IEEE, 98(5), 735-751. https://doi.org/10.1109/JPROC.2009.2034765

AHN, 2023. Actueel Hoogtebestand Nederland (AHN). Available at: https://www.ahn.nl (accessed 4 July 2025).

Baltsavias, E. P. (1999). Airborne laser scanning: basic relations and formulas. ISPRS Journal of photogrammetry and remote sensing, 54(2-3), 199-214. https://doi.org/10.1016/S0924-2716(99)00015-5

Bürgmann, R., Rosen, P. A., & Fielding, E. J. (2000). Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. Annual review of earth and planetary sciences, 28(1), 169-209. https://doi.org/10.1146/annurev.earth.28.1.169

da Silva Júnior, J. C., Medeiros, V., Garrozi, C., Montenegro, A., & Gonçalves, G. E. (2019). Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. Computers and electronics in agriculture, 166, 105017. https://doi.org/10.1016/j.compag.2019.105017

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6, e5518. https://doi.org/10.7717/peerj.5518

Li, B., Xie, H., Tong, X., Tang, H., & Liu, S. (2023). A global-scale DEM elevation correction model using ICESat-2 laser altimetry data. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-15. https://doi.org/10.1109/TGRS.2023.3321956

Li, Y., Fu, H., Zhu, J., Wu, K., Yang, P., Wang, L., & Gao, S. (2022). A method for SRTM DEM elevation error correction in forested areas using ICESat-2 data and vegetation classification data. Remote Sensing, 14(14), 3380. https://doi.org/10.3390/rs14143380

Liu, A., Cheng, X., & Chen, Z. (2021). Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. Remote Sensing of Environment, 264, 112571. https://doi.org/10.1016/j.rse.2021.112571

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., ... & Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. Remote sensing of environment, 190, 260-273. https://doi.org/10.1016/j.rse.2016.12.029

Neuenschwander, A., Popescu, S., Nelson, R., Harding, D., Pitts, K., & Robbins, J. (2019). ATLAS/ICESat-2 L3A Land and Vegetation Height, Version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center (DAAC) data set, ATL08-001. https://ui.adsabs.harvard.edu/link_gateway/2019nsid.data.1250N/doi:10.5067/ATLAS/ATL08.001

Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., ... & Thomas, T. C. (2019). The Ice, Cloud, and Land Elevation Satellite–2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. Remote sensing of environment, 233, 111325. https://doi.org/10.1016/j.rse.2019.111325

Pancholi, S., Abhinav, A., Maithani, S., & Bhardwaj, A. (2023). Simulation of DEM Based on ICESat-2 Data Using Openly Accessible Topographic Datasets. Environmental Sciences Proceedings, 29(1), 66. https://doi.org/10.3390/ECRS2023-16189

Pronk, M., Eleveld, M., & Ledoux, H. (2024). Assessing vertical accuracy and spatial coverage of ICESat-2 and GEDI spaceborne lidar for creating global terrain models. Remote Sensing, 16(13), 2259. https://doi.org/10.3390/rs16132259

Smith, B., Fricker, H. A., Holschuh, N., Gardner, A. S., Adusumilli, S., Brunt, K. M., ... & Siegfried, M. R. (2019). Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter. Remote Sensing of Environment, 233, 111352. https://doi.org/10.1016/j.rse.2019.111352