

Comment

Comment on Guth et al. Benchmarking Elevation Plus Land Surface Parameters Finds FathomDEM and Copernicus DEM Win as Best Global DEMs. *Remote Sens.* 2025, 17, 3919

Yu-Feng Ho , Qin Xu , Octavian Puiu Chiriac , Mustafa Serkan Isik  and Tomislav Hengl 

OpenGeoHub Foundation, 6865 HK Doorwerth, The Netherlands; qin.xu@opengeohub.org (Q.X.); octavian.chiriac@opengeohub.org (O.P.C.); serkan.isik@opengeohub.org (M.S.I.); tom.hengl@opengeohub.org (T.H.)

* Correspondence: yu-feng.ho@opengeohub.org

Abstract

A formal comment on a current Global Digital Elevation Model (GDEM) benchmark by Guth et al. is presented. It reproduces the original study, emphasizing and modifying the methodologies in two key aspects: (1) adding statistical testing and assessment, and (2) making the processing workflow (code) and test site tiles (data) publicly available and hence reusable. To test whether the performance of different GDEMs is statistically significant, the Kruskal–Wallis test was used. In addition, a pixel-alignment strategy to resolve pixel-shifting issues commonly observed in GDEMs was introduced. The results align with the research outcomes in Guth et al. regarding ranking FathomDEM, a restricted license product, first. However, a notable discrepancy was observed in the performance evaluation of COPDEM and GEDTM v1.2 within barren areas. Contrary to the findings of Guth et al., these results indicate no degradation in the performance of GEDTM v1.2 versus COPDEM. Furthermore, statistical comparison of the quality of GDEMs across various land surface parameters indicates no significant differences among GDEMs. The results of Guth et al. are potentially affected by the high sensitivity of the proposed FUV approach and the limited sample size (representing parts of the USA and western Europe). All code used to produce the results and all input datasets analyzed are provided, allowing for full reusability and reproducibility of these results.

Keywords: accuracy assessment; global data; geomorphometry; digital terrain model; statistical testing



Academic Editor: Prasad S. Thenkabail

Received: 3 March 2026

Revised: 18 May 2026

Accepted: 2 July 2026

Published: 17 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The Digital Elevation Model (DEM) is an essential data source for biogeography, hydrological monitoring, soil and biological surveys, predictive mapping of soil and vegetation, and soil erosion modeling [1–4]. With JAXA releasing ALOS AW3D30 (ALOS) [5] and ESA/European Commission releasing Copernicus DEM (COPDEM) [6] under open data licenses, the doors have opened for a wide range of applications of geomorphometry, hydrological modeling, and similar techniques. Several teams around the world have recently focused on producing 1-arcsec (~30 m spatial resolution) DTMs (vegetation and buildings removed) with full global terrestrial coverage [7–9]. Given their widespread application, continuous quality assessment is essential to ensure the high usability and proper use of Global DEM products.

Guth et al. [10] recently published a comprehensive benchmarking study of GDEMs, which builds on established evaluation frameworks and utilizes the Fraction of Unexplained Variance (FUV) as a primary metric [11–13]. Their study highlights FathomDEM and COPDEM as leading models based on elevation accuracy and derived terrain variables; GEDTM v1.2 (terrain model) is evaluated as having poor performance and being even worse than COPDEM (surface model). Although this work contributes significantly to the field, several methodological aspects warrant further examination to ensure reproducibility and global applicability. We identify four primary areas of concern:

1. **Concern 1: Limited reproducibility and transparency of reported tests.** Although the analysis utilized open-source software (MICRODEM), the specific implementation, including code for cropping, reprojection, data aggregation, and analytical operations, was not explicitly provided in the publication [10]. Furthermore, the available source code lacks a clear operational sequence, making it difficult to interpret and reuse. Additionally, the code relies on the Delphi IDE, which is restricted to the Windows operating system. This limitation creates a significant barrier, excluding a substantial portion of the open-source developer community.
2. **Concern 2: High sensitivity and subjectivity of the selected metric.** The previous evaluation utilized qualitative descriptors that may not fully reflect the underlying data. Furthermore, the FUV metric appears sensitive to processing parameters; however, the study omitted a formal sensitivity analysis and uncertainty quantification. In the absence of a rigorous statistical framework, the performance of the reported model remains susceptible to subjective interpretation and can lead to an overestimation of the accuracy of the model.
3. **Concern 3: Error in the technical description of GDEMs.** Certain technical descriptions in Guth et al. [10] regarding GEDTM v1.2 warrant correction. Specifically, the resolution of GEDTM v1.2 is 1 arcsec, and its production logic is a fusion of ALOS and COPDEM rather than the purely radar-based approach incorrectly listed in Table 3 of Guth et al. [10]. Clarifying these specifications is necessary to accurately interpret the fusion logic and the performance of the model.
4. **Concern 4: Geographical over-representation of Europe/America.** The validation dataset relies heavily on airborne lidar from Europe and North America, representing only ~23–24% of global landmass. This geographic concentration limits the generalizability of the results to global contexts.

To try to address some of these concerns and cross-check results of Guth et al. [10], we have prepared a reproducible pipeline using independent validation datasets (national DTMs). This methodology directly addresses Concerns 1 and 2. Concern 3 is addressed directly by the updated GEDTM v1.2 documentation [14]. Finally, regarding Concern 4, the issue could be mitigated if Guth et al. [10] explicitly acknowledged the geographical and potential land cover biases inherent in its validation dataset. Consequently, extrapolating global model performance from such localized coverage warrants significant caution. Although our results (to address Concerns 1 and 2) partially align with the finding that FathomDEM performs well, our analysis diverges regarding the relative performance of GEDTM v1.2 compared to COPDEM and ALOS. We discuss possible reasons for this discrepancy and provide access to all of our code and datasets used. Finally, we propose a statistical framework for assessing the FUV metric to provide a more robust basis for GDEM benchmarking to support conclusions based on unbiased geographical representation and statistical significance.

2. Materials and Methods

2.1. Data Collection and Pre-Processing

The reference dataset consists of high-resolution national DTMs sourced from the Netherlands, Switzerland, the United States, and Australia. These are compared against four Global DEMs (GDEMs): COPDEM, ALOS, GEDTM30, and FathomDEM (see Table 1 for details). To maintain spatial consistency, national Digital Elevation Models (DEMs) were reprojected to a common vertical datum (EGM2008, EPSG:3855). Simultaneously, the data were aggregated to a 1-arcsecond grid (EPSG:4326) using average resampling via the Geospatial Data Abstraction Library (GDAL). COPDEM and ALOS increase the longitude grid spacing from 1-arcsec starting from 50°N/S and 60°N/S up to 10-arcsec and 6-arcsec respectively [5,6]. We implemented cubic resampling to create 1-arcsec tiles. Cubic resampling produced reasonable results as it preserves sharper terrain features than the bilinear one and prevents over-smoothing the surface. FathomDEM was downloaded in 1-arcsec squared tiles from Zenodo [8,15], which does not require resampling. The resampled tiles were mosaicked into Cloud-Optimized GeoTIFFs (COGs) and hosted on the OpenGeoHub S3 server. GEDTM30 is already available on the OpenGeoHub S3 server as a single 1-arcsec endpoint. To facilitate complete reproducibility and ensure reusability, the download and mosaicking processes for the input data are documented, and the metadata and endpoint locations for all processed DTMs are defined in a YAML configuration file within the project repository (<https://codeberg.org/openlandmap/GEDTM30/src/branch/main/benchmark>, accessed on 1 July 2026).

Table 1. Characteristics of Global and National DEMs based on the project coverage database.

DEM	Coverage	Vertical Datum	Horizontal Datum	Ref.
COPDEM v2023-1	Global	EGM2008	EPSG:4326	[6]
AW3D30 v4.1	Global	EGM96	EPSG:4326	[5]
FathomDEM v1-0	Global	EGM2008	EPSG:4326	[8,15,16]
GEDTM v1.2	Global	EGM2008	EPSG:4326	[7,14]
AHN4—DTM	the Netherlands	EPSG:5730	EPSG:28992	[17]
swissALTI3D	Switzerland	EPSG:5728	EPSG:2056	[18]
USGS DTM	USA	EPSG:5703	NAD83 UTM	[19]
Australia	Coastal Zone ¹	EPSG:5711	GDA94/MGA	[20]

¹ Refers specifically to Australia's populated coastal zone.

2.2. Modified Benchmarking Pipeline

The original source code for the benchmark study by Guth et al. [10] is hosted at MICRODEM (https://github.com/prof-pguth/git_microdem, accessed on 1 July 2026). However, as the original Pascal compilation is constrained to modern Windows 64-bit operating systems, we then implemented an independent, reproducible pipeline that follows their core conceptual framework while introducing modifications to enhance generalizability. The workflow follows these six (6) generic steps:

- (1) Retrieval of 10 × 10 km tiles from the benchmark research dataset [21];
- (2) Selection of tiles that overlap with the reference DTMs available (in total, 36 tiles in various landscapes);
- (3) Extraction and reprojection of all GDEMs, reference DTMs, and ESA Land Cover data [22] to the Equi7 grid [23] at 30 m resolution using mean resampling, noting that ESA Land Cover data is resampled by using mode;
- (4) Calculation of land surface parameters (slope, tangential curvature, and plan curvature) using lsp-calculator v1.2.0 and WhiteboxTools v2.4.0 (Table 2) [24,25];
- (5) Derivation of tile-level statistics;

- (6) Derivation of the Fraction of Unexplained Variance (FUV) ($1 - r^2$) and normalized Mean Absolute Error (nMAE) for each GDEM and terrain parameter against the reference data. FUV is described by Guth et al. [10]. The nMAE is computed by first calculating the Mean Absolute Error over all valid paired pixel values and then dividing it by the standard deviation of the corresponding reference data. Thus, the normalization factor is the standard deviation of the reference dataset after excluding nodata values:

$$\text{nMAE} = \frac{\text{MAE}}{\sigma_x} = \frac{\frac{1}{N} \sum_{i=1}^N |x_i - y_i|}{\sigma_x}, \quad (1)$$

where x_i denotes the reference values, y_i denotes the comparison values, N is the number of valid paired observations, and σ_x is the standard deviation of the valid reference values.

Table 2. Summary of elevation and derived terrain parameters, including processing window sizes and software tools used for computation.

ID	Parameter Name	Window	Program
ELEV	Elevation		
SLOPE	Slope in degrees	5×5	lsp-calculator
TANGC	Tangential curvature	5×5	lsp-calculator
PLANC	Plan curvature	5×5	WhiteboxTools

FUV metrics are highly sensitive to sub-pixel misalignment, a phenomenon noted in Guth et al. [10]. To minimize the impact of one-pixel horizontal/vertical shifting on accuracy metrics, Guth et al. [10] carefully resampled the reference DTM to each GDEM considering the GDEMs' geometric models. However, spatial mismatches are unavoidable during the cropping of a global mosaic into 1-degree $\times$ 1-degree tiles, because a one-degree span cannot be perfectly divided into a 1-arcsecond grid. Guth et al. [10] created tiles that use the GDEMs' bounding boxes, and then resampled the high-resolution reference DTM to match each GDEM. We instead implemented a 9-point moving window alignment at Step 6. By increasing the extraction window by one pixel (e.g., from 100×100 to 101×101), we tested nine possible pixel alignments (top-left, top-center, top-right, etc.). The pipeline automatically selects the alignment that yields the minimum FUV. This ensures that the reported accuracy reflects the model's vertical performance rather than artifacts of coordinate transformation or grid-centering discrepancies.

2.3. Statistical Inference and Significance Testing

Previous benchmarking efforts relied on subjective estimates of tolerance to determine whether differences between DEMs were significant [10]. To respond to Concern 2, we provide a more objective evaluation. We transition from subjective thresholds to a formal statistical framework. Since the FUV (calculated as $1 - R^2$) does not necessarily follow a normal distribution, we employed the Kruskal–Wallis H-test, a non-parametric method, to determine if the observed differences between the four GDEMs are statistically significant [26]. This approach allows for a data-driven determination of ties or winners in model performance, reducing the reliance on internal heuristic judgments. In order to conduct a Kruskal–Wallis H-test, we altered the scoring system from Guth et al. [10], which relies on the frequency of “first-place” rankings (lowest FUV), to a Rank-Sum Scoring System. In this framework, performance is weighted across the entire cohort: the top-performing DEM is awarded 3 points, with subsequent ranks receiving 2, 1, and 0 points.

2.4. Reproducible Development Environment

The entire pipeline is organized as a self-contained extension of the GEDTM30 project (<https://codeberg.org/openlandmap/GEDTM30/src/branch/main/benchmark>, accessed on 1 July 2026). The code is hosted on Codeberg; Python 3.12.0 is used for data orchestration and statistical analysis, as well as the generation of terrain parameters via whitebox-workflows [25], whereas terrain parameters generated by lsp-calculator are employed in a Bash script [24]. To ensure full reproducibility, we have also archived the produced code and data on Zenodo with a unique DOI (<https://zenodo.org/records/18836249>, accessed on 1 July 2026).

3. Results

3.1. Impact of Pixel Shifting on DEM Accuracy

To investigate whether pixel shifting contributes to accuracy uncertainty during the analysis, we evaluated all four target DEMs both with and without a 9-point moving window. As shown in Figure 1, applying the 9-point moving window yields lower FUV values by mitigating errors caused by pixel shifting during resampling. Notably, the FUVs for slope and tangential curvature exhibit greater sensitivity to a one-pixel shift than those for plan curvature and elevation. In this comparison, the control group (C, dark shades) extracts pixels based on exact tile alignment, whereas the experimental group (E, light shades) selects the optimal FUV from the 9-point neighborhood to account for any potential single-pixel misalignment.

3.2. Concordance with Existing Benchmarks

Our initial analysis largely replicates the general performance trends reported by Guth et al. [10]. Figure 2 illustrates the relationship between tile-wise FUV and empirical cumulative distribution functions (ECDFs) of all four GDEMs tested. Consistent with previous findings, FathomDEM frequently occupies the lower-FUV spectrum, showing the same tendency as reported by Guth et al. [10].

3.3. Evaluation in Barren Terrain: DSMs Versus GEDTM v1.2

A critical point of contention in the recent literature is the performance of GEDTM30 in unvegetated and non-urban areas. Although Guth et al. [10] reported that GEDTM30 performs poorly compared to COPDEM in a barren Nevada tile, our analysis of a comparable tile in Joshua Tree (ID: N33YW116D_REC3575444; 95% barren cover) yields different results. Figure 3 compares elevation and derived terrain parameters (slope, tangential, and plan curvature from left to right) for ALOS, COPDEM, GEDTM30 (version 1.2) and FathomDEM (from top to bottom). The calculated FUVs for GEDTM v1.2 were 0.0003, 0.033, 0.217 and 0.869 for elevation and terrain variables, respectively, while COPDEM recorded 0.0004, 0.0509, 0.308, and 0.881. Quantitatively, these metrics indicate that there is no significant visual performance gap between the models in high-aridity, barren environments, suggesting that the “poor performance” previously reported may not be globally representative of barren terrains. In addition, we computed nMAE, which shows that GEDTM v1.2 achieves accuracy between that of ALOS and COPDEM.

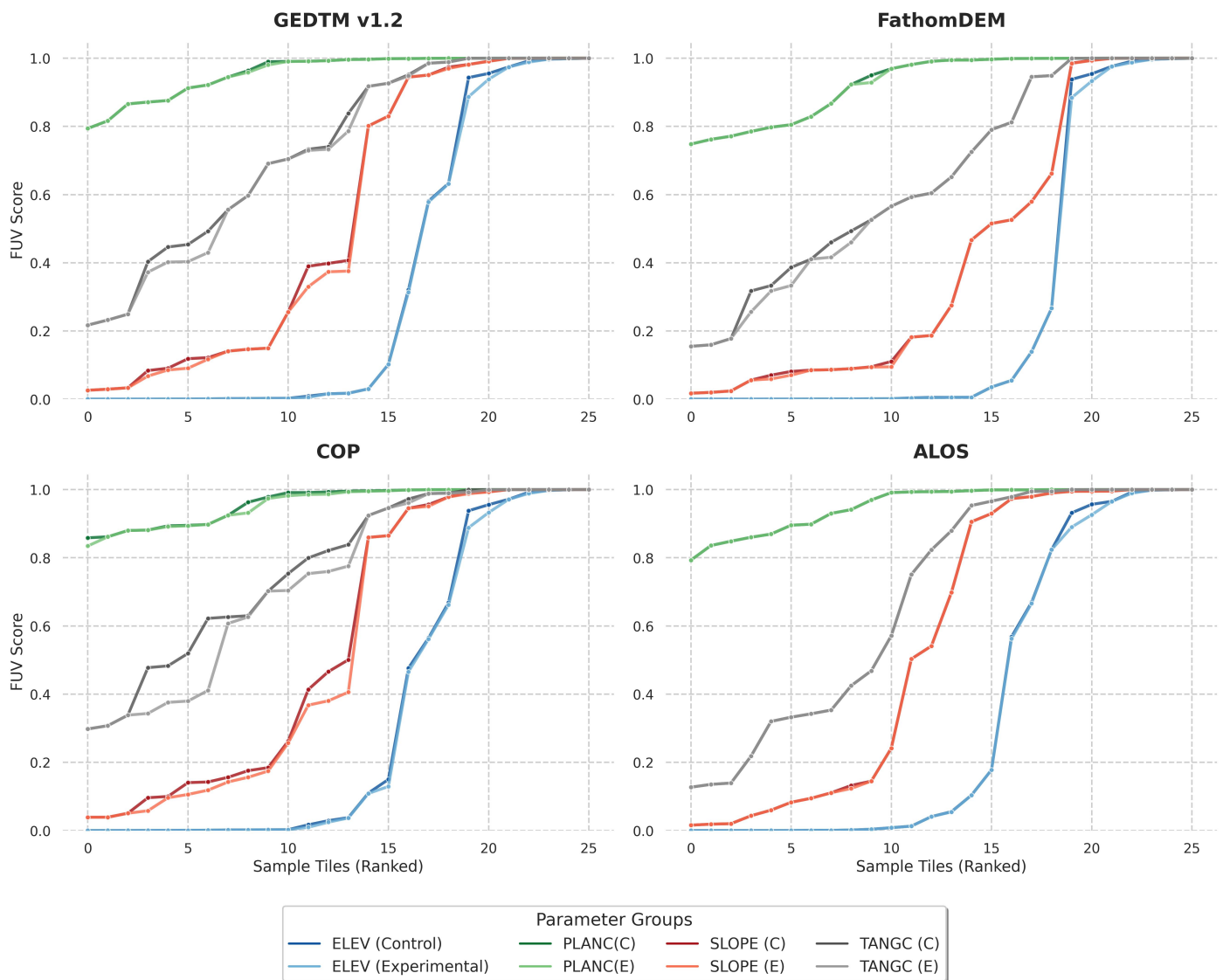


Figure 1. Comparison of FUV scores across four Global DEMs: GEDTM v1.2, FathomDEM, COPDEM, and ALOS AW3D30. Profiles are categorized into four target parameter groups: elevation (blue), plan curvature (green), slope (red), and tangential curvature (grey). Sample tiles are ranked along the x-axis by ascending FUV values. The analysis contrasts the control (C, dark shades), representing unaligned data, against the experimental (E, light shades), which utilizes a 9-point moving window alignment. Lower FUV values indicate better results.

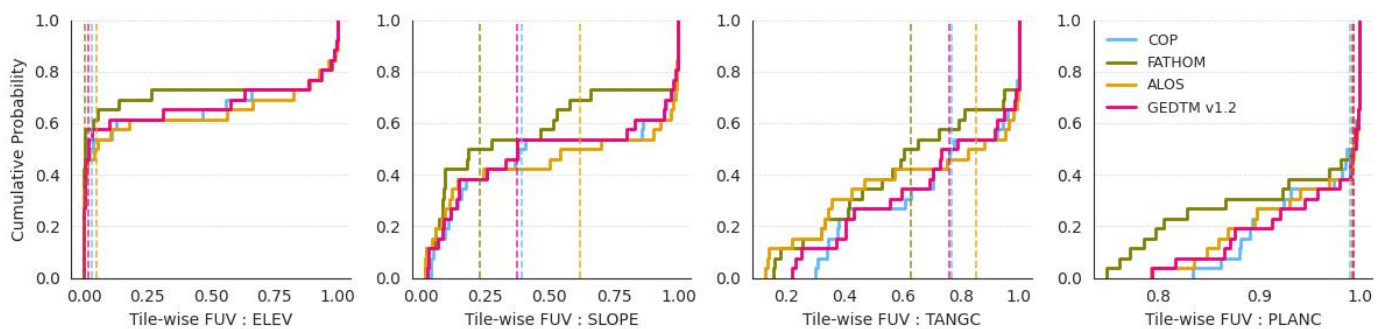


Figure 2. Empirical cumulative distribution functions (ECDFs) of tile-wise FUV for the four parameters (ELEV, SLOPE, TANGC and PLANC), with dashed lines denoting median values.

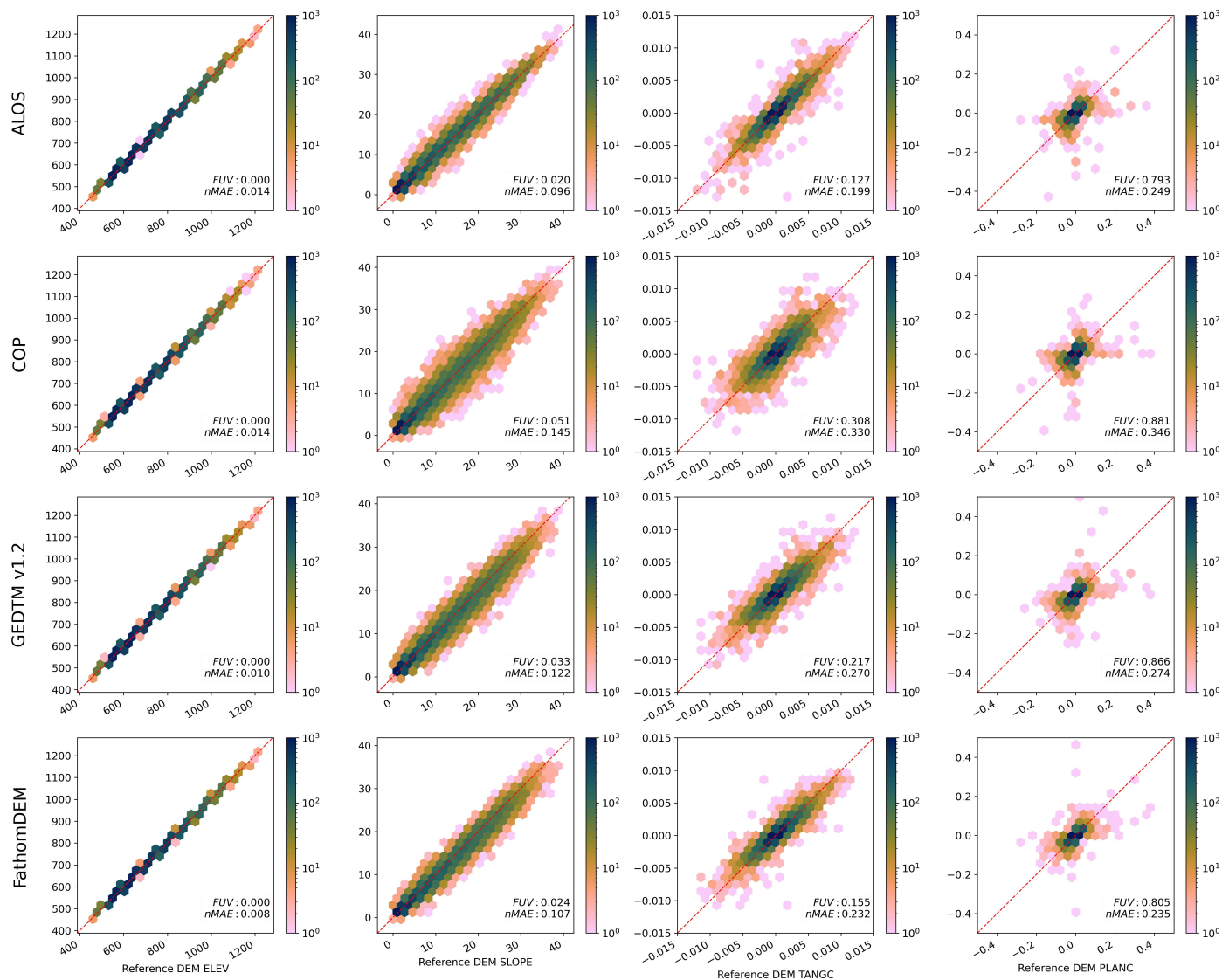


Figure 3. Accuracy assessment of COPDEM, ALOS, GEDTM v1.2 and FathomDEM at a barren tile in Joshua Tree (ID: N33YW116D_REC3575444), comparing elevation, slope, tangential curvature and plan curvature against reference DTM and its derivatives. The results show that GEDTM v1.2 performed similarly to other GDEMs, in contrast to the findings of Guth et al. [10].

3.4. Comparative Ranking and Frequency Analysis

Guth et al. [10] evaluated GDEMs using a scoring system based on the frequency of “first-place” rankings (lowest FUV) per tile and parameter. It suggested that among COPDEM, ALOS, and GEDTM v1.2, the superior general-purpose model was COPDEM, and ALOS excelled in high-slope areas. In our revised scoring system, the results (Figure 4) show the average ranking scores of GEDTM v1.2, FathomDEM, COPDEM and ALOS. Among the target GDEMs, FathomDEM achieves the high ranking score in the four parameters, and GEDTM v1.2 performs better in elevation, slope and plan curvature among the three unrestricted (open data) GDEMs; however, it stays between COP and ALOS in tangential curvature within our test sites.

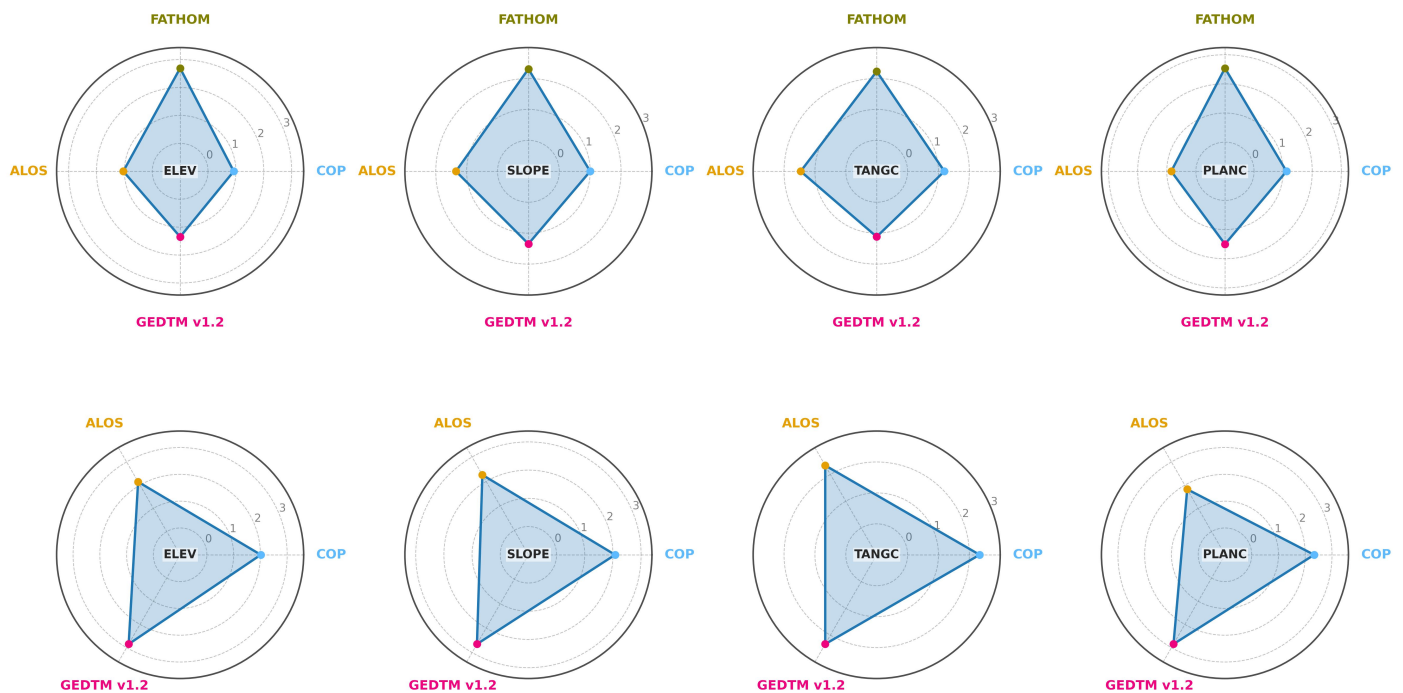


Figure 4. Average ranking score in comparison of FUV per tile across ELEV, SLOPE, TANGC and PLANC. The upper plot is the comparison among the four target GDEMs. The lower plot shows the comparison among the three unrestricted (open data) GDEMs. The closer the dataset accuracy metric to the edge of the circle, the better the performance.

3.5. Statistical Significance of Performance Gaps

We next ran the Kruskal–Wallis test on the score value (Figure 5). This shows no statistically significant difference in the distribution of FUV values among the four models. The rank distributions yielded the following p -values: elevation: $p = 0.832$, slope: $p = 0.789$, tangential curvature: $p = 0.895$, plan curvature: $p = 0.707$. Although FathomDEM may show a lower median FUV for elevation, slope, and tangential curvature, the overlap in performance distributions suggests that claiming one model as definitively superior to or poorer than the others lacks a robust statistical basis. This highlights the need to use formal inference rather than heuristic ranking systems in GDEM benchmarking.

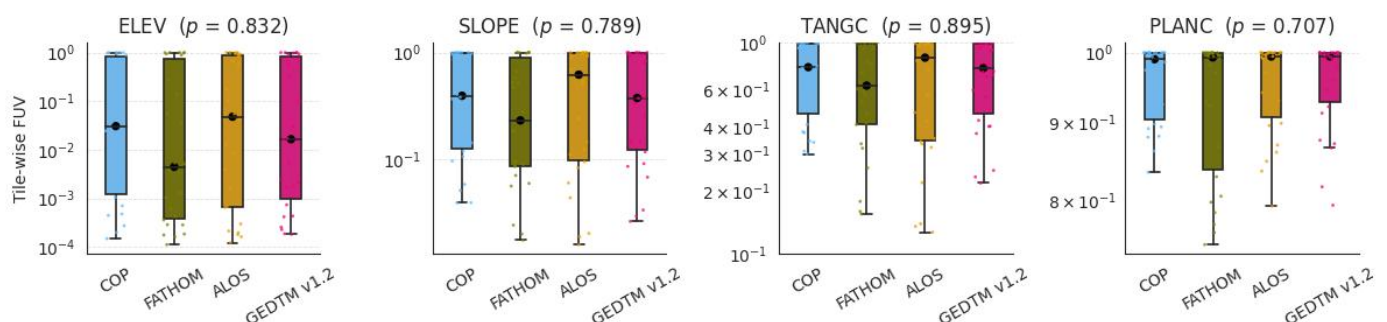


Figure 5. Distributions of FUV per tile for ELEV, SLOPE, TANGC and PLANC. Boxplots summarize the median and interquartile variability in explanatory power. The p -value indicates the Kruskal–Wallis H-test on FUVs of tiles. This indicates a high overlap in the accuracy metric and marginal statistical difference.

4. Discussion

4.1. Machine Learning Optimization and Its Impact on High-Order LSP Accuracy

Random Forest (used to construct ensemble GEDTM30) is considered a black-box model [27], in which the prediction objective is explicitly defined, but the ensemble-based decision-making process underlying individual predictions remains difficult to interpret. The observed performance of GEDTM30, which consistently ranks between ALOS and COPDEM in such high-order LSPs (e.g., plan curvature, tangential curvature), is consistent with that underlying architectural logic. GEDTM30 has been trained with a limited number of covariates across canopy and building height and extent models, long-term Landsat-derived indices, and coarse topographic derivatives. The Random Forest model is made up of 100 global trees and 50 local trees and is optimized to ensure minimal error between the predicted values and the training data from GEDI and ICESat-2 [7].

Because the GEDTM30 objective function primarily minimizes vertical elevation error rather than optimizing for derivative terrain variables (slope or curvature), the resulting LSPs, especially high-order ones tend to represent a numerical compromise between its primary inputs, COP and ALOS. This finding relates to a misconception in Guth et al. [10], who misclassified GEDTM as purely radar-based, whereas it is actually derived from fused radar and optical data (respond to Concern 3). This could explain why GEDTM30 may not frequently secure the “first-place” rank in frequency-based scoring systems [10], i.e., this is expected behavior for any ensemble estimate. Therefore, these findings suggest that, while fusion models effectively mitigate outliers in primary elevation data, they may inherit topographic characteristics from their constituent datasets.

4.2. A Reproducible and Decentralized Framework

This study introduces an open-source pipeline designed to inherit and standardize the benchmarking methodologies developed by DEMIX from the geomorphometry community [28]. By providing a transparent script-based workflow, we move toward a decentralized evaluation framework. This pipeline is not intended to replace existing benchmarking efforts, but rather to serve as a common asset for the community. This transparency is essential to resolve discrepancies in model performance reports and ensures that any researcher can verify the accuracy of a GDEM using locally relevant independent reference data. The inclusion of a 9-point alignment check specifically addresses the sensitivity of FUV metrics to coordinate transformations, providing a more robust technical standard for future comparisons.

4.3. Limitations and Future Outlook

A primary limitation of this research is the sample size of the reference DTM tiles, which is significantly smaller than that employed by Guth et al. [10]. Although the observed trends are aligned with the existing literature, the lower tile count may reduce the statistical power of the Kruskal–Wallis tests. Furthermore, the geographic distribution of reference data remains heavily skewed toward North America, Europe, and Australia, leaving a significant gap in the validation of GDEMs in tropical, densely vegetated, and developing-world contexts.

The modular and open-source nature of our framework, on the other hand, allows for iterative updates. As more airborne lidar and national DTMs become available, the global community can integrate them into the existing pipeline. Although the lack of a high-quality lidar-derived DTM globally remains a fundamental constraint, our pipeline provides a scalable solution to this limitation. Once new high-quality data are available, users can fork the repository, contribute new data sources and download scripts, update the YAML configuration file, and run the workflow natively on all major operating systems

(Linux, macOS, and Windows). We believe that continuously updated and FAIR (Findable, Accessible, Interoperable and Reusable) [29] benchmarks will be crucial for the continuous quality assessment of global elevation products as new versions are released.

5. Conclusions

In this study, we have proposed and implemented a fully reproducible pipeline that adapts and extends the methodology established by Guth et al. [10]. By providing a script-based implementation and mitigating sub-pixel shifting, this framework allows the broader geomorphometry community to contribute independent reference data, refine validation metrics, and collectively enhance the robustness of global elevation assessments. Our results partially corroborate previous benchmarks, confirming the same conclusion that FathomDEM generally demonstrates superior performance across a range of land surface parameters. However, our findings diverge from previous reports on the performance of GEDTM30. Specifically, in high-aridity barren terrains, our case study shows no significant performance deficit for GEDTM v1.2 compared to COPDEM. Furthermore, applying the Kruskal–Wallis H-test, we could not conclude that the performance differences between FathomDEM, GEDTM30, COPDEM, and ALOS are statistically significant. In addition, by replacing the scoring system based on the frequency of “*first-place*” rankings with the Rank-Sum Scoring System, the results, contrary to Guth et al. [10], show that GEDTM v1.2 in fact leads in accuracy for elevation and slope and is often between ALOS and COPDEM in high-order LSPs (as expected for an ensemble GDEM). Last but not least, we encourage geomorphometry and digital terrain modeling communities to publish their benchmarks using FAIR frameworks and to make their code available with instructions for scrutiny and easy reproducibility and reusability.

Author Contributions: Conceptualization, Y.-F.H. and T.H.; methodology, Y.-F.H., Q.X. and O.P.C.; software, Y.-F.H., Q.X. and O.P.C.; validation, Y.-F.H., M.S.I. and T.H.; formal analysis, Y.-F.H., M.S.I. and T.H.; investigation, Y.-F.H. and T.H.; resources, T.H.; data curation, Y.-F.H., Q.X., O.P.C. and M.S.I.; writing—original draft preparation, Y.-F.H. and M.S.I.; writing—review and editing, Y.-F.H., O.P.C., M.S.I. and T.H.; visualization, Y.-F.H., M.S.I. and O.P.C.; supervision, T.H.; project administration, T.H.; funding acquisition, T.H. All authors have read and agreed to the published version of the manuscript.

Funding: The Open-Earth-Monitor Cyberinfrastructure project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101059548. The Global Pasture Watch under the Land Carbon Lab project has received funding from the Bezos Earth Fund. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: The research data and code are documented and stored in a self-contained repository at <https://codeberg.org/openlandmap/GEDTM30/src/branch/main/benchmark> (accessed on 1 July 2026), which is part of the GEDTM30 repository (<https://codeberg.org/openlandmap/GEDTM30>, accessed on 1 July 2026). Please refer to our Zenodo repository (<https://zenodo.org/records/18836249>, accessed on 1 July 2026) if you would like to cite the workflow [30]. Comprehensive analytical tables, including tile-specific metrics and FUV scores for the evaluated GDEMs, are included in the Zenodo repository.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used artificial intelligence chatbots to help improve the clarity and readability of this manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: All authors are employed by OpenGeoHub Foundation. Yu-Feng Ho is the main author of GEDTM30’s work. The funders had no role in the design of the study; in the collection,

analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

GDEM	Global Digital Elevation Model
DEM	Digital Elevation Model
LSP	Land Surface Parameter
DTM	Digital Terrain Model
DSM	Digital Surface Model
ALOS	AW3D30 DEM
FAIR	Findability, Accessibility, Interoperability, and Reusability
COP, COPDEM	Copernicus DEM
FUV	Fraction of Unexplained Variance
nMAE	normalized Mean Absolute Error

References

- Moore, I.D.; Grayson, R.; Ladson, A. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Processes* **1991**, *5*, 3–30. [\[CrossRef\]](#)
- Pike, R.; Evans, I.; Hengl, T. Chapter 1 Geomorphometry: A Brief Guide. In *Geomorphometry*; Developments in Soil Science; Hengl, T., Reuter, H.I., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; Volume 33, pp. 3–30. [\[CrossRef\]](#)
- Behrens, T.; Schmidt, K.; MacMillan, R.A.; Viscarra Rossel, R.A. Multi-scale digital soil mapping with deep learning. *Sci. Rep.* **2018**, *8*, 15244. [\[CrossRef\]](#) [\[PubMed\]](#)
- Maxwell, A.E.; Shobe, C.M. Land-surface parameters for spatial predictive mapping and modeling. *Earth-Sci. Rev.* **2022**, *226*, 103944. [\[CrossRef\]](#)
- Earth Observation Research Center Japan Aerospace Exploration Agency (JAXA EORC). *AW3D30 Version 4.1 Product Description*; Japan Aerospace Exploration Agency: Chofu, Japan, 2024.
- European Space Agency. *Copernicus Global Digital Elevation Model*; European Space Agency: Paris, France, 2024. [\[CrossRef\]](#)
- Ho, Y.F.; Grohmann, C.H.; Lindsay, J.; Reuter, H.I.; Parente, L.; Witjes, M.; Hengl, T. GEDTM30: Global ensemble digital terrain model at 30 m and derived multiscale terrain variables. *PeerJ* **2025**, *13*, e19673. [\[CrossRef\]](#) [\[PubMed\]](#)
- Uhe, P.; Lucas, C.; Hawker, L.; Brine, M.; Wilkinson, H.; Cooper, A.; Saoulis, A.A.; Savage, J.; Sampson, C. FathomDEM: An improved global terrain map using a hybrid vision transformer model. *Environ. Res. Lett.* **2025**, *20*, 034002. [\[CrossRef\]](#)
- Hawker, L.; Uhe, P.; Paulo, L.; Sosa, J.; Savage, J.; Sampson, C.; Neal, J. A 30 m global map of elevation with forests and buildings removed. *Environ. Res. Lett.* **2022**, *17*, 024016. [\[CrossRef\]](#)
- Guth, P.L.; Trevisani, S.; Grohmann, C.H.; Lindsay, J.B.; Reuter, H.I. Benchmarking Elevation Plus Land Surface Parameters Finds FathomDEM and Copernicus DEM Win as Best Global DEMs. *Remote Sens.* **2025**, *17*, 3919. [\[CrossRef\]](#)
- Bielski, C.; López-Vázquez, C.; Grohmann, C.H.; Guth, P.L.; Hawker, L.; Gesch, D.; Trevisani, S.; Herrera-Cruz, V.; Riazanoff, S.; Corseaux, A.; et al. Novel approach for ranking DEMs: Copernicus DEM improves one arc second open global topography. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–22. [\[CrossRef\]](#)
- Guth, P.L.; Trevisani, S.; Grohmann, C.H.; Lindsay, J.; Gesch, D.; Hawker, L.; Bielski, C. Ranking of 10 global one-arc-second DEMs reveals limitations in terrain morphology representation. *Remote Sens.* **2024**, *16*, 3273. [\[CrossRef\]](#)
- Guth, P.L.; Geoffroy, T.M. LiDAR point cloud and ICESat-2 evaluation of 1 second global digital elevation models: Copernicus wins. *Trans. GIS* **2021**, *25*, 2245–2261. [\[CrossRef\]](#)
- Ho, Y.; Hengl, T. *Global Ensemble Digital Terrain Model 30m (GEDTM30)*; Zenodo: Geneva, Switzerland, 2025. [\[CrossRef\]](#)
- Fathom, U.K. *FathomDEM v1-0 Americas*; Zenodo: Geneva, Switzerland, 2025. [\[CrossRef\]](#)
- Fathom, U.K. *FathomDEM v1-0 Eurasia and Africa*; Zenodo: Geneva, Switzerland, 2024. [\[CrossRef\]](#)
- Actueel Hoogtebestand Nederland (AHN). Data Feed—Digital Terrain Model (DTM) 0.5 m. 2025. Available online: https://service.pdok.nl/rws/ahn/atom/dtm_05m.xml (accessed on 12 May 2026).
- Federal Office of Topography (swisstopo). swissALTI3D—Download. 2025. Available online: <https://www.swisstopo.admin.ch/de/hoeihenmodell-swissalti3d#swissALTI3D&ÄŤDownload> (accessed on 12 May 2026).
- National Geospatial Technical, Earth Resources Observation, National Geospatial Program. *Seamless 1 Meter Digital Elevation Models (DEMs)—USGS National Map 3DEP Downloadable Data Collection: U.S. Geological Survey Data Release*; U.S. Geological Survey: Reston, VA, USA, 2025. [\[CrossRef\]](#)

20. Geoscience Australia. *Digital Elevation Model (DEM) of Australia derived from LiDAR 5 Metre Grid*; Zenodo: Geneva, Switzerland, 2015. [[CrossRef](#)] [[PubMed](#)]
21. Guth, P. *DEMIX GIS Database Version 4*; Zenodo: Geneva, Switzerland, 2025. [[CrossRef](#)]
22. Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. *ESA WorldCover 10 m 2021 v200*; Zenodo: Geneva, Switzerland, 2022. [[CrossRef](#)]
23. Bauer-Marschallinger, B.; Sabel, D.; Wagner, W. Optimisation of global grids for high-resolution remote sensing data. *Comput. Geosci.* **2014**, *72*, 84–93. [[CrossRef](#)]
24. Feciskanin, R.; Hajdúchová, V. *New Tool for Calculating Land Surface Parameters*; Zenodo: Geneva, Switzerland, 2025. [[CrossRef](#)]
25. Lindsay, J. Whitebox GAT: A case study in geomorphometric analysis. *Comput. Geosci.* **2016**, *95*, 75–84. [[CrossRef](#)]
26. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [[CrossRef](#)]
27. Aria, M.; Cuccurullo, C.; Gnasso, A. A comparison among interpretative proposals for Random Forests. *Mach. Learn. Appl.* **2021**, *6*, 100094. [[CrossRef](#)]
28. Strobl, P.A.; Bielski, C.; Guth, P.L.; Grohmann, C.H.; Muller, J.P.; López-Vázquez, C.; Gesch, D.B.; Amatulli, G.; Riazanoff, S.; Carabajal, C. The Digital Elevation Model Intercomparison eXperiment DEMIX, a community-based approach at global DEM benchmarking. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 395–400. [[CrossRef](#)]
29. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
30. Ho, Y.F.; Qin, X.; Chiriack, O.; Isik, M.S. *Open-Source Global DEM Comparison Workflow*; Zenodo: Geneva, Switzerland, 2026. [[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.