

# Novel Approach for Ranking DEMs: Copernicus DEM Improves One Arc Second Open Global Topography

Conrad Bielski<sup>1</sup>, Member, IEEE, Carlos López-Vázquez<sup>2</sup>, Senior Member, IEEE, Carlos H. Grohmann<sup>3</sup>, Member, IEEE, Peter L. Guth<sup>4</sup>, Laurence Hawker<sup>5</sup>, Dean Gesch<sup>6</sup>, Sebastiano Trevisani<sup>7</sup>, Virginia Herrera-Cruz<sup>8</sup>, Member, IEEE, Serge Riazanoff<sup>9</sup>, Axel Corseaux<sup>10</sup>, Hannes I. Reuter<sup>11</sup>, and Peter Strobl<sup>12</sup>

**Abstract**—We present a practical approach to intercompare a range of candidate digital elevation models (DEMs) based on predefined criteria and a statistically sound ranking approach. The presented approach integrates the randomized complete block design (RCBD) into a novel framework for DEM comparison. The method presented provides a flexible, statistically sound, and customizable tool for evaluating the quality of any raster—in this case, a DEM—by means of a ranking approach, which takes into account a confidence level and can use both quantitative and qualitative criteria. The users can design their own criteria for the quality evaluation in relation to their specific needs. The application of the RCBD method to rank six 1" global DEMs, considering a wide set of study sites, covering different morphological and land cover settings, highlights the potentialities

of the approach. We used a suite of criteria relating to the differences in the elevation, slope, and roughness distributions compared to reference DEMs aggregated from 1- to 5-m light detection and ranging (LiDAR)-derived DEMs. Results confirmed the significant superiority of Copernicus DEM (CopDEM) 1" and its derivative forests and buildings removed DEM (FABDEM) as the overall best 1" global DEMs. They are slightly better than Advanced Land Observing Satellite (ALOS) and clearly outperform NASADEM and SRTM, which are, in turn, much better than advanced spaceborne thermal emission and reflection radiometer (ASTER).

**Index Terms**—Friedman statistics, global digital elevation model (DEM), intercomparison, open-source tools.

Manuscript received 5 September 2023; revised 6 December 2023 and 25 January 2024; accepted 9 February 2024. Date of publication 20 February 2024; date of current version 14 March 2024. The work of Carlos H. Grohmann was supported in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development (CNPq)] under Grant 311209/2021-1 and in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo [São Paulo Research Foundation (FAPESP)] under Grant 2019/26568-0 and Grant 2023/11197-1. The work of Laurence Hawker was supported by the Natural Environment Research Council (NERC) EVOFLOOD Project under Grant NE/S015795/1. (Corresponding author: Carlos H. Grohmann.)

Conrad Bielski is with EOXPLORE, 8204 Oberhaching, Germany (e-mail: conrad.bielski@eoxplore.com).

Carlos López-Vázquez is with the LatinGEO Lab IGM+ORT, Universidad ORT Uruguay, 11100 Montevideo, Uruguay (e-mail: carloslopez@uni.ort.edu.uy).

Carlos H. Grohmann is with the Spatial Analysis And Modelling Lab/Institute of Energy and Environment (SPAMLab/IEE), Universidade de São Paulo, São Paulo 05508-010, Brazil (e-mail: guano@usp.br).

Peter L. Guth, retired, was with the Department of Ocean and Atmospheric Sciences, United States Naval Academy, Annapolis, MD 21402 USA (e-mail: prof.pguth@gmail.com).

Laurence Hawker is with the School of Geographical Sciences, University of Bristol, BS8 1SS Bristol, U.K. (e-mail: laurence.hawker@bristol.ac.uk).

Dean Gesch is with the U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD 57198 USA (e-mail: gesch@usgs.gov).

Sebastiano Trevisani is with the Department of Culture del Progetto, Iuav University of Venice, 30135 Venice, Italy (e-mail: strevisani@iuav.it).

Virginia Herrera-Cruz is with Airbus Defence and Space, 88039 Friedrichshafen, Germany (e-mail: virginia.herrera@airbus.com).

Serge Riazanoff and Axel Corseaux are with VisioTerra, 77420 Champ-sur-Marne, France (e-mail: serge.riazanoff@visioterra.fr; axel.corseaux@visioterra.fr).

Hannes I. Reuter is with the European Commission, Eurostat, L-2721 Luxembourg City, Luxembourg (e-mail: hannes.reuter@ec.europa.eu).

Peter Strobl is with the European Commission, Joint Research Center (JRC), 21027 Ispra, Italy (e-mail: peter.strobl@ec.europa.eu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2024.3368015>, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2024.3368015

## I. INTRODUCTION

OVER the past two decades, several Earth observation missions have resulted in finer than 100-m resolution global digital elevation models (DEMs), most of which are shared freely and openly worldwide. These data revolutionized Earth sciences and spurred many applications that require accurate information about the shape of Earth's surface. Consequently, the demand for high-quality DEMs across multiple disciplines continues to grow, and users around the world are faced with the challenge of understanding the DEMs' intrinsic characteristics and making an informed choice for their particular application.

At this time, at least six different global medium-resolution (i.e., 10–100 m) DEMs have been produced using a variety of techniques. We expect more to come in the near future as new technologies and methods are developed. However, most users do not have the resources and expertise to perform an in-depth comparison of different DEMs. Therefore, expert advice that provides information pertinent to the user's need will benefit the community to identify the most appropriate dataset. This challenge identified by the Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (WGCV) was taken up by the Terrain Mapping Subgroup (TMSG) in its Digital Elevation Model Inter-comparison eXercise (DEMIX), by developing open and transparent tools to compare DEMs and rank them based on their choice of criteria.

As a major step toward this goal, this article presents a novel and flexible tool to provide geospatial data users with a practical approach to intercompare a set of candidate global DEMs based on predefined criteria and a statistically sound ranking approach. During the discussions and development,

the DEMIX group referred to the method as the “wine contest” as an analogy to a comparison of wines by human judges, but with DEMs instead of wines, and both qualitative and subjective criteria instead of judges. The framework provides the wider geospatial community with tailored recommendations regarding available DEM products that are not limited to one domain, geographic area, or landscape type. The method is flexible and customizable in relation to the specific needs and requirements of users in their particular application. We then rank global 1” DEMs and show that three of them are clearly superior based on the chosen criteria.

#### A. Brief Review of Comparing and Ranking Global DEMs

The Shuttle Radar Topography Mission (SRTM), which flew in 2000 and released edited DEMs starting in 2004, revolutionized digital topography [1]. Previously DEMs had been created from stereo photogrammetry, either directly from the imagery or indirectly from the contours on topographic maps, covered only a single country, and were often not freely available [2]. SRTM produced and released a globally consistent DEM for latitudes lower than 60°, initially at 1 arc second (1”, ~30 m at the equator) grid spacing in the United States and at 3” (~90 m) everywhere else, which, starting in 2014, was also made available globally at 1” spacing. The U.S. Geological Survey (USGS) held a workshop in 2005, and the published papers set the standard for assessing SRTM using reference data and determining the quality of the DEM [3]. The approaches used either sparsely located control points [4], [5] or comparison to reference DEMs using all points in the grid, comparing the elevations or including derived geomorphometric parameters such as slope [6]. These workshop papers also emphasized that results vary with different land cover and terrain slope categories [5], [7].

The release of the 1” advanced spaceborne thermal emission and reflection radiometer (ASTER) global DEM [8] led to a large number of papers comparing SRTM and ASTER to each other and reference data and assessing whether SRTM or ASTER was “better” [9], [10], [11]; generally, finding SRTM was superior although, at that point, the 1” SRTM was still not available globally. A new milestone was achieved with the release of the TanDEM-X DEM [12], [13], a global digital surface model (DSM) at 0.4” latitude pixel spacing and 0.4”–4” in longitude. The data set remained limited to scientific use but produced a number of comparisons to SRTM, ASTER, and Advanced Land Observing Satellite (ALOS) [14]. TanDEM-X DEM is also the basis for the 1” and 3” DEMs of the same name and the basis for the commercial WorldDEM and, therefore, also for the Copernicus DEM (CopDEM) [15]. Sequential releases of ALOS World 3D [16], NASADEM [17], TanDEM-X DEM, and CopDEM led to additional comparisons of the freely available 1” DEMs trying to establish which was “best.”

Most studies compare DEMs against each other or with reference data (Ice, Cloud and Land Elevation Satellite (ICESat)/global navigation satellite system (GNSS) point elevations or light detection and ranging (LiDAR) datasets) for a limited number of sites usually located in one country and considering a single parameter [i.e., elevation differences (ELVDs)] and a single criterion, typically based on popular

statistics such as mean, median, standard deviation (STD), and root mean square error (RMSE) [4], [5], [6], [14], [18], [19], [20], [21], [22], [23]. Other criteria include comparisons of contour lines [14], [22], drainage network [19], [23], and topographic profiles [14], [21], [22].

In all the cases the “best” DEM is the one with the lowest value for the metrics of difference (that is, closest to the reference data), but, even with the same criterion, different sites might produce a different ranking. In that case, there is no reason to expect that there will be an agreement for the ranking using different criteria. As a result, these studies provide only very limited recommendations of which DEM performs better for a specific use and in a specific region [18], [24].

The availability of 3” DEMs like the public version of the TanDEM-X DEM and MERIT [25] produced additional comparisons [26], complicated by the difference in resolution. The free ALOS and CopDEM represent downsampled versions of commercial DEMs. Commercial DEMs are beyond the scope of this article because of their cost and the many difficulties of comparing DEMs with different spatial resolutions.

All of the global 1” DEMs have gone through sequential editions, starting with the use of ASTER to fill the voids in SRTM, and are all to some degree composite DSMs [27]. The next evolution in DEMs has been the application of machine learning to “improve” current global DEMs considering additional external data; Hawker et al. [28] created the first freely available, global digital terrain model [27] named forests and buildings removed DEM (FABDEM).

While the conceptual difference between a DSM and a digital terrain model (DTM) is clear, the practical matter of creating the DEM from a satellite-based sensor is not so easy. How the sensor integrates the signal over a pixel area and how much the signal penetrates vegetation vary with the sensor and the characteristics of the area. Guth et al. [27] showed how the global DEM elevations varied within a high-resolution LiDAR point cloud, and Guth and Geoffroy [21] showed statistical distributions of the DEM elevations above, below, or within a LiDAR point cloud, and the results differed among the global 1” DEMs. Calling these DEMs DSMs is a simplification; current technology does not allow for the selection of a DSM or DTM but simply reports what the sensor measured. Even the derived FABDEM, seeking to adapt CopDEM and create a DTM, cannot create a perfect DTM and does not improve on CopDEM everywhere. The other DEMs are closer to DSMs but most likely fall somewhere between a real DTM and DSM [21]. We compare the global DEMs to both reference DTMs and DSMs; the user must clearly understand the implications of the results and decide which are valid for their purpose.

## II. METHODS

Table I summarizes the steps to evaluate six global 1” DEMs with 236 widely distributed test tiles from 24 test areas. We produce a geographic information system (GIS) database that contains an evaluations table with numerical metrics comparing the test DEM to a reference DEM and an opinions table that translates the evaluations to an ordinal ranking that allows for ties and statistical confidence in the results (see Section II-F2 for details).

TABLE I  
GENERAL DEMIX STEPS AS EXECUTED FOR THE COMPARISON OF GLOBAL DEMs

Step	Section	Software	Task	Details
1	II-B	Web downloads	Obtain high quality reference source data and candidate DEMs	6 candidate global DEMs, 1", geographic Source reference DTM: 1-5m UTM projection Source reference DSM: if available
2	II-C	MICRODEM	Prepare reference DEMs	Pixel-is-point Pixel-is-area For high lat COPDEM, separate Pixel-is-point For high lat ALOS, separate Pixel-is-area
3	II-D	MICRODEM	Prepare candidate DEMs	Convert to ellipsoid, then to geoid Use grids from national/regional mapping agency Use GDAL for areas in the USA
4	II-E	MICRODEM	Build GIS Database	Compute test area statistics Create difference outputs Compute metrics and evaluations table Produce opinions table from initial tolerances
5	II-F	Jupyter Notebook	Rank global DEMs	Choose criteria to use Adjust tolerances if required Recreate opinions table Filter database for land types or tiles Produce final rankings Compute confidence levels Produce graphics

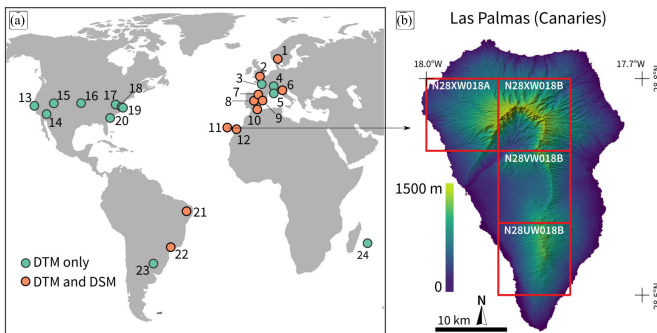


Fig. 1. (a) Location of the 24 test areas made up of 236 DEMIX tiles. (b) Distribution of DEMIX tiles over the Las Palmas Island. The names of the test areas shown on map (a): 01: Norway (none tiles); 02: Oxford (four); 03: Caen (six); 04: Valonne (nine); 05: Vanoise (four); 06: Trentino (one); 07: Pyrenees (two); 08: Madrid (35); 09: Ebro Delta (27); 10: Almeria (30); 11: Las Palmas (four); 12: Canary East (18); 13: Redwoods (six); 14: State Line (13); 15: Canyon Range (14); 16: Republican River (19); 17: Shenandoah (two); 18: Blackwater (six); 19: Chincoteague (five); 20: Charleston (six); 21: Pernambuco (four); 22: São Paulo (five); 23: Uruguay (one); and 24: La Reunion (six).

### A. Test Areas and DEMIX Tiling

The reference data for the intercomparison of global DEMs have significantly higher accuracy elevation values and much smaller sampling spacing than the DEMs to be tested. These reference DEMs were produced by a variety of mapping agencies from around the world with sampling intervals between 1 and 5 m, which is more than five times finer spatial resolution than the global DEMs considered. The distribution of the test areas with reference DEMs is shown in Fig. 1; some of the reference DEMs provided separate DSMs and DTMs. We decided not to alter the global DEMs beyond converting all to the EGM2008 vertical datum and only transformed the reference data to match the global DEMs.

For each of the test areas, one or more DEMIX tiles were extracted. A DEMIX tile covers approximately  $10 \times 10$  km in size defined on a geographic latitude/longitude grid. This grid covers the entire globe and was developed in the context

of DEMIX to allow anyone to use a consistent sampling. The DEMIX tiles provide a global tessellation at a scale fine enough to produce locally significant results, coarse enough to keep the total number of tiles manageable, and nearly equal in size to allow for statistical aggregation and comparative analysis. Technical details, naming convention, and grid definition files can be found in [29].

The area covered by the reference DEMs varied by location and, therefore, the number of DEMIX tiles covering each area differs. Reference DEMs do not have to fit perfectly within DEMIX tiles, and the expectation is that most will not because of the irregular boundaries of the reference DEMs due to mapping project extent, administrative boundaries, or DEM boundaries in universal transverse Mercator (UTM) projection rather than geographic lines. Furthermore, it is not necessary to fill entirely a DEMIX tile; otherwise, it would be impossible to include many interesting geographical features such as coastal areas or missing data because some mapping agencies do not edit water and leave voids in LiDAR-derived DEMs. However, in order to include the tile in our database, we require that at least 75% of the DEMIX tile is covered by reference and candidate data. Fig. 1(b) shows the Las Palmas island area (Location 11). Two of the four tiles are full, whereas two tiles covering the coast are  $> 75\%$  full. The other tiles will not be included because they do not fulfill the 75% DEMIX sample tile coverage rule.

Both DTMs and DSMs were used in this evaluation where available. While DTMs were available for all 24 test areas and 236 tiles, DSMs are only available in 11 of those areas (134 tiles) because some mapping agencies do not provide them (see Fig. 1).

### B. Step 1—Obtain High-Quality Reference Source Data and Candidate DEMs

For a comparison of global DEMs, the reference DEMs must cover a representative sample of the different landforms

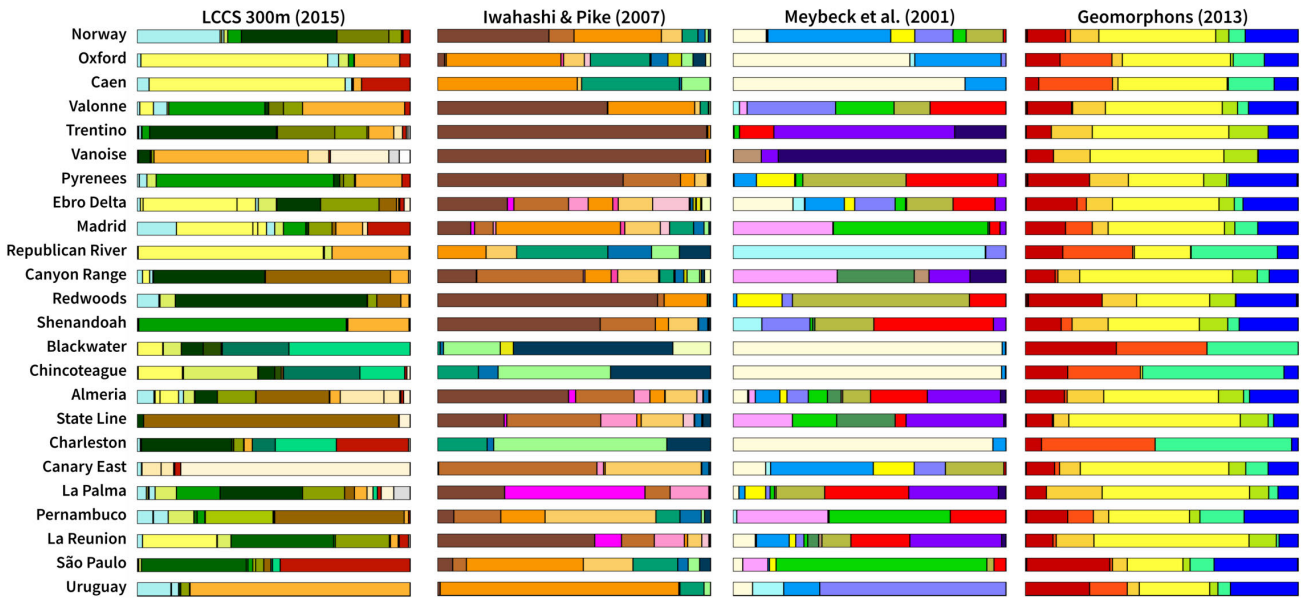


Fig. 2. Tile classifications for the 24 test areas in our sample for one land cover classification and three geomorphometric landform classifications. Areas are arranged from north to south. Legend for colors is available in Fig. 1 in the Supplementary Material.

observed on Earth. The quality of the global DEMs being compared is not expected to differ across national or continental borders because the same acquisition technique and processing were used to create each of the DEMs under consideration. Therefore, it was necessary to be sure that a representative sample was available and covered landforms available on all continents. As a first approximation, this was achieved by calculating statistics and classifying landforms found within sample tiles. Summary statistics were calculated across all 236 tiles using the reference 1" DTMs. The results are summarized in Table II for the following characteristics:

- 1) mean elevation in meters;
- 2) mean slope in percentage [30];
- 3) mean roughness in percent, with roughness defined as the STD of the slope in a  $5 \times 5$  window [31];
- 4) relief—elevation range in meters;
- 5) percent forest—percentage that is in a forest category [32];
- 6) percent urban—percentage that is urban [32];
- 7) percent barren—percentage that is one of the barren categories [32].

The results show that a wide range of surface characteristics are captured in the test tiles. These summary statistics are further supported by land cover and geomorphometric classifications.

Fig. 2 presents the classification associated with each of the 24 test areas. The left column displays the land cover classes using the United Nations Land Cover Classification System with 38 categories [32]. The color denotes the class, whereas the width of the bar denotes the overall land cover percentage of that particular cover class. The right three columns of Fig. 2 present the geomorphometric classifications from left to right [33], [34] and geomorphons [35]. These geomorphometric classes range between 7 and 16 categories each. Furthermore, Fig. 2 reveals the challenge in trying to neatly

characterize a 100-km<sup>2</sup> tile because no two DEMIX tiles, even adjacent ones, are alike from their surface characteristics. Our tile selection clearly represents a large range of different landscapes and provides a much better diversity of surface characteristics than previous efforts comparing global DEMs.

### C. Step 2—Prepare Reference DEMs

In order for the reference DEMs to be compared directly with the global DEMs, they must be transformed into the same spatial sampling (grid) and vertical reference (datum). The following key global DEM characteristics were managed to produce the reference DEMs. The pixels in the geographic DEMs are rectangular (larger in the north–south direction once away from the equator) and rotated slightly compared to the UTM projection of the source DEM.

*Match Grid Type:* The global DEMs under consideration have different grid representations that affect their spatial location: pixel-is-point and pixel-is-area [27]. Only one-quarter of the area is shared by pixels defined by the two representations, and although they have the same nominal coordinates in uneven terrain, they are unlikely to have the same elevation.

*Spatial Sampling:* A 1" DEM cannot be compared directly with a reference DEM having a spatial resolution of 1–5 m because of the significant differences in the spatial scale. The global DEMs have a spatial sampling of approximately 30 m (1") at the equator, and their sensing techniques can be assumed to deliver an average elevation in each given cell. Therefore, the reference DEMs must be aggregated in order to match the spatial sampling of the global DEM.

*Aggregation:* Global DEMs are provided in geographic coordinates, whereas reference DEMs typically use local projections such as UTM to avoid spatial distortions. While interpolation has been used to solve the reprojection issues, Guth and Kane [36] showed that reinterpolation changed the characteristics of the DEM and care in the algorithms allows

TABLE II  
REFERENCE DTM COLLECTION SUMMARY STATISTICS FOR ALL 236 DEMIX TILES

	Average elevation (m)	Average slope (%)	Average roughness (%)	Relief (m)	Forest in tile (%)	Urban in tile (%)	Barren in tile (%)
Maximum	2226.1	76.7	22.7	2363.8	99.7	96.9	98.7
Mean	564.9	15.4	4.8	530.8	29.3	6.7	22.3
Median	461.5	10.4	3.6	356.1	12.2	1.2	8.2
Minimum	-0.9	0.1	0.0	2.8	0.0	0.0	0.0
Std Dev	527.7	14.3	4.0	516.1	32.3	15.2	27.9

accurate computations with the geographic coordinates without introducing significant interpolation errors. The projected coordinates of each pixel in the source reference DEMs are converted to WGS84 latitude and longitude and mapped to the corresponding pixel in the test DEM. The value for the reference 1" DEM is the mean of the points in the high-resolution source DEM, ranging from about 36 points for a 5-m DEM to about 900 for a 1-m DEM.

*Latitudinal Effects:* The pixel spacing changes with latitude for CopDEM and ALOS but remains constant for the others.

*Vertical Datum:* An elevation model is referenced to a vertical datum. We moved the reference DEMs to the EGM2008 vertical datum using bilinear interpolation and the highest resolution available transformation grids from the PROJ library (<https://proj.org/>) or using GDAL to provide combined vertical and horizontal datum shifts for the United States DEMs. Both the PROJ library and GDAL came via OSGeo4W install (<https://www.osgeo.org/projects/osgeo4w/>, downloaded on April 11, 2023). In a few cases, we used a local transformation grid, not included in the PROJ library, which does not have an European Petroleum Survey Group (EPSG) code.

The DEM characteristics were carefully monitored during the preparation of the reference DEMs. This was achieved by using MICRODEM version 2023.6.20 [37], [38] as the reference software for this experiment because it allowed access to the source code to analyze and fix any issues with the preparation of the reference DEMs and/or transformations of the global DEMs. The authors are not aware of other software that will do the aggregations correctly from UTM to geographic grids in both pixel-is-area and pixel-is-point geometries with the same simplicity and transparency.

For each source DTM, the following reference DTMs were produced to cover the different global DEMs in the test areas:

- 1) 1"  $\times$  1" pixel-is-point DTM;
- 2) 1"  $\times$  1" pixel-is-area DTM;
- 3) 1"  $\times$  CopDEM spacing pixel-is-point for high-latitude DEMs;
- 4) 1"  $\times$  ALOS spacing pixel-is-area for high-latitude DEMs.

If the reference DEM also had a DSM, the same types of reference DSMs were produced. Up to eight reference DEMs could be required covering each DEMIX tile. This makes it possible to perform quantitative analysis between reference DEMs and the global DEMs without the need for further interpolation or adjustment of the candidate DEM. The reference DEMs, with source metadata, are available in [39].

#### D. Step 4—Prepare Candidate DEMs

We moved NASADEM, ALOS, SRTM, and ASTER from the EGM96 to the EGM2008 vertical datum, forcing them to be of floating point representation, using the 15 arc-minute grid for EGM96 (the best available) and the 1 arc-minute grid for EGM2008. CopDEM and FABDEM are delivered in floating point precision for elevation with EGM2008 and did not require any changes.

#### E. Build the GIS Database

Creating the GIS database used for the rankings involves a number of design considerations, including the selection of criteria and the tolerances to decide if small numerical differences represent significant differences.

1) *Criteria Used for DEM Comparison:* For this article, a pixel-by-pixel comparison was applied for the assessment of global DEMs, using a reference DEM with higher accuracy elevation values. A pixel-by-pixel comparison computes the difference between a parameter calculated using the global DEM and the reference DEM; the parameter can be either the elevation or a derived focal parameter. The reference DEM (DTM and DSM if available) is assumed to be the best representation of elevation and derived parameters for a particular sample location. Consequently, any differences will be considered as the error of the candidate global DEM.

The quantitative assessments carried out on the global DEMs were based on pixel-by-pixel differences of three geomorphometric parameters: elevation, slope, and roughness index. The following parameters were computed for the six candidate DEMs for all DEMIX tiles with respect to the reference DTM and reference DSM where available: ELVDs, slope differences (SLPDs) using the [30] algorithm, and roughness differences (RUFDS) from the STD of slope in a 5  $\times$  5 window [31].

These parameters were chosen because they measure different aspects of the DEM and the underlying terrain and are not closely correlated. As derivatives of elevation, the slope and roughness parameters are more sensitive to nonsystematic DEM elevation errors and are also among the most used key parameters applied in many analyses for the characterization of DEMs. Among the various roughness indices [31], the STD of slope can be considered a flow-directional roughness index [40] being computed in the direction of the gradient.

The differences between the reference DEM and global DEM were always computed in the same manner

$$\text{difference} = \text{global}_{\text{DEM}} - \text{reference}_{\text{DEM}}$$

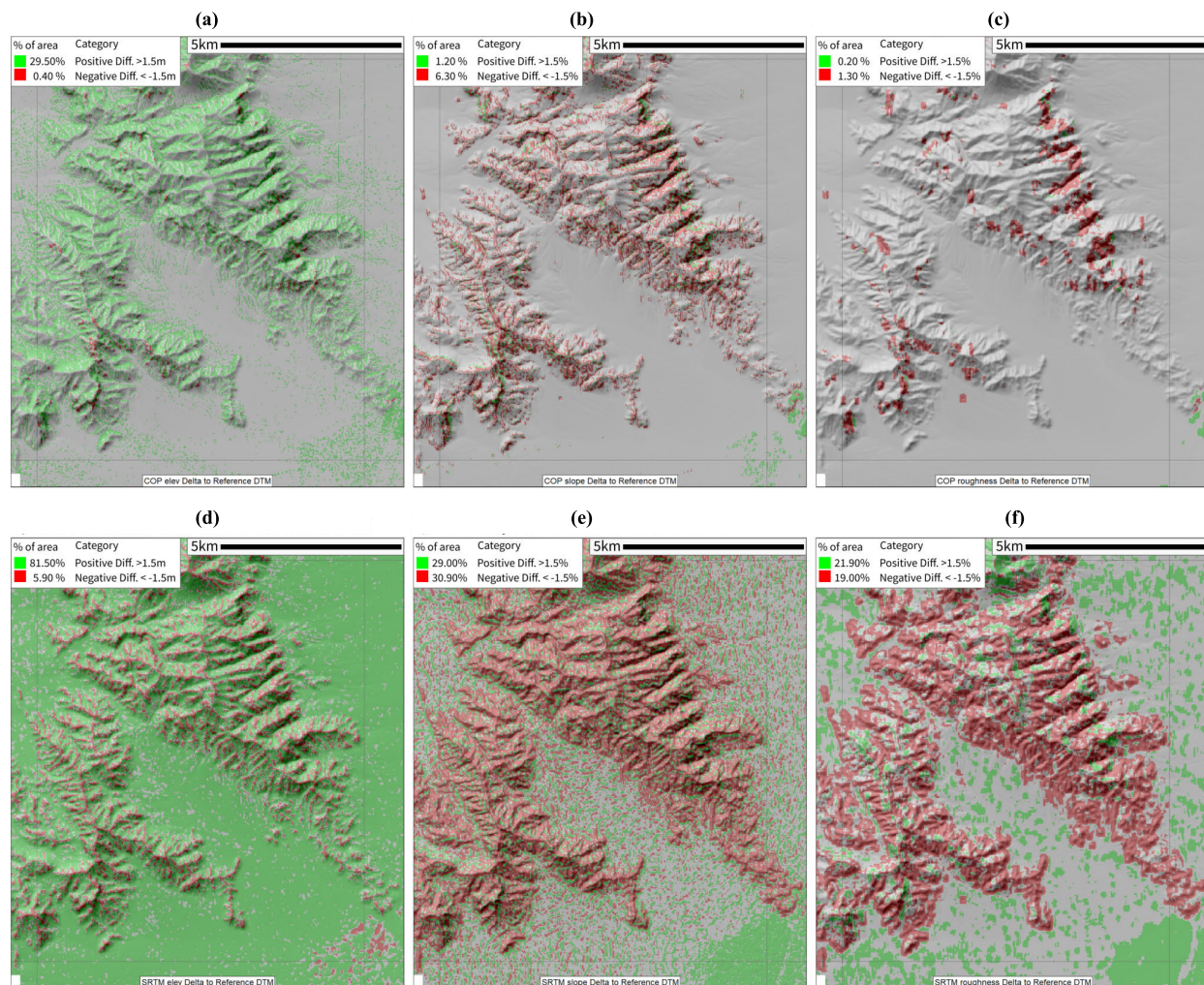


Fig. 3. Difference maps for tile N35VW116G for CopDEM and SRTM compared to the reference DTM. Difference maps for all six global DEMs are presented in Fig. 2 in the Supplementary Material. (a) COPDEM elevation difference to reference DTM. (b) COPDEM slope difference to reference DTM. (c) COPDEM roughness difference to reference DTM. (d) SRTM elevation difference to reference DTM. (e) SRTM slope difference to reference DTM. (f) SRTM roughness difference to reference DTM.

where positive values indicate that the global DEM has a higher value than the reference DEM at a specific pixel. The ELVD parameter cannot be ranked because the range of outputs will be signed and ideally distributed around zero if the DEM is unbiased. Fig. 3 shows the difference maps for tile N35VW116G for CopDEM and SRTM, draped on a hillshade. Fig. 4 shows the difference distributions for this tile for all six DEMs. In this tile, there is a slight positive bias for elevation and a negative bias for slope for CopDEM, which is a variable result among all the tiles that we considered. The very well-defined modes close to 0 occur in almost all the 236 tiles.

In order to achieve a quantitative assessment required for the randomized complete block design (RCBD) method based on the above parameters, it was necessary to compute derived statistics from the distribution of the differences for each of the 236 DEMIX tiles.

- 1) *STD*: A characteristic of the difference distribution, which is not related to the tolerance that we use to create ties.
- 2) *Average Deviation (AVD)*: The mean difference was subtracted from each difference, and the average of its absolute value was computed [41].

- 3) RMSE.
- 4) Mean absolute error (MAE).
- 5) *Linear error 90 (LE90)*: The 90th percentile of the distribution of the absolute differences.

For all of the chosen metrics, the quantitative assessment rule is that lower is better. These metrics output only nonnegative values, which allows them to be easily ranked. These metrics measure different properties of the difference distribution, putting differing weights on the tails of the distribution. The database also includes the mean and median for each difference distribution because they provide additional context in interpreting the results, despite not being useful for ranking purposes.

The three parameters and five metrics for each DEMIX tile give 15 different criteria on which to apply the RCBD methodology. These are presented in the following manner: ELVD\_RMSE, where the first acronym specifies the computed parameter and the second acronym specifies the metric applied.

2) *Per-Pixel Land Types*: Global DEMs perform differently depending on the underlying characteristics of the surface. As shown in Fig. 2, any single DEMIX tile incorporates a variety of surface characteristics whose tendencies can be lost

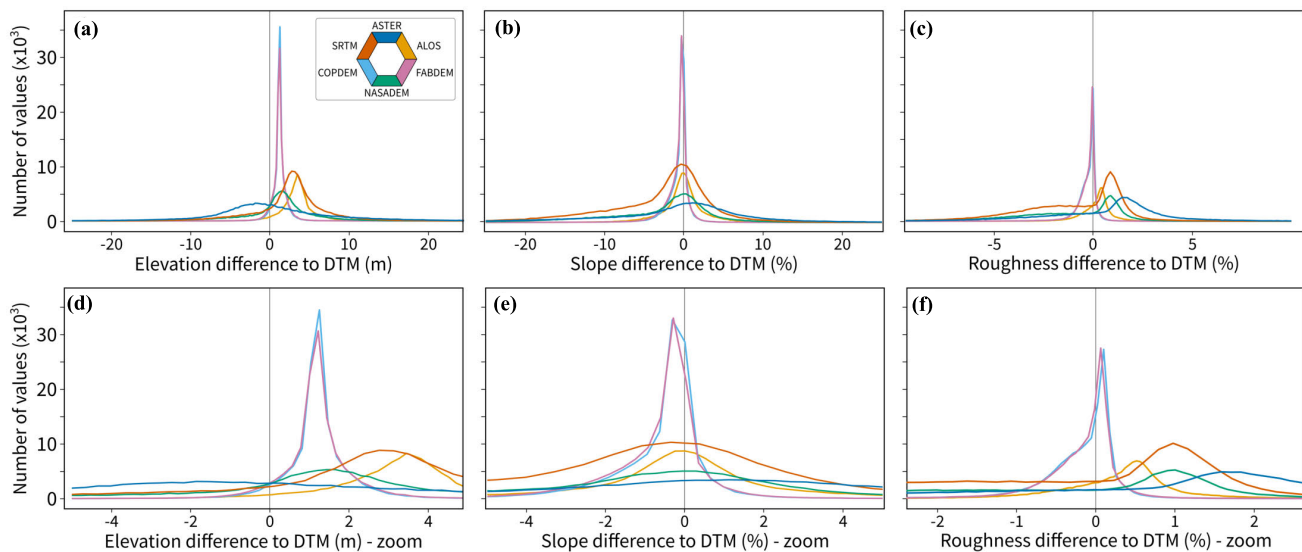


Fig. 4. Distribution of elevation (a) and (d), slope (b) and (e) and roughness (c) and (f) differences for tile N35VW116G, compared to the reference DTM. Positive values mean that the candidate DEM has higher values than the reference. (Top) Entire distribution. (Bottom) Region around zero.

when looking at overall metrics. In addition to computing statistics for every pixel in the tile, we computed separate pixel-by-pixel statistics for different slope and land cover classes. We used three land cover classes (forest, urban, and barren) from the Copernicus Global Land Cover Layers—Collection 2 [32] and four slope categories: cliff (slope > 50%), steep (slope > 25% and < 50%), gentle (slope < 25% and > 12.5%), and flat (slope < 12.5%).

3) *DEMIX Database*: MICRODEM [37], [38] produced the DEMIX GIS database [42], combining the tile characteristics, an evaluations table, and an opinions table with an initial set of tolerances (see Section II-F2). The evaluations table contains the numerical results for each criterion for every tile and will not change during the ranking. Calculating the evaluations table and tile characteristics is a relatively slow process but will not have to be recomputed unless the DEMs or criteria change. It can be expanded at any time by additional tiles or criteria.

The opinions table ranks the six DEMs from 1 to 6 (a low score is best), with the possibility of ties. We assign a tolerance for each criterion; evaluations within the tolerance will be treated as a tie. Many of the candidate global DEMs are rounded to integer type; therefore, it does not make sense to treat differences in the order of centimeters or even decimeters as significant for any elevation-related metric. Based on our experience with global DEMs, we have chosen the tolerances below; for other applications, different tolerances might be appropriate. For the elevation parameters, we chose a tolerance of 0.5 m. For slope, an elevation error of 0.5 m over a 30-m horizontal distance would lead to a slope error of about 1%, and we picked a tolerance of 0.5%. We used a tolerance of 0.2% for the roughness. Given the evaluations table and the tolerances, the opinions table can be created on the fly very rapidly after a change of the tolerances or filter of the database. Output from the database can be extracted in an interchangeable, human-readable, and comma-separated values (.csv) file format.

The current DEMIX database [42] includes 55 699 opinions that are rows found within the database. The following fields are included to help read the database:

- 1) tile name, test area, and centroid coordinates;
- 2) tile characteristics required for per tile filters;
- 3) whether the reference DEM is a DTM or a DSM (II-B);
- 4) the land type, with ALL for every pixel, or one of the seven categories defined above if there are at least 100 pixels of that type within the tile and the percentage of the tile in the land type;
- 5) criterion (quantitative assessment metric) name for the 15 applied metrics (plus six other signed metrics for analysis purposes not applicable to the RCBD method) defined above;
- 6) evaluations table using all six candidate DEMs;
- 7) tolerance used to populate the opinions table and identify ties;
- 8) opinions table for each computed metric including the identification of potential ties due to the predefined tolerances.

A summary of the composition of the DEMIX database used for the ranking of the candidate global DEMs is presented in Table III. The table shows the number of tiles with more than 100 pixels (out of roughly 130 000 pixels) within a tile based on the eight defined surface types. This adds up to the total number of opinions available in the database to produce the final ranking. For example, of the 236 reference DTM tiles (the “ALL” surface type), only 152 have the minimum required 100 “cliff” surface type pixels, whereas 181 tiles have the minimum required number of “urban” surface type pixels.

Any combination of opinions can be used in the final ranking. The database can be filtered to show only opinions for a single land type, such as “steep” or “barren,” or a user may only be interested in the tiles where the forest cover type covers at least 50%–75% of the tile or where the relief is greater than 1000 m. This would filter the evaluations table to only include tiles that fulfill the desired requirement.

TABLE III  
COMPOSITION OF THE DEMIX DATABASE IN TERMS OF LAND TYPES

REF_TYPE	AREA	Land Types								Total Types Present	Total Opinions
		ALL	BARREN	FOREST	URBAN	FLAT	GENTLE	STEEP	CLIFF		
DSM	11	134	129	116	121	134	132	131	108	897	15080
DTM	24	236	231	201	181	236	219	190	152	1494	24705
Total	24	370	360	317	302	370	351	321	260	2391	39785
											20640

The database stores all the needed outputs, including the evaluations and opinions tables, for the RCBD method applied to the comparison of the global DEMs. Consequently, they do not need to be recomputed when a subset is used to compute the final ranking.

#### F. Rank Global DEMs

We implemented statistical tests to use the GIS database to rank the DEMs and provide statistical significance. A Jupyter notebook performs the calculations and creates graphical output.

1) *Choice of Criteria:* This step is one of the most challenging parts of the method because it relates to the choice and calculation of the quantitative criteria, which, in the end, will be the basis for the ranking of the global DEMs. For this study, the chosen metrics are based on the computed differences to the reference DEMs. The following will underpin the choice of the quantitative assessments for this DEMIX ranking.

First, it is advantageous to use metrics that are not highly correlated because the amount of diverse information is maximized for the comparison. To illustrate this, correlation matrices were computed based on the 15 quantitative assessment metrics presented for all 236 DEMIX tiles using CopDEM and the reference DTM and DSM where available (see Fig. 5, leftmost column matrices) for the land types flat, steep, and barren. The five metrics appear to be correlated within each parameter but less correlated between parameters. For example, generally high correlations (light green) are observed particularly among the five roughness metrics and the five slope metrics. The lowest correlations (darker blue) observed are between the elevation and roughness parameters.

Second, although the design of the RCBD method precludes the use of quantitative magnitudes that cannot be ranked (e.g., the mean), it is of interest to explore magnitudes that provide insights into the comparison of global DEMs. They provide valuable information to interpret the outcomes, as shown in Fig. 4, where the distribution of pixel differences between the reference DSM and DTM is presented for a single tile. Such information provides insights into whether a candidate's global DEM is higher or lower, steeper or gentler, and rougher or smoother compared to the reference DEM.

2) *Statistical Confidence With Mixed Quantitative and Qualitative Opinions:* Ranking a collection of DEMs from a given set of candidates leads to a conceptually similar problem to comparing wines. We use this analogy in order to develop a novel, extensible, and user-tunable comparison of global DEMs. Since it is the object to be ranked, we identify

DEMIX with wines. For every combination of DEMIX tile and criterion, there will be one ranking, formally similar to the one produced by a judge (tile and criterion) analyzing wines (DEMIX). We are aware that, for wines, most criteria are subjective and require considerations over many sensory aspects. However, objective opinions can be addressed as well. We wish to underline that the method can also be applied to comparisons of other geospatial products, is in no way limited to the ranking of global DEMs, and only requires that the products are ranked according to defined criteria.

The proposed DEMIX RCBD ranking provides a statistical procedure to use the results from quantitative and/or qualitative criteria to rank DEMs and provide statistical confidence to the final ranking. The adopted statistical procedure is formally known as the RCBD [43]. We apply the RCBD in the context of an intercomparison of global 1'' DEMs to identify the most appropriate DEM for a given set of criteria and sample sites while providing statistical confidence to the outcomes. We use multiple criteria that can be either qualitative or quantitative.

Based on this design (RCBD), two tables will be built. The first one will be denoted here as the evaluations table (see Table IV, left-hand side). It has  $k$  columns (one per global DEM) and  $N$  rows (one for each criterion applied to each site), holding the numerical results from evaluating the criterion for a site and for a DEM. For example, one row might hold the elevation RMSE values for a particular site. A ranking for each row of the evaluations table is produced and stored as the corresponding row of the opinions table (see Table IV, right-hand side). Each row of the opinion table holds a ranking, which (in a case without ties) is a permutation of the integer values from 1 to  $k$ . The opinion table can be interpreted as the starting point of the RCBD, so no requirements are posed over how it was produced. Thus, the evaluations can be either qualitative (i.e., subjective based on the opinion of an expert) or quantitative based on some objective method that can produce a numerical result amenable to be sorted (like the RMSE). If a criterion cannot produce a ranking, it cannot be part of the RCBD method. The ranking is built with the rule of "lower is better." In the RCBD context, the following associations can be made.

- 1) The DEMs are evaluated in an arbitrary order, and therefore, the test is randomized. This is important only for subjective criteria.
- 2) There must be a valid opinion for every wine based on their individual taste, so the test is complete. There are no voids in the opinions table.
- 3) There is no reason to expect the criteria (block) to produce identical opinions for the same set of wines.



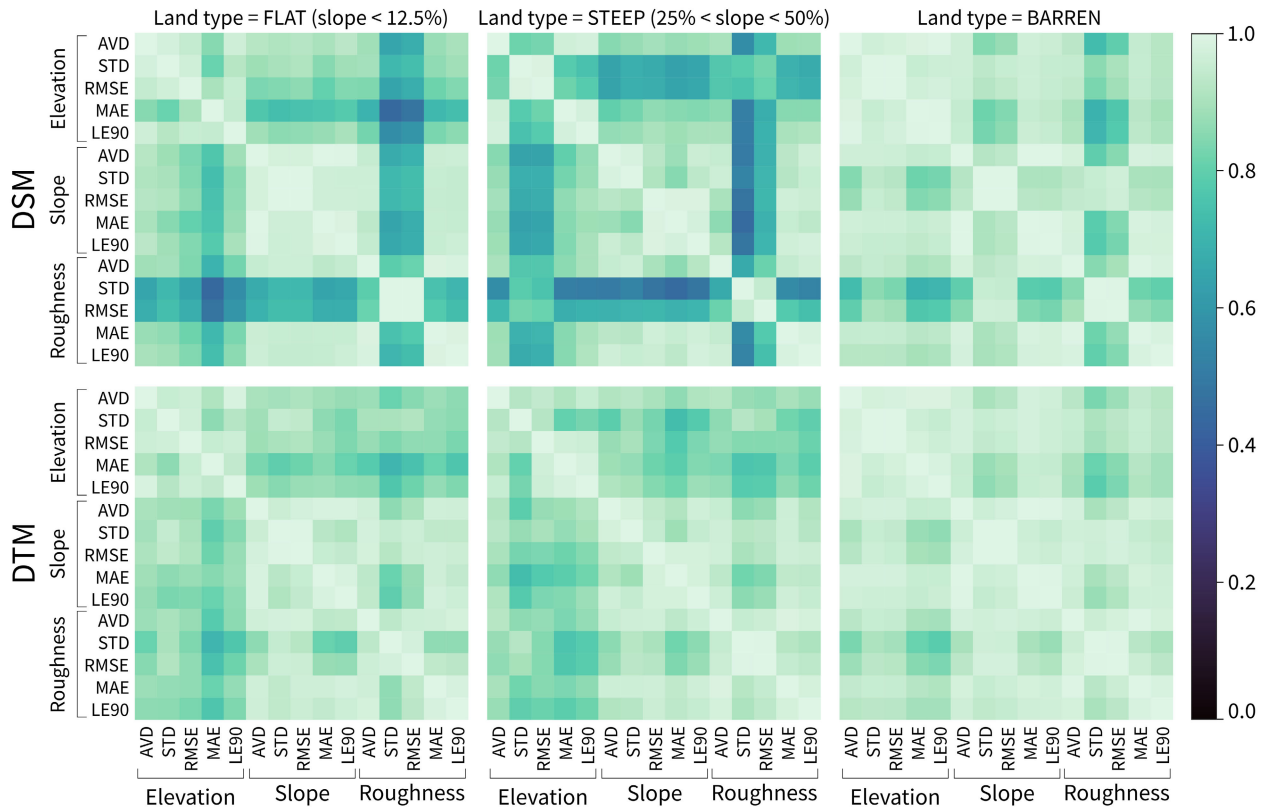


Fig. 5. Correlation matrix among the 15 criteria for 236 DEMIX tiles using reference DTMs and DSMs for three different land types. Note the generally high correlations, particularly among the five roughness criteria and the five slope criteria. The lowest correlations are between the elevation and roughness criteria. Matrices are computed using the Pearson correlation coefficient.

TABLE IV

REALISTIC EXAMPLE APPLIED TO THE INTERCOMPARISON OF DEMs—THE EVALUATIONS TABLE (LEFT-HAND SIDE) RECORDS THE ASSESSMENT OUTCOMES FOR EACH OF THE CANDIDATE DEMs. THE OPINIONS TABLE (RIGHT-HAND SIDE) TRANSLATES THE EVALUATIONS TABLE TO RANKED OPINIONS. ELVD: ELEVATION DIFFERENCE FROM REFERENCE; SLPD: SLOPE DIFFERENCE FROM REFERENCE; RUFD: ROUGHNESS DIFFERENCE FROM REFERENCE; RMSE: ROOT MEAN SQUARE ERROR; LE90: LINEAR ERROR AT 90%; MAE: MEAN ABSOLUTE ERROR; AND AVD: AVERAGE DEVIATION

Criterion	Evaluations						tolerance	Opinions						
	FABDEM	CopDEM	ALOS	NASADEM	SRTM	ASTER		FABDEM	CopDEM	ALOS	NASADEM	SRTM	ASTER	
ELVD_RMSE (m)	2.46	2.44	4.42	5.36	5.78	9.00	0.5	1.5	1.5	3	4.5	4.5	6	
ELVD_LE90 (m)	2.38	2.29	6.04	8.36	8.67	14.20	0.5	1.5	1.5	3	4.5	4.5	6	
SLPD_RMSE (%)	3.52	3.47	3.79	9.22	9.04	12.48	0.5	2	2	2	4.5	4.5	6	
SLPD_MAE (%)	1.28	1.19	2.20	5.83	5.70	8.26	0.5	1.5	1.5	3	4.5	4.5	6	
RUFD_AVD (%)	0.51	0.52	0.63	2.28	2.22	2.61	0.2	2	2	2	4.5	4.5	6	
RUFD_RMSE (%)	1.24	1.25	1.03	3.15	3.09	3.70	0.2	2.5	2.5	1	4.5	4.5	6	
$R_i$								11.0	11.0	14.0	27.0	27.0	36.0	
$R_j^2$								121.0	121.0	196.0	729.0	729.0	1296.0	sum=3192.0

The null hypothesis is that there is no difference among the DEMs, and a consensus based on the opinions cannot be achieved. However, if the null hypothesis is rejected, then the contest ranking is not based on chance (given a chosen confidence level), and some conclusions can be obtained.

Criteria that, in general, cannot be ranked, i.e., whose results cannot be interpreted in the sense of better or worse, are not applicable. Ties, individual results that are considered equally good (or equally bad), can be considered by applying the mid-rank procedure [44] and arise after considering tolerances.

This example illustrates some salient features of the procedure that can help users compare DEMs.

- 1) The evaluations table can hold both qualitative and quantitative evaluations; qualitative ones must be presented in such a manner that allows one to rank them accordingly;

- 2) Quantitative outcomes recorded in the evaluations table are translated to the opinions table according to the mid-rank procedure.
- 3) Numerical tolerances should be included in the application of quantitative criteria that reflect the uncertainty of the resulting values.

When applicable, tolerances should account for measurement uncertainty to identify minor differences in the evaluations table that will have an impact on the opinions table but are not really different. For example, results in the evaluations table are recorded to two decimal places, but many of these DEMs only record elevation to the nearest meter, and thus, differences of centimeters or even decimeters would not imply a significant difference among DEMs. Tolerances can also reflect chosen reference data comparison algorithms and

thematic requirements when justified to support the creation of the opinions table.

With the opinions table in-hand, the overall opinion ranking can be computed where  $R_j$  is the sum of the ranked value for a given DEM. Lower is better since the rank order is from 1 = best to  $k$  = worst. Sorting the  $R_j$  values might produce the requested ranking result. However, there is yet no statistical confidence associated with these rankings, so the next step is to apply the appropriate statistical tests.

3) *Integrating Statistical Confidence to the Ranking*: Statistical confidence must be associated with the final rankings because it is imperative that the outcomes produced are not due entirely to chance (as might happen if the opinions table rankings are taken naively).

One of the strengths of the RCBD method is that both quantitative and qualitative evaluations can be integrated into the contest. In addition, quantitative evaluations using different criteria need not be commensurate, thus precluding the ANOVA approach. Therefore, a nonparametric test should be used, which can operate over both types of rank results. The chosen nonparametric test is the Friedman test [45] because, among the nonparametric choices at hand, it is the best alternative for minimizing the risk of paradoxical results [46]. The general Friedman's statistic  $\chi_r^2$  is presented in (1), valid with or without ties ( $N$  = number of opinions;  $k$  = number of DEMs)

$$\chi_r^2 = \frac{N(k-1) \left[ \sum_{j=1}^k \frac{R_j^2}{N} - C_F \right]}{\sum_{i=1}^N \sum_{j=1}^k r_{ij}^2 - C_F}; \quad C_F = \frac{Nk(k+1)^2}{4}. \quad (1)$$

The entries in the opinions table (see Table IV, right-hand side) are denoted as the elements  $r_{ij}$ , and the final row presents the column sums by  $R_j$ . The remaining values to compute the Friedman statistic ( $\chi_r^2$ ) can be extracted from the same table: with  $C_F = 441$ , sum of  $r_{ij}^2 = 537.0$ , and sum of  $R_j^2 = 3192.0$ ; then,  $\chi_r^2 = 28.871$ .

For a given  $k$ ,  $N$ , and confidence level alpha, the Friedman's statistic value  $\chi_r^2$  is compared to the critical value  $\chi_{r,crit}$ ; if  $\chi_r^2$  is larger, the null hypothesis is rejected, and the conclusion is that we cannot accept that the DEMs are equivalent. Since we are willing to accept ties, the standard critical values tables from the Friedman test are not suitable. From the table for  $k = 6$  provided by López-Vázquez et al. [47] and López-Vázquez and Hochstain [48], and at the 95% confidence level, the row  $N = 6$  offers a critical value of  $\chi_{crit} = 10.489$ . Compared to the  $\chi_r$  result of 28.871, which is larger, one should reject the null hypothesis that the opinions table entries are purely random, implying that the DEMs are not equivalent.

The null hypothesis has been rejected based on the Friedman test, so there are statistically significant differences among the set of DEMs under consideration. The lower the values of  $R_j$ , the better, but it is still necessary to assess whether the difference between the pairs of ranked candidate DEMs is statistically significant or otherwise conclude that the pair is tied. The process is denoted as post-hoc analysis, and there exist different options to carry out such an analysis. In this case and following [49], we propose to use the test by Dunn [50] applying the Bonferroni correction.

A pair of DEMs is considered significantly different if

$$|R_i - R_j| \geq z_{1-\alpha/k/(k-1)} \sqrt{\frac{Nk(k+1)}{6}} \quad (2)$$

where  $z$  stands for the inverse of the cumulative normal distribution and  $\alpha$  is the prescribed confidence level.

4) *Step 5—Jupyter Notebook Implementation*: The final step ranks the candidate global DEMs. This ranking is based on statistical computations just described in order to provide confidence in the obtained result.

The needed functionality to read the database and implement the statistical procedures required to produce the final rankings was made available through a Jupyter Notebook [51] using Python version 3.10.x and the Pandas, Numpy, Matplotlib, and Seaborn libraries [52], [53], [54], [55].

The DEMIX Jupyter notebook can be run on a user's local machine or in the cloud using the Google Colab platform [56]. As input, this Jupyter notebook takes the GIS database produced by MICRODEM ignoring the signed mean and median statistics.

The information provided in the database allows the user to directly select opinions most appropriate to their requirements. For each run, the Jupyter notebook computes the final rankings based on the chosen opinions including the required confidence levels to support the outcomes. Furthermore, the DEMIX Jupyter notebook provides tools to analyze outputs by creating graphics and figures to help understand the final rankings. Some results are highlighted in Section III.

### III. RESULTS: APPLYING THE RCBD TO COMPARE GLOBAL DEMS

One goal of the DEMIX RCBD method is to provide DEM users with a procedure that can be applied anywhere for the intercomparison of global DEMs so that a choice can be made based on a final ranking that is statistically significant, taking into account criteria that are important to the user. For this reason, the procedure must be flexible in order to allow users to choose the criteria applicable to their particular requirements and deliver a practical outcome. The following results demonstrate these valuable characteristics. Furthermore, the results show what we think are generally valid rankings of these six DEMs.

The output of the Jupyter notebook is illustrated in Fig. 6. Fig. 6 (top) shows the overall ranking using all tiles and criteria based on the DTM and DSM reference DEMs, and the lower portion shows various filters of the database. The statistical significance of the ranking is presented in Fig. 6 by drawing a box around those global DEMs whose rank cannot be differentiated from a random result, i.e., the pairwise rank outcome does not pass the significance test as such ranking outcomes are essentially ties. For the overall set of evaluations (ALL land type), the DTM winner is FABDEM and the DSM is CopDEM, both at the 95% confidence level. For both, the DSM and DTM ALOS is ranked only slightly lower, and for some comparisons such as DSM with high average roughness, ALOS is tied with CopDEM for first place.

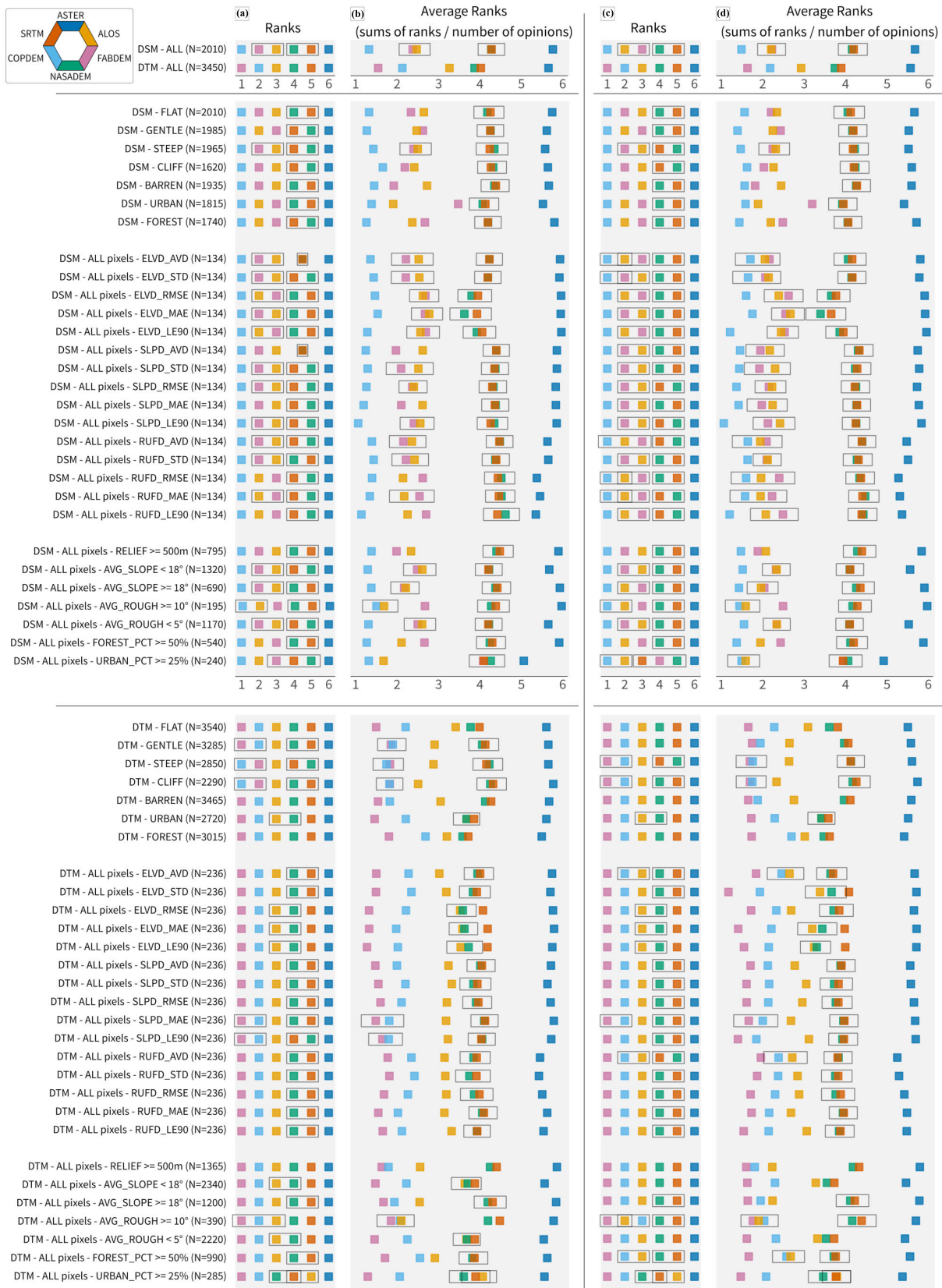


Fig. 6. Final rankings of DEMs based on the intercomparison of six 1'' global DEMs. The different outcomes are based on the chosen land types, the 15 contest criteria, and filters for tiles with specific characteristics. Tolerances for (a) and (b): ELVD = 0.5, SLPD = 0.5, and RUFSD = 0.2. Tolerances for (c) and (d): ELVD = 1.0, SLPD = 1.0, and RUFSD = 0.4; the increase in tolerances led to a greater number of ties in (c) and (d) compared to the rankings shown in (a) and (b).

In Fig. 6(a) and (c), the intercomparison rankings are ordered from best (1) to worst (6). From a statistical perspective, these rankings are ordinal and do not provide any information about how much better or worse any of the

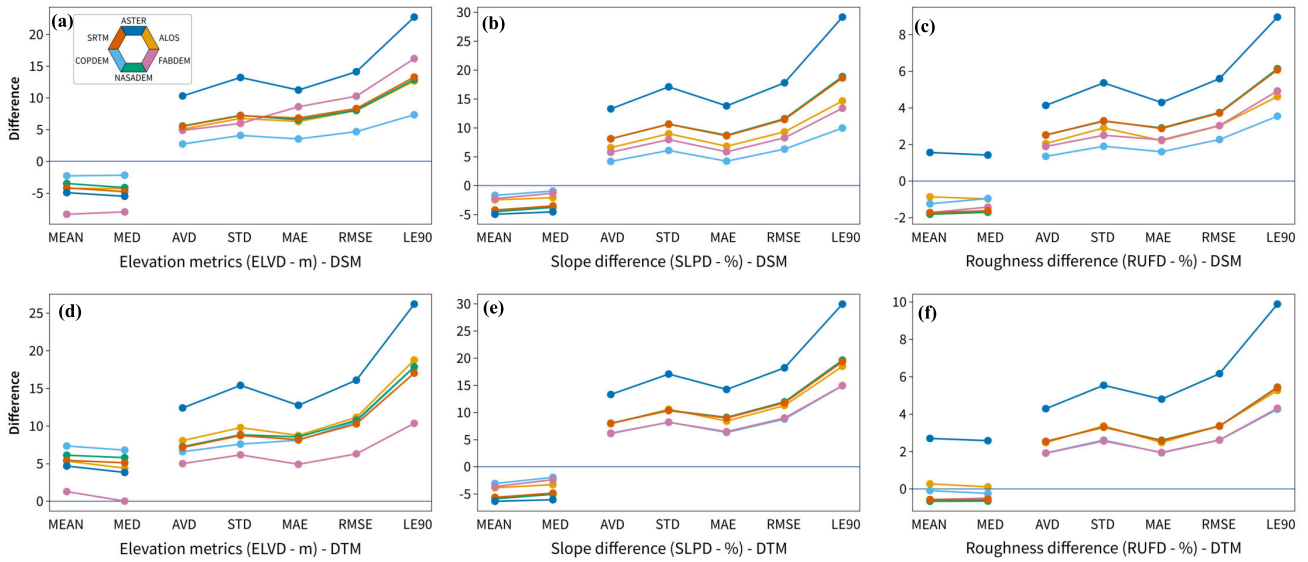


Fig. 7. Distribution of seven elevation (a) and (d), slope (b) and (e), and roughness (c) and (f) metrics comparing the difference between the test and reference DSMs (top) and DTMs (bottom) for tile N43PW002B in Spain. The mean and median on the left of each graph are signed; only the five rightmost parameters are used in the rankings. The lines connecting the points, while not statistically valid, show the effect on the parameters by their weighting of the tail of the distribution.

candidate DEMs are compared to each other. This is the expected output from DEMIX because it allows users to easily understand the best opinion based on their chosen criteria. A good impression of the variety of results is gained when analyzing the other contest rankings based on the subsequent rows. These have been grouped by reference DEM (DTM or DSM), land type filter (FLAT, GENTLE, FOREST, and so on), and criterion (ELVD\_RMSE, SLPD\_LE90, RUFDM\_AE, and so on). In different rows the order among DEMs changes. Some ties show up, with no clear preference among which DEMs are tied. This illustrates the power of the procedure in the context of the intercomparison of global DEMs because, depending on the user requirements, the “best” DEM will emerge based on the set of chosen criteria, far from a situation of one-option-fits-all.

To try and visualize the final ranking outcomes with a bit of context, Fig. 6(b) presents the same outcomes with the goal of showing how many times a test DEM ranks higher/lower over the number of opinions. The output is the sum of the ranks for any particular test DEM for each row in the opinions table (denoted as  $R_j$  in Section III), divided by the number of opinions. This visualization shows that the test DEM does not always have to be ranked number 1 in the opinions table to be the overall winner. For example, FABDEM is not the best DEM in very steep terrain. Columns (a)/(b) and (c)/(d) differ in showing the rankings on the left and the scores on the right, which provide a qualitative indication of how different the DEMs are. For instance, ASTER is always the worst, but it is much worse than NASADEM and SRTM, which are actually very close.

Fig. 6(c) and (d) demonstrate the effects that changing tolerances can have on the final rankings. The increasing of the tolerances produces more ties between the different global DEMs but actually does not produce a tie for the first place DEM very often. This is an important result for the intercomparison of global DEMs.

The results presented in Fig. 6 are not exhaustive and help to demonstrate the power and flexibility of the RCBM method for the intercomparison of global DEMs. The results also provide an important overview of the global DEM leaders based on the criteria used in this exercise: ELVDs, SLPDs, and RUFDMs.

#### A. DEMIX Tile Results

The final rankings presented in Fig. 6 integrate almost 40 000 opinions in the DEMIX database (see Table I), which have been produced from 236 DEMIX tiles in 24 areas. These visualizations support the overall results produced through DEMIX and provide specific outcomes that support the results of the intercomparison of the global DEMs.

Fig. 7 displays seven metrics based on elevation [see Fig. 7(a) and (d)], slope [see Fig. 7(b) and (e)], and roughness [see Fig. 7(c) and (f)] differences between the reference and test DEMs for tile N43PW002B in Spain. The first two results displayed in each case (left-hand side of each graph) are the mean and the median differences. These signed results cannot be ranked and, therefore, are not directly applicable; however, they provide valuable summary statistics for the tile under consideration. Specifically, they show how candidate DEMs compare to the reference DEM in terms of the direction of the differences: positive or negative. The remaining results are those used in the preparation of the evaluations table. Note that we show only one tile as a representative scenario of the metrics. While this is not a standard graphic, we show it for two important reasons. First, there is little crossing of the lines, and the ranking of the DEMs is the same regardless of the criterion used—ASTER is always the most different from the reference DEM. Second, popular metrics such as MAE, RMSE, and LE90 produce different numerical results but consistently lead to the same ranking of the DEMs.

For the signed results, one observes some negative bias for results based on the elevations (compared to a reference DSM),

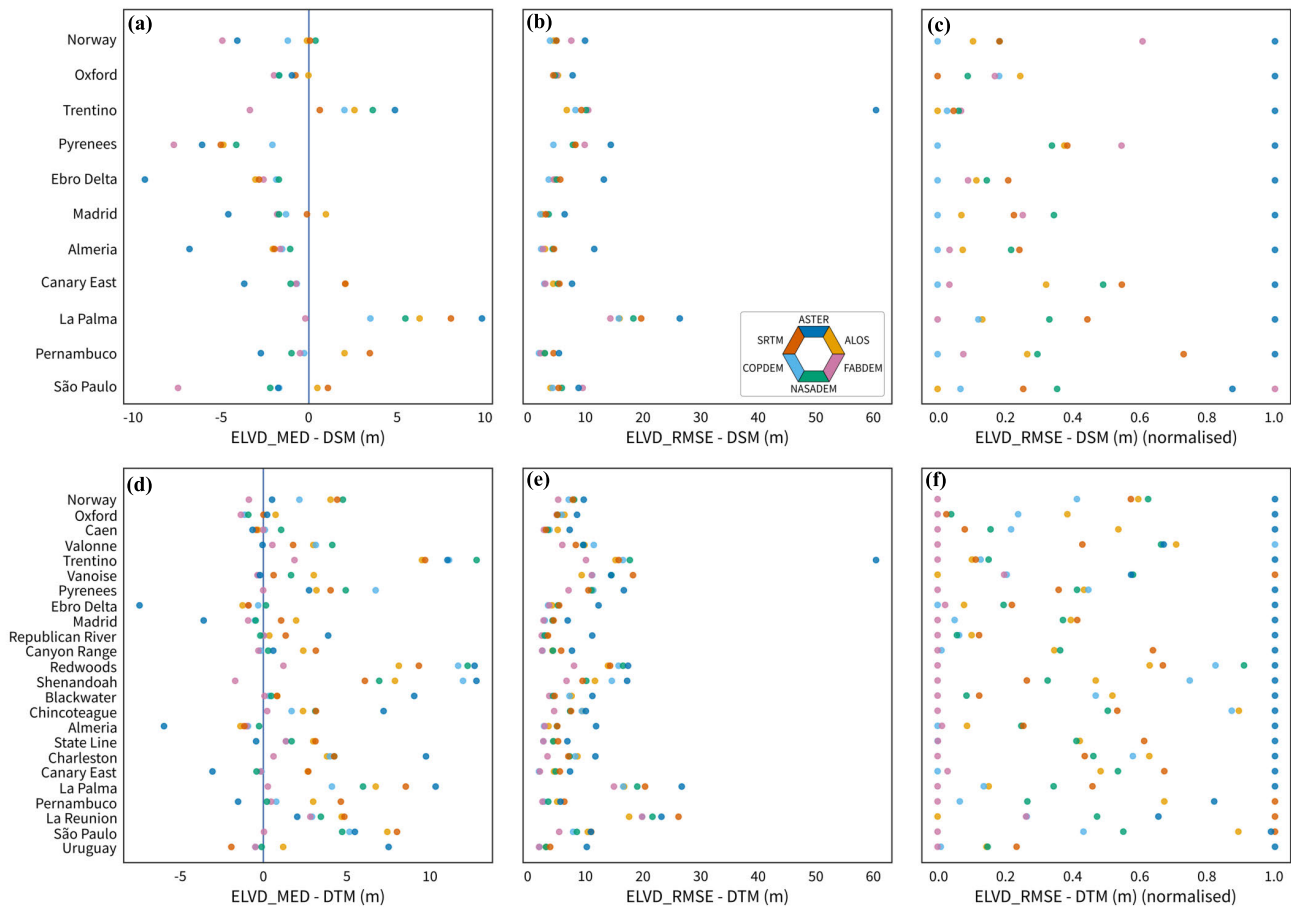


Fig. 8. Distribution of two signed metrics for all areas. (a) and (d) ELVD\_MED. (b) and (e) ELVD\_RMSE. (c) and (f) ELVD\_RMSE with normalized values. Negative values for the median indicate that the DEM is lower in elevation than the reference DEM, while positive values indicate that the DEM has higher elevations than the reference DEM. The areas are arranged by latitude from north to south.

slope, and roughness parameters. ASTER results tend to be outside of the group of results being an outlier in terms of not corresponding with the reference DEM compared to the other global DEMs, as clearly shown in Fig. 6.

The five unsigned metrics used as criteria show similar trends across the chosen tile and within each geomorphometric parameter. It would, however, be difficult to choose the best global candidate DEM based only on these graphs, especially with any statistical confidence.

Fig. 8(a) and (d) shows the medians of the ELVD distributions for the 24 test areas, which average the results from 1 to 35 tiles in each area, highlighting a general negative bias for results based on the elevations compared to a reference DSM and a positive bias compared to the DTM. Fig. 4 showed histograms with the distribution of elevation, slope, and roughness for a single tile, and Fig. 7 showed all seven parameters that we computed for a different tile with elevation, slope, and roughness parameters. In all of these figures, and in virtually all comparisons that we ran, ASTER results tend to be well outside the group of other results.

Based on the DEMIX final rankings, it is apparent that CopDEM and FABDEM performed best in the global DEM intercomparison, with ALOS being a close third. However, such a conclusion could not be readily made based only on the visualization of the different metrics, especially as the

number of locations increases. Unlike typical DEM comparison exercises, the DEMIX RCBD method provides statistical significance to the conclusions.

#### IV. DISCUSSION

This article has three objectives: 1) introduce the DEMIX RCBD method; 2) validate an open-source tool chain to apply the methodology; and 3) apply the method for the first time to an intercomparison of six global 1" DEMs over a wide range of test areas with general recommendations on which of the DEMs performs best.

The results of DEMIX produced a ranking with the prescribed confidence level of the global DEMs based on a set of quantitative criteria. The quantitative assessments were based on popular metrics applicable to intrinsic and derived parameter values of DEMs, such as elevation, slope, and roughness.

The different contest final rankings produced a variety of results depending on the choice of criteria. For the DTM rankings (see Fig. 6), the leaders (top three) in order were FABDEM, CopDEM, and ALOS. For some filtered criteria choices in the lower part of Fig. 6, there was also a number of ties between these top three. NASADEM even moved into ties for third place in some particular cases where ALOS does not compare well with FABDEM and CopDEM. Overall

NASADEM slightly improved SRTM but only for limited land types and mostly for elevation metrics and not slope or roughness. For the DTM (lower tolerances) contests, ASTER, NASADEM, and SRTM were always in the lower three places with ASTER always being clearly last.

The DSM rankings results differ from those of the DTM rankings both overall and across the different criteria. While the overall leader was CopDEM, ALOS and FABDEM share second and third places, sometimes tied. NASADEM and SRTM generally tied for fourth place; however, ASTER remained in sixth place throughout the DSM contests as well.

Varying tolerances [see Fig. 6(c) and (d)] generally increased the number of ties but did not change the general trends observed in the DSM and DTM contest and barely changed the first place results of CopDEM compared to a reference DSM and FABDEM compared to a reference DTM.

#### A. Comparing the DEMIX Method and Rankings to Previous Work

This work differs quite significantly from previous studies in a number of ways and most importantly in the total combination of factors that were applied for the intercomparison and analysis of the candidate global DEMs. The following list compares and contrasts the DEMIX method to previous work.

- 1) Computed distributions of the differences between the reference and global DEMs for elevation, slope, and roughness. Many earlier studies just compare the distributions of the DEM with respect to the distribution of reference elevation data. Furthermore, the derived geomorphometric parameters in this study increase the effect of differences among the DEMs, and the derived characteristics are often more important than the original DEM elevations for many applications.
- 2) Integrated seven parameters for each difference distribution—most previous studies integrate fewer parameters typically only RMSE and MAE [14]. While the chosen seven parameters are correlated, all the relationships are included in the database and show how little difference the choice makes.
- 3) Only 1" DEMs were considered; changes in DEM spacing impact DEM characteristics making reliable inter-comparisons across multiple resolutions difficult. There are enough good quality global 1" DEMs applicable to land-based processes that one should only consider 3" DEMs such as MERIT [25] when lower spatial resolution is explicitly required or if the user feels that MERIT resembles a better DTM than FABDEM aggregated to 3".
- 4) Created reference 1" DEMs from the fine spatial resolution source reference DEMs based on open data from national/local mapping agencies. These DEMs have significantly better spatial resolution and detail compared to the target 1" resolution. However, the choice to use 1" reference DEMs requires accepting this resolution and any limitations that it imposes on the subsequent analysis.
- 5) No reinterpolation of the candidate global DEMs was made after downloading other than adjusting to the EGM2008 vertical datum. Interpolation cannot improve a DEM or create additional elevations that were not captured in the original data.
- 6) All computations were performed in native geographic coordinates [36]. While there are differences in the statistics between the rectangular 1" pixels and square 30-m pixels, interpolating to UTM coordinates adds its own changes to computations done on the DEM, which we avoided.
- 7) Created reference DEMs to consider the half-pixel offset between ALOS and CopDEM (and the other lower quality DEMs) for the intercomparison, a consequence of the decision not to resample candidate DEMs. This issue is generally not discussed in the literature (or addressed by interpolation) although Purinton and Bookhagen [57] designed their comparisons to avoid the need for DEM coregistration.
- 8) Both DSM and DTM reference data are considered separately. It turns out that, in most of our test areas, those not dominated by urban or forested land cover, the DTM and DSM vary little. Some of the minor differences arise from the different characteristics of the source data for the reference DEM: for example, in Spain, these have different resolutions of 2 and 5 m, respectively, and were created at different times.
- 9) The intercomparison was pixel-by-pixel at every point in the sample tiles, providing approximately 130 000 points per tile. This is a significantly greater number of points compared to studies using geodetic benchmarks, airport runways, or even measurements from laser altimeters, such as Global Ecosystem Dynamics Investigation (GEDI) and ICESat-2. In addition to providing often orders of magnitude more points for the inter-comparisons, the pixel-by-pixel option provides data in both steep and forested parts of the test area, where point-based reference data are often not available [58].
- 10) A large number of diverse test areas—in comparison, for example, Uuemaa et al. [59] reported using 608 million 1-m pixels (after interpolation of the 1" test DEMs), which corresponds to only six DEMIX tiles. Purinton and Bookhagen [57], on the other hand, used the equivalent of 400 contiguous DEMIX tiles, but from a single mountainous test area.
- 11) The DEMIX intercomparison contest covered a wide range of land types (see Fig. 2), which, combined with a large number of test areas, greatly increases the validity of the presented results. In contrast, some studies, such as [28] and [60], focused their analysis primarily on coastal or floodplain terrain, which probably accounts for our finding that FABDEM often ranked lower than the CopDEM from which it was derived, particularly in mountainous terrain. At the other elevation extreme in mountain areas, Purinton and Bookhagen [57] preferred CopDEM, while Florinsky et al. [24] and Trevisani et al. [61] found that relative performances between CopDEM and ALOS were dependent on slope and roughness, with, in general, better performance of CopDEM except for very steep slopes.

Previous work did not quantify the ranking as developed within the DEMIX RCBD method. Our approach quantifies the intercomparisons with additional metrics beyond RMSE or MAE and also considers elevation, slope, and roughness. Furthermore, the DEMIX methodology was applied to a representative land type sample covering a total area of 23 600 km<sup>2</sup> across three different continents. The closest approach to a DEMIX methodology applied to DEM characterization is shown in [62, Table IV], where they created a “confusion matrix” that ranked the DEMs from best to worst on four criteria.

### B. DEM User Driven Contest

A DEM user understands how to apply a DSM or DTM in their application but is not necessarily familiar with how different DEMs were created or the different characteristics that can affect a user’s DEM choice. The DEMIX methodology was developed to help DEM users make more informed decisions about the elevation data that they plan to use for their particular domain. This is a user-oriented method where experts can contribute novel criteria or provide reference data to be shared with the community to improve user choices. The outcomes from this intercomparison experiment demonstrate this ability of the method through the numerous rankings of the different criteria and regions of interest.

After DEMIX was started, it soon became evident that equally challenging to test and compare DEMs at a global scale would be to present the volume of results generated. Raw quality and validation figures alone usually require expert judgment on whether and to what extent a DEM is suitable for a certain application. The methodology offers a transparent and versatile solution to this problem. Not only does each criterion require that the results can be ranked but also the real advantage to users comes with the possibility to combine criteria and test areas in any sensible way. This produces a statistically sound recommendation about which DEM to use that can be tailored to intended uses. For example, DEM experts could advise on which criteria to use and users could then rank available DEMs for the area or type of terrain that they are most interested in.

Our study gives a first-ordered ranking of DEMs. As further criteria and test areas are added to the database by the community, more comprehensive rankings and coverage of additional applications will be possible. However, we expect our general conclusions to remain valid.

How can DEM users take advantage of this wealth of scientific information? The DEMIX initiative is rectifying this issue with a methodology that provides open access to high-quality scientific results and provides DEM users:

- 1) the ability to choose regions and locations based on standardized sampling areas (DEMIX tiles);
- 2) the ability to allow users to choose relevant criteria to characterize and rank their DEMs of interest;
- 3) the possibility to change the number of tiles and the criteria to produce new rankings among their DEMs of interest for comparison.

#### 1) Thematic Applications of the DEMIX Methodology:

The DEMIX methodology, including its flexibility and the

possibility to evaluate DEM quality according to ad hoc chosen criteria, can be relevant in a multitude of application contexts, such as the study of geoenvironmental processes, including the critical area of natural hazards. In the analysis and modeling of natural hazards, DEM derivatives are routinely adopted as geoenvironmental proxies or as input features in supervised learning approaches [63], [64], [65]. For example, a basic terrain variable such as slope represents a key factor in landslide susceptibility models [66], [67], as well as in the context of seismic hazard, where it has been adopted for the derivation of a proxy of shear wave velocity [68], [69]. It is clear that the possibility to evaluate and compare which DEM is better according to the representation of slope can provide insightful and useful suggestions for the users. For example, the results of the present DEMIX experiment suggest that CopDEM and ALOS should be preferred over the usually adopted SRTM and ASTER DEMs. The same discussion can be true in many other Earth science-related issues and topics such as fluvial morphology [70], geodiversity studies [71], flood hazards [72], and other domains.

No temporal criteria were taken into account during this first ranking of global DEMs, such as the date on which the measurements were taken or the length of time that it took to acquire the global coverage. Temporal differences between global DEMs could be relevant to thematic applications, especially if those changes occurred in the last 20 years and produced significant changes to the surface. Examples where significant elevation change could have occurred include volcanic eruptions, landslides, deforestation, urbanization, open pit mining, or severe erosion. The comparison should be carried out over comparable DEMs, and users must decide on what is comparable. If the user thinks that higher quality, newer DEMs are not representative of temporal changes under investigation, they can simply use the older DEMs. DEMIX proposes a paradigm shift, moving the focus from data producers to users. Traditional comparison exercises were expected to produce a definitive ranking, likely to be carved in stone. Now, we expect to bring the power to the user, who will be able to justify their own choices with sound statistical arguments.

DEM users should also be aware that reference data used in this work were not acquired at the same time as the global DEMs under consideration. The data collection date for the reference data used is provided in the metadata.

2) *Contest Criteria Ranking Requirements:* One of the powerful features of the methodology is the fact that users can define their own criteria. The only requirement is that the chosen metric must be designed and defined in such a way that the final outcomes can be ranked from best to worst/most appropriate to least appropriate. For example, in the above discussion, it was mentioned that differences in elevation between the test and reference DEM, even if they contain useful information, cannot be used directly for ranking because these are signed quantities.

For this reason, five criteria were based on the absolute differences of elevation instead. Strictly speaking, the adoption of the absolute value of differences implies a symmetrical weighting scheme in which the same importance is given to positive and negative ELVDs. However, in some applications,

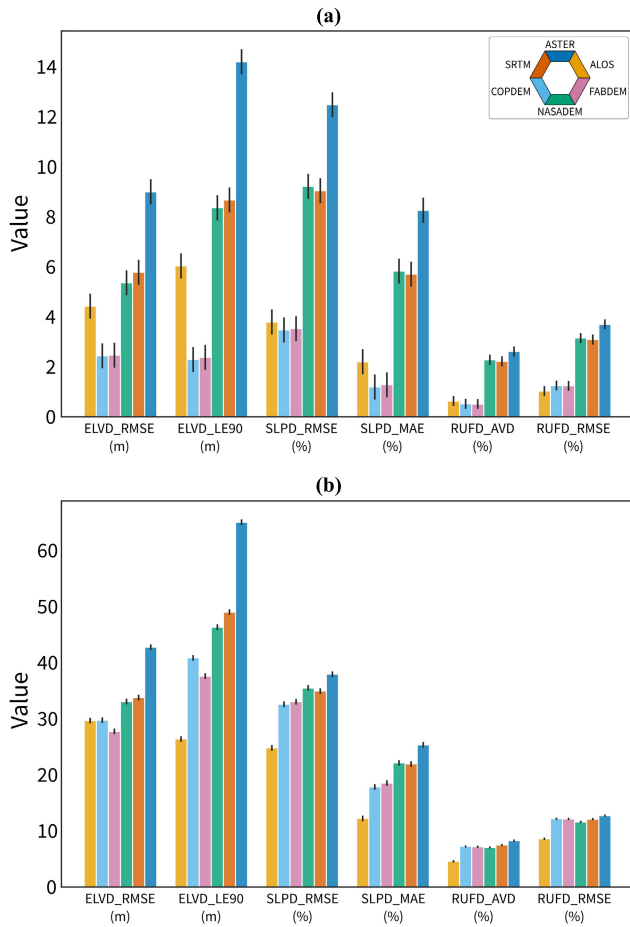


Fig. 9. Bar plots of the evaluation results for six criteria with two tiles compared to a reference DTM. The small black line atop each bar shows the tolerance that we used for that criterion and shows that the evaluation of another candidate DEM would need to be tied. (a) DEMIX tile—N35VW116G. (b) DEMIX tile—N28XW018B.

such as coastal flood risk analysis [73], [74], negative and positive errors can impact our evaluations differently, for example, in terms of cost functions. Accordingly, signed differences can be transformed into unsigned and ordered rank quantities, weighting positive, and negative errors differently (i.e., non-symmetrically), taking advantage of a suitable transforming function.

A major rationale for developing the methodology, and indeed for informally calling it a wine contest, was the ability to include subjective criteria from expert judges. While this study did not include any subjective criteria, Guth et al. [75] compared hillshade maps of these six DEMs in three of the DEMIX tiles included here. In addition to validating our conclusions about the relative rankings of these DEMs, they point out the logistical testing challenges in setting up a subjective ranking for just three tiles, let alone the 236 tiles used for this work.

Figs. 7 and 9 provide some guidance on the selection of criteria and the tolerances for accepting ties. In Fig. 7, the criteria used for the intercomparison on the right are separated from the signed means and medians. The five unsigned criteria are arranged from left to right in rough order by the effect of outliers; LE90 always has a much larger value compared to the

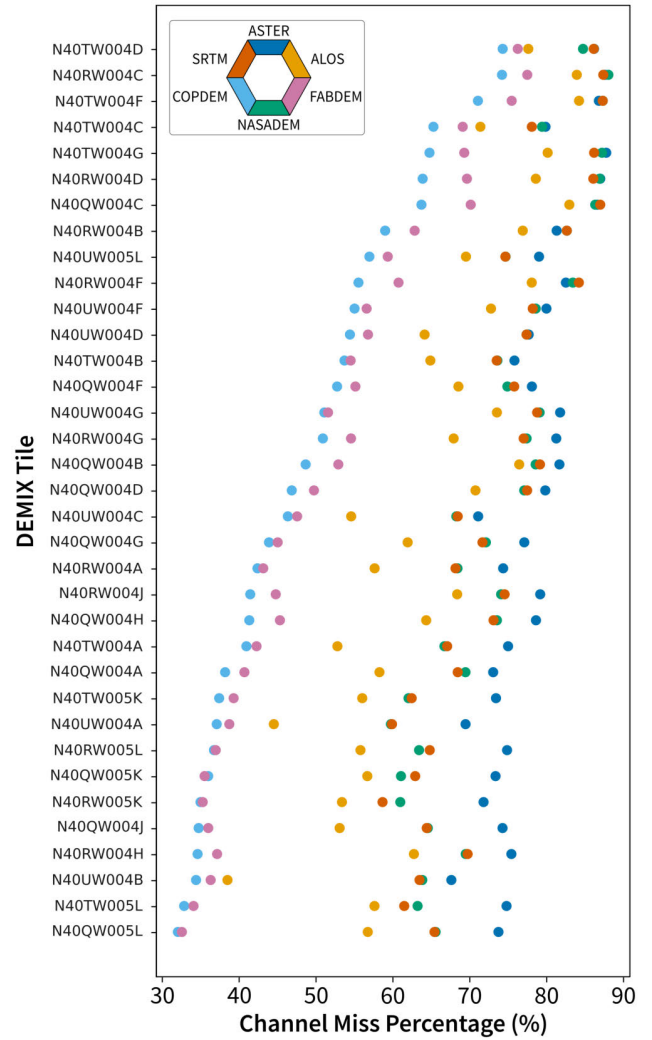


Fig. 10. Channel location comparison computed as the channel miss percentage for 35 DEMIX tiles in the Madrid (Spain) test area. A low score is better: 0% means that the channels derived from the candidate DEM perfectly match the reference DTM, and 100% means that they do not match at all.

TABLE V  
POSSIBLE OUTCOMES OF THE BINARY PATTERN MEASURES COMPUTED BETWEEN THE PIXELS IN THE REFERENCE CHANNEL DATASET AND THE SIX GLOBAL DEMS

	Test DEM channel	Test DEM inter-channel
Reference DEM channel	Correct channel	Underprediction (misses)
Reference DEM inter-channel	Overpredictions (false positive)	Correct inter-channel

others. With rare exceptions, the lines connecting the values do not cross, and the rankings for the criteria are the same. The tolerances used for ties apply to the difference scale on the vertical axis; the amount of “white” space between the lines shows the tolerance that would be required to consider the DEMs tied. In most cases, visual assessment of graphs like this reinforces our conclusions about the rank ordering of the DEM: CopDEM and FABDEM, ALOS, NASADEM and SRTM, and, finally, ASTER.

Fig. 9 shows the evaluation results for six criteria and two different tiles. Note the very different vertical scales; tile



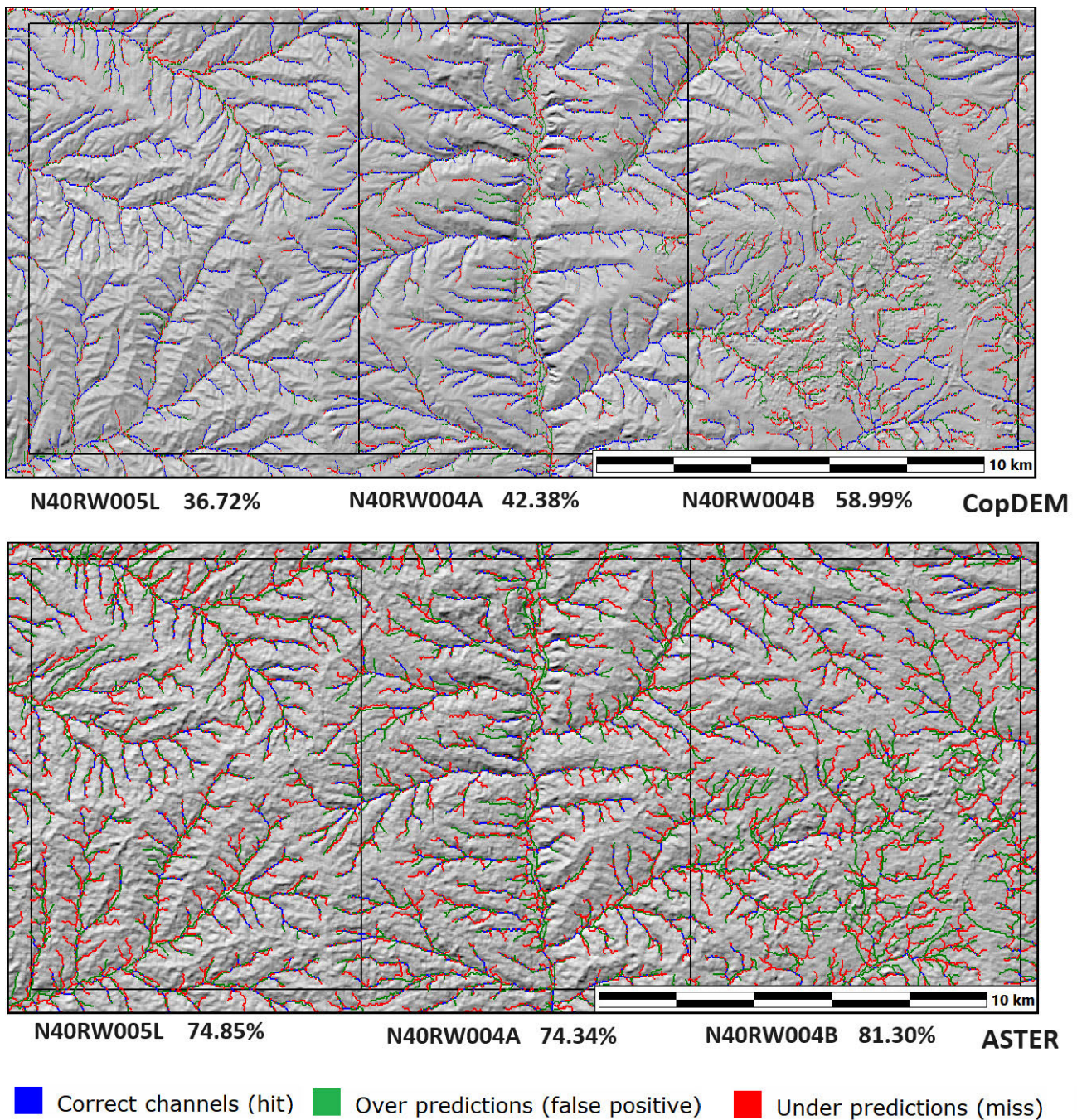


Fig. 11. Three tiles in the Madrid (Spain) test area, showing the miss percentages (low score best) for the channel networks on hillshades for (top) CopDEM and (bottom) ASTER.

N28XW018B in the Canaries is very steep and rough. Scores for the different criteria are also very different although this also depends on the local topography. Salient results from this graph are that most of our results are immune to any reasonable choice of tolerances or which criteria are used, and our ranking of the DEMs is robust. Tile N28XW018B is anomalous in showing ALOS as the best DEM for five of the six criteria; the results for tile N35VW166G are more typical (see Table IV).

*C. Practical Application—Hydrology*

Most users want to use DEMs for a wide range of practical applications, ranging from modeling landslide susceptibility

and making flood models, to creating hillshade base maps. Statistical measures such as the 15 criteria used in our rankings do not always intuitively relate to what users want to apply, and most comparisons in the scientific literature use only a few specific metrics comparing elevations to reference data. We have argued that our combined assessment of elevation, slope, and roughness of the DEMs results in a more comprehensive and broadly applicable ranking of global DEMs compared to using only elevation metrics focusing on small regions.

To demonstrate the relevance of the wine contest results, we briefly look at a common application of DEMs in this section and show that the global DEM rankings produced

TABLE VI  
FREE QUASI-GLOBAL DEMS AT 1" (AFTER [27])

DEM	Primary Source	Producer	Vertical Datum	Precision	Grid Storage	Longitudinal Spacing	Acquired
SRTM v3	C band radar	NASA	Orthometric EGM96	Integer	Pixel-is-Point	Constant	2000 (11 days)
ASTER GDEM v3	Stereo NIR imagery	NASA/METI	Orthometric EGM96	Integer	Pixel-is-Point	Constant	2000-2013
ALOS World 3D AW3D30 v3.2	Stereo pan imagery	JAXA	Orthometric EGM96	Integer	Pixel-is-Area	Variable	2006-2011
NASADEM	Reprocessed C band radar	NASA	Orthometric EGM96	Integer or floating point	Pixel-is-Point	Constant	2000 (11 days)
Copernicus DEM GLO30	X band radar, Edited WorldDEM	ESA/Airbus	Orthometric EGM2008	Floating point	Pixel-is-Point	Variable	2011-2015
FABDEM v1-2	Edited Copernicus DEM		Ellipsoidal WGS84	Floating point	Pixel-is-Point	Constant	2011-2015

can also apply to the creation of a drainage channel network. For this demonstration, the test area contains 35 DEMIX tiles located near Madrid, Spain. Because channel networks cover larger areas than the 100-km<sup>2</sup> DEMIX tiles and channels can come into and out of individual tiles, using larger areas provides a better channel network by considering more of the full contributing drainage area. The SAGA (version 9.3.0) open-source software for geospatial analysis [76], [77] was used with two modules: “Sink Removal” and “Channel network,” accepting all default parameters and using as input our 1 arc second reference DTMs derived from the Spanish national 2-m data [39] for both pixel-is-point and pixel-is-area geometries. The resulting vector channel network was rasterized to a 1 arc second grid in MICRODEM filtering to the channels with PLOT\_ORDER > 1.

Table V shows the four categories resulting from the analysis and how the hit percentage [Correct channel/(Correct channel + Underprediction)] can be used to assess the results [78]. In Figs. 10 and 11, we use a missing percentage (100—hit percentage) where lower is better, which matches the criteria used in the wine contest (see Figs. 6 and 8).

Fig. 10 shows the results of comparing the river networks produced using the six global DEMs to the reference DTM and plotting the channel pixel miss percentage (low score is better). The relative order of the outcomes and the separations between the six global DEMs are constant for all 35 tiles. Differences among the tiles reflect the different terrain characteristics.

A visual representation of the river network-based comparison is shown for three tiles in Fig. 11. The top plot (CopDEM) shows the consistently lowest miss percentages (best depiction of the channel network), and the bottom plot (ASTER) presents an example of the DEM, which consistently has the highest miss percentage (worst channel network). The easternmost tile, N40RW004B, is largely urban and has the largest miss percentage compared to the reference DTM. Flat areas are a known problem in channel extraction [79], and all DEMs are not good as seen in the urban tile on the right and the main river valley in the center tile. Even in the hilly areas, the ASTER DEM tiles that are shown miss more of the small order streams. In addition to the channel networks, the hillshade from ASTER is much rougher and does not show the ridges and valleys, as well as that from CopDEM.

The global DEM intercomparison that produced the rankings in this article did not directly consider hydrology. The example presented in this section based on channel networks in the Madrid area closely matches the rankings of the six global DEMs in our holistic set of elevation, slope, and roughness

criteria. A qualitative assessment of hillshade maps for several DEMIX tiles had similar results and ranking of the six global DEMs [75]. Both the channel network and hillshade results are based on significantly smaller sample sizes, but we find the outcomes support the wine contest DEM intercomparison rankings (see Fig. 6), which we expect to be generally valid for a broad range of applications. This hydrology experiment verifies the applicability of the global wine contest rankings beyond having better RMSE or LE90 statistical values, and the ranking suggests that most practical applications will benefit from using the DEMs recommended by our rankings.

## V. CONCLUSION

This article presents the novel “DEMIX wine contest methodology,” for the intercomparison of DEMs, which produces a final ranking with prescribed confidence levels based on given criteria. We applied the method to six global 1" DEMs: ALOS, ASTER, CopDEM, FABDEM, NASADEM, and SRTM. The intercomparison was done using 15 criteria related to elevation, slope, and roughness measures derived from reference 1" DEMs. The final rankings with confidence level based on the choice of criteria and land type demonstrate the powerful features of this method, including the ability for the user to choose the most relevant criteria and areas.

The main outputs from the DEMIX global DEM intercomparison are presented in Fig. 6 and further supported by other figures included in the text to provide the experts and users transparency with respect to the following conclusions.

- 1) From an overall final ranking of the global 1" DEM intercomparison, CopDEM and FABDEM are clearly the frontrunners based on the chosen criteria and test areas: CopDEM compared to a DSM and FABDEM compared to a DTM. The best rankings depend on the choice of parameters and what the user wants to do with the DEM. ALOS was generally the third best but sometimes moving into the second place.
- 2) FABDEM improves on CopDEM in most cases where a DTM is required, except for steep terrain that was not an emphasis in its generation.
- 3) SRTM and NASADEM are distinctly in the lower half of the wine contest rankings indicating lower quality than the top three. Consequently, unless compelling reasons (such as surface changes) are presented, we encourage users to move away from using them. While SRTM revolutionized global digital topography, NASADEM only produced modest improvements and actually decreased scores for some criteria. The improvements in

NASADEM are mostly for the elevation measures, and it did little to change the derived slope and roughness measures.

- 4) As many prior studies have also shown, the ASTER DEM is clearly the lowest performer.
- 5) Certain areas show quite anomalous DEM rankings. For example, ALOS on average was the best in about 8.4% of the opinions, but, in four areas (Vanoise, São Paulo, La Reunion, and La Palma), it was better at least twice as often. This supports the DEMIX goal to help DEM users choose appropriate products for their areas of interest and the criteria most relevant to their usage.
- 6) All the global 1" DEMs that were evaluated by DEMIX are composite, i.e., they required other data sources for in-filling gaps because the original data acquisition did not provide full coverage. For example, CopDEM makes extensive use of high-quality national LiDAR data for infilling and improving TanDEM-X data over Norway. Results in this area are, therefore, not representative of most other test areas.
- 7) The three high-quality DEMs (CopDEM, FABDEM, and ALOS) are all global, and there is no longer an issue with the lack of coverage at high latitudes. However, the local availability of the main instrument data varies significantly among ALOS, CopDEM, and FABDEM and may affect their spatial consistency.
- 8) The three global high-quality DEMs are much closer in time of collection: CopDEM and FABDEM: 2010–2015 and ALOS: 2006–2011. Except for SRTM, all data used to produce the global DEMs were collected over a number of years. The reference DEM data used here also come from a wide range of dates. Global DEM users requiring data from 2000 can take advantage of these results to decide whether SRTM or NASADEM meets their elevation needs.

The breakout of the six global 1" DEMs into two groups corresponds to the technologies and sensor characteristics used to acquire the elevation data. SRTM, NASADEM, and ASTER are delivered having a pixel size of 1", but, from the beginning, it was recognized that the actual resolution did not match comparable cartographic DEMs and that the "real" resolution was coarser than 1" [6], [9], [80], [81]. In contrast, ALOS and CopDEM are derived from commercial DEM products having resolutions of 0.15" (~5-m spatial resolution) and 0.4" (~12 m), respectively. Consequently, ALOS and CopDEM have a resolution appropriate to their pixel size and, therefore, better match the reference DEMs.

In this experiment, CopDEM consistently ranks the best compared to the reference DSM, and FABDEM ranks the best compared to a DTM. Guth and Geoffroy [21] showed that CopDEM consistently had elevations tightly clustered near the center of the LiDAR point cloud and, thus, was intermediate between a DSM and a DTM. The other DEMs investigated in that work had much greater dispersion in the point cloud. Histograms from the difference distributions (see Fig. 4) for all the tiles that we examined support the general applicability of this generalization, and it applies to slope and roughness, as well as elevation.

## APPENDIX A GLOBAL DEM CHARACTERISTICS

See Table VI.

## APPENDIX B GLOBAL DEM DATA SOURCES

The following list contains the names of all candidate DEMs evaluated in this article together with a web reference for access to further information and download of the data.

- 1) *SRTM*: <https://earthexplorer.usgs.gov>, [https://search.earthdata.nasa.gov/search/granules?p=C1000000240-LPDAAC\\_ECS](https://search.earthdata.nasa.gov/search/granules?p=C1000000240-LPDAAC_ECS).
- 2) *ASTER GDEM*: <https://ssl.jspacesystems.or.jp/ersdac/GDEM/E/>, <https://search.earthdata.nasa.gov/search/granules?p=C1711961296-LPCLLOUD>.
- 3) *ALOS AW3D30*: <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm>.
- 4) *NASADEM*: [https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM\\_SHHP.001](https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM_SHHP.001), [https://search.earthdata.nasa.gov/search?q=C1546314043-LPDAAC\\_ECS](https://search.earthdata.nasa.gov/search?q=C1546314043-LPDAAC_ECS).
- 5) *Copernicus DEM*: <https://doi.org/10.5270/ESA-c5d3d65>.
- 6) *FABDEM*: <https://doi.org/10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn>.

## ACKNOWLEDGMENT

Support for the Digital Elevation Model Inter-comparison eXercise (DEMIX) project was provided by the European Space Agency (ESA) in the framework of the EDAP+ project. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## REFERENCES

- [1] T. G. Farr et al., "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, no. 2, Jun. 2007, Art. no. RG2004, doi: [10.1029/2005rg000183](https://doi.org/10.1029/2005rg000183).
- [2] D. B. Gesch, K. L. Verdin, and S. K. Greenlee, "New land surface digital elevation model covers the Earth," *Eos, Trans. Amer. Geophys. Union*, vol. 80, no. 6, pp. 69–70, Feb. 1999, doi: [10.1029/99eo00050](https://doi.org/10.1029/99eo00050).
- [3] D. B. Gesch, J. P. Müller, and T. G. Farr, "Special Issue Foreword: The shuttle radar topography mission—Data validation and applications," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 3, pp. 233–235, 2006.
- [4] E. Rodríguez, C. S. Morris, and J. E. Belz, "A global assessment of the SRTM performance," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 3, pp. 249–260, Mar. 2006, doi: [10.14358/PERS.72.3.249](https://doi.org/10.14358/PERS.72.3.249).
- [5] M. Hofton, R. Dubayah, J. B. Blair, and D. Rabine, "Validation of SRTM elevations over vegetated and non-vegetated terrain using medium footprint LiDAR," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 3, pp. 279–285, Mar. 2006, doi: [10.14358/pers.72.3.279](https://doi.org/10.14358/pers.72.3.279).
- [6] P. L. Guth, "Geomorphometry from SRTM," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 3, pp. 269–277, Mar. 2006, doi: [10.14358/pers.72.3.269](https://doi.org/10.14358/pers.72.3.269).
- [7] C. C. Carabajal and D. J. Harding, "ICESat validation of SRTM C-band digital elevation models," *Geophys. Res. Lett.*, vol. 32, no. 22, Nov. 2005, Art. no. L22S01, doi: [10.1029/2005gl023957](https://doi.org/10.1029/2005gl023957).
- [8] M. Abrams, B. Bailey, H. Tsu, and M. Hato, "The ASTER Global DEM," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 4, pp. 344–348, 2010.
- [9] H. I. Reuter, A. Neison, P. Strobl, W. Mehl, and A. Jarvis, "A first assessment of aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2009, pp. 240–243, doi: [10.1109/IGARSS.2009.5417688](https://doi.org/10.1109/IGARSS.2009.5417688).
- [10] P. L. Guth, "Drainage basin morphometry: A global snapshot from the shuttle radar topography mission," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 7, pp. 2091–2099, Jul. 2011, doi: [10.5194/hess-15-2091-2011](https://doi.org/10.5194/hess-15-2091-2011).

- [11] J. A. Slater et al., "Global assessment of the new ASTER global digital elevation model," *Photogrammetric Eng. Remote Sens.*, vol. 77, no. 4, pp. 335–349, Apr. 2011, doi: [10.14358/pers.77.4.335](https://doi.org/10.14358/pers.77.4.335).
- [12] B. Wessel, *TanDEM-X Ground Segment—DEM Products Specification Document*, document TD-GS-PS-0021, Issue 3.1, 2016. [Online]. Available: <https://tandemx-science.dlr.de/>
- [13] P. Rizzoli et al., "Generation and performance assessment of the global TanDEM-X digital elevation model," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 119–139, Oct. 2017, doi: [10.1016/j.isprsjprs.2017.08.008](https://doi.org/10.1016/j.isprsjprs.2017.08.008).
- [14] C. H. Grohmann, "Evaluation of TanDEM-X DEMs on selected Brazilian sites: Comparison with SRTM, ASTER GDEM and ALOS AW3D30," *Remote Sens. Environ.*, vol. 212, pp. 121–133, Jun. 2018, doi: [10.1016/j.rse.2018.04.043](https://doi.org/10.1016/j.rse.2018.04.043).
- [15] P. Strobl, "The new Copernicus digital elevation model," *GSICS Quart.*, vol. 14, no. 1, pp. 1–20, 2020, doi: [10.25923/enp8-6w06](https://doi.org/10.25923/enp8-6w06).
- [16] T. Tadono et al., "Generation of the 30 M-mesh global digital surface model by ALOS PRISM," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, pp. 157–162, 2016, doi: [10.5194/isprsarchives-XLI-B4-157-2016](https://doi.org/10.5194/isprsarchives-XLI-B4-157-2016).
- [17] R. Crippen et al., "Nasadem global elevation model: Methods and progress," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLI-B4, pp. 125–128, Jun. 2016, doi: [10.5194/isprs-archives-xli-b4-125-2016](https://doi.org/10.5194/isprs-archives-xli-b4-125-2016).
- [18] K. Zhang et al., "Accuracy assessment of ASTER, SRTM, ALOS, and TDX DEMs for Hispaniola and implications for mapping vulnerability to coastal flooding," *Remote Sens. Environ.*, vol. 225, pp. 290–306, May 2019, doi: [10.1016/j.rse.2019.02.028](https://doi.org/10.1016/j.rse.2019.02.028).
- [19] M. D. R. González-Moradas and W. Viveen, "Evaluation of ASTER GDEM2, SRTMv3.0, ALOS AW3D30 and TanDEM-X DEMs for the Peruvian Andes against highly accurate GNSS ground control points and geomorphological-hydrological metrics," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111509, doi: [10.1016/j.rse.2019.11.1509](https://doi.org/10.1016/j.rse.2019.11.1509).
- [20] K. Gdulová, J. Marešová, and V. Moudrý, "Accuracy assessment of the global TanDEM-X digital elevation model in a mountain environment," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111724, doi: [10.1016/j.rse.2020.11.1724](https://doi.org/10.1016/j.rse.2020.11.1724).
- [21] P. L. Guth and T. M. Geoffroy, "LiDAR point cloud and ICESat-2 evaluation of 1 second global digital elevation models: Copernicus wins," *Trans. GIS*, vol. 25, no. 5, pp. 2245–2261, Aug. 2021, doi: [10.1111/tgis.12825](https://doi.org/10.1111/tgis.12825).
- [22] E. S. Ebinne, O. I. Apeh, E. C. Moka, and E. J. Abah, "Comparative analysis of freely available digital elevation models for applications in multi-criteria environmental modeling over data limited regions," *Remote Sens. Applications: Soc. Environ.*, vol. 27, Aug. 2022, Art. no. 100795, doi: [10.1016/j.rsase.2022.100795](https://doi.org/10.1016/j.rsase.2022.100795).
- [23] M. del Rosario González-Moradas, W. Viveen, R. A. Vidal-Villalobos, and J. C. Villegas-Lanza, "A performance comparison of SRTM v. 3.0, AW3D30, ASTER GDEM3, Copernicus and TanDEM-X for tectono-geomorphic analysis in the south American Andes," *CATENA*, vol. 228, Jul. 2023, Art. no. 107160, doi: [10.1016/j.catena.2023.107160](https://doi.org/10.1016/j.catena.2023.107160).
- [24] I. V. Florinsky, T. N. Skrypitsyna, S. Trevisani, and S. V. Romaikin, "Statistical and visual quality assessment of nearly-global and continental digital elevation models of Trentino, Italy," *Remote Sens. Lett.*, vol. 10, no. 8, pp. 726–735, Apr. 2019, doi: [10.1080/2150704x.2019.1602790](https://doi.org/10.1080/2150704x.2019.1602790).
- [25] D. Yamazaki et al., "A high-accuracy map of global terrain elevations," *Geophys. Res. Lett.*, vol. 44, no. 11, pp. 5844–5853, Jun. 2017, doi: [10.1002/2017GL072874](https://doi.org/10.1002/2017GL072874).
- [26] L. Hawker, J. Neal, and P. Bates, "Accuracy assessment of the TanDEM-X 90 digital elevation model for selected floodplain sites," *Remote Sens. Environ.*, vol. 232, Oct. 2019, Art. no. 111319, doi: [10.1016/j.rse.2019.11.1319](https://doi.org/10.1016/j.rse.2019.11.1319).
- [27] P. L. Guth et al., "Digital elevation models: Terminology and definitions," *Remote Sens.*, vol. 13, no. 18, p. 3581, Sep. 2021, doi: [10.3390/rs13183581](https://doi.org/10.3390/rs13183581).
- [28] L. Hawker et al., "A 30 m global map of elevation with forests and buildings removed," *Environ. Res. Lett.*, vol. 17, no. 2, Feb. 2022, Art. no. 024016, doi: [10.1088/1748-9326/ac4d4f](https://doi.org/10.1088/1748-9326/ac4d4f).
- [29] P. L. Guth, P. Strobl, K. Gross, and S. Riazanoff, 2023, "DEMIX 10k Tile Data Set (1.0)," *Zenodo*, doi: [10.5281/zenodo.7504791](https://doi.org/10.5281/zenodo.7504791).
- [30] I. S. Evans, "An integrated system of terrain analysis and slope mapping," *Zeitschrift für Geomorphologie*, vol. Suppl.-Bd. 36, pp. 274–295, 1980.
- [31] C. H. Grohmann, M. J. Smith, and C. Riccomini, "Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1200–1213, Apr. 2011.
- [32] M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus global land cover layers—Collection 2," *Remote Sens.*, vol. 12, no. 6, p. 1044, Mar. 2020, doi: [10.3390/rs12061044](https://doi.org/10.3390/rs12061044).
- [33] J. Iwahashi and R. J. Pike, "Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature," *Geomorphology*, vol. 86, nos. 3–4, pp. 409–440, May 2007, doi: [10.1016/j.geomorph.2006.09.012](https://doi.org/10.1016/j.geomorph.2006.09.012).
- [34] M. Meybeck, P. Green, and C. Vörösmarty, "A new typology for mountains and other relief classes," *Mountain Res. Develop.*, vol. 21, no. 1, pp. 34–45, 2001, doi: [10.1659/0276-4741\(2001\)021\[0034:ANTFMA\]2.0.CO;2](https://doi.org/10.1659/0276-4741(2001)021[0034:ANTFMA]2.0.CO;2).
- [35] J. Jasiewicz and T. F. Stepinski, "Geomorphons—A pattern recognition approach to classification and mapping of landforms," *Geomorphology*, vol. 182, pp. 147–156, Jan. 2013, doi: [10.1016/j.geomorph.2012.11.005](https://doi.org/10.1016/j.geomorph.2012.11.005).
- [36] P. Guth and M. Kane, "Slope, aspect, and hillshade algorithms for non-square digital elevation models," *Trans. GIS*, vol. 25, no. 5, pp. 2309–2332, Oct. 2021, doi: [10.1111/tgis.12852](https://doi.org/10.1111/tgis.12852).
- [37] P. Guth, "Geomorphometry in MicroDEM," in *Geomorphometry: Concepts, Software, Applications* (Developments in Soil Science), vol. 33. Amsterdam, The Netherlands: Elsevier, 2009, ch. 15, pp. 351–366, doi: [10.1016/S0166-2481\(08\)00015-9](https://doi.org/10.1016/S0166-2481(08)00015-9).
- [38] P. L. Guth. (2023). *MICRODEM*. Accessed: Feb. 10, 2023. [Online]. Available: [https://github.com/prof-pguth/git\\_microdem](https://github.com/prof-pguth/git_microdem)
- [39] P. L. Guth, S. Riazanoff, A. Corseaux, C. H. Grohmann, S. Trevisani, and C. López-Vázquez, 2023, "DEMIX 1," Reference DEMs version 2.0," *Zenodo*, doi: [10.5281/zenodo.8086806](https://doi.org/10.5281/zenodo.8086806).
- [40] S. Trevisani and M. Cavalli, "Topography-based flow-directional roughness: Potential and challenges," *Earth Surf. Dyn.*, vol. 4, no. 2, pp. 343–358, Apr. 2016, doi: [10.5194/esurf-4-343-2016](https://doi.org/10.5194/esurf-4-343-2016).
- [41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipes," in *The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007. [Online]. Available: <http://numerical.recipes/>
- [42] P. L. Guth, 2023, "DEMIX GIS database (2.0)," *Zenodo*, doi: [10.5281/zenodo.7402618](https://doi.org/10.5281/zenodo.7402618).
- [43] S. Adelman, "The generalized randomized block design," *Amer. Statistician*, vol. 23, no. 4, p. 35, Oct. 1969, doi: [10.2307/2681737](https://doi.org/10.2307/2681737).
- [44] I. L. Amerise and A. Tarsitano, "Correction methods for ties in rank correlations," *J. Appl. Statist.*, vol. 42, no. 12, pp. 2584–2596, Dec. 2015, doi: [10.1080/02664763.2015.1043870](https://doi.org/10.1080/02664763.2015.1043870).
- [45] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937, doi: [10.2307/2279372](https://doi.org/10.2307/2279372).
- [46] M. Fey and K. A. Clarke, "Consistency of choice in nonparametric multiple comparisons," *J. Nonparametric Statist.*, vol. 24, no. 2, pp. 531–541, Jun. 2012, doi: [10.1080/10485252.2012.675436](https://doi.org/10.1080/10485252.2012.675436).
- [47] C. López-Vázquez, A. Tasistro, and E. Hochstzain, "Exact tables for the Friedman rank test: Case with ties," *Chilean J. Statist.*, vol. 12, no. 1, pp. 1–23, 2021. [Online]. Available: <http://soche.cl/chjs/volumes/12/ChJS-12-01-02.pdf>
- [48] C. López-Vázquez and E. Hochstzain. (2020). *Tables for the Friedman's Test With Ties*. Accessed: Jan. 20, 2020. [Online]. Available: <https://dspace.ort.edu.uy/handle/20.500.11968/4208>
- [49] D. G. Pereira, A. Afonso, and F. M. Medeiros, "Overview of Friedman's test and post-hoc analysis," *Commun. Statist. Simul. Comput.*, vol. 44, no. 10, pp. 2636–2653, Aug. 2014, doi: [10.1080/03610918.2014.931971](https://doi.org/10.1080/03610918.2014.931971).
- [50] O. J. Dunn, "Multiple comparisons among means," *J. Amer. Stat. Assoc.*, vol. 56, no. 293, pp. 52–64, Mar. 1961, doi: [10.2307/2282330](https://doi.org/10.2307/2282330).
- [51] T. Kluyver et al., "Jupyter notebooks—A publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Amsterdam, The Netherlands: IOS Press, 2016, pp. 87–90, doi: [10.3233/978-1-61499-649-1-87](https://doi.org/10.3233/978-1-61499-649-1-87).
- [52] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [53] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [54] W. McKinney, "Pandas: A foundational Python library for data analysis and statistics," *Python High Perform. Sci. Comput.*, vol. 14, no. 9, pp. 1–9, 2011.
- [55] Python Software Foundation. (2021). *Python Programming Language, Version 3.10.x*. Accessed: Jan. 20, 2020. [Online]. Available: <http://www.python.org/>
- [56] C. H. Grohmann, "The DEMIX Wine Contest Jupyter notebook," in *Proc. Geomorphometry2023 Conf.*, Iasi, Romania, 2023. <https://dx.doi.org/10.5281/zenodo.7779255>

- [57] B. Purinton and B. Bookhagen, "Beyond vertical point accuracy: Assessing inter-pixel consistency in 30 m global DEMs for the arid central Andes," *Frontiers Earth Sci.*, vol. 9, Oct. 2021, Art. no. 758606, doi: [10.3389/feart.2021.758606](https://doi.org/10.3389/feart.2021.758606).
- [58] D. B. Gesch, M. J. Oimoen, and G. A. Evans, "Accuracy assessment of the U.S. geological survey national elevation dataset, and comparison with other large-area elevation datasets: SRTM and ASTER," U.S. Geological Surv., Reston, VA, USA, Open-File Tech. Rep. 2014-1008, 2014, doi: [10.3133/ofr20141008](https://doi.org/10.3133/ofr20141008).
- [59] E. Uuemaa, S. Ahi, B. Montibeller, M. Muru, and A. Kmoch, "Vertical accuracy of freely available global digital elevation models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM)," *Remote Sens.*, vol. 12, no. 21, p. 3482, Oct. 2020, doi: [10.3390/rs12213482](https://doi.org/10.3390/rs12213482).
- [60] D. B. Gesch, "Best practices for elevation-based assessments of sea-level rise and coastal flooding exposure," *Frontiers Earth Sci.*, vol. 6, p. 230, Dec. 2018, doi: [10.3389/feart.2018.00230](https://doi.org/10.3389/feart.2018.00230).
- [61] S. Trevisani, T. N. Skrypitsyna, and I. V. Florinsky, "Global digital elevation models for terrain morphology analysis in mountain environments: Insights on Copernicus GLO-30 and ALOS AW3D30 for a large Alpine area," *Environ. Earth Sci.*, vol. 82, p. 198, Apr. 2023, doi: [10.1007/s12665-023-10882-7](https://doi.org/10.1007/s12665-023-10882-7).
- [62] B. Dobre, I. P. Kovács, and T. Bugya, "Comparison of digital elevation models through the analysis of geomorphic surface remnants in the desatoya mountains, Nevada," *Trans. GIS*, vol. 25, no. 5, pp. 2262–2282, Aug. 2021, doi: [10.1111/tgis.12819](https://doi.org/10.1111/tgis.12819).
- [63] J. P. Wilson and J. C. Gallant, *Terrain Analysis: Principles and Applications*. Hoboken, NJ, USA: Wiley, 2000.
- [64] T. Hengl and H. I. Reuter, *Geomorphometry: Concepts, Software, Applications* (Developments in Soil Science), vol. 33. Amsterdam, The Netherlands: Elsevier, 2008.
- [65] I. V. Florinsky, *Digital Terrain Analysis in Soil Science and Geology*. Amsterdam, The Netherlands: Elsevier, 2016.
- [66] G. Titti, L. Borgatti, Q. Zou, P. Cui, and A. Pasuto, "Landslide susceptibility in the belt and road countries: Continental step of a multi-scale approach," *Environ. Earth Sci.*, vol. 80, no. 18, Sep. 2021, doi: [10.1007/s12665-021-09910-1](https://doi.org/10.1007/s12665-021-09910-1).
- [67] F. Huang et al., "Landslide susceptibility prediction considering neighborhood characteristics of landslide spatial datasets and hydrological slope units using remote sensing and GIS technologies," *Remote Sens.*, vol. 14, no. 18, p. 4436, Sep. 2022, doi: [10.3390/rs14184436](https://doi.org/10.3390/rs14184436).
- [68] D. J. Wald and T. I. Allen, "Topographic slope as a proxy for seismic site conditions and amplification," *Bull. Seismological Soc. Amer.*, vol. 97, no. 5, pp. 1379–1395, Oct. 2007, doi: [10.1785/0120060267](https://doi.org/10.1785/0120060267).
- [69] D. C. Heath, D. J. Wald, C. B. Worden, E. M. Thompson, and G. M. Smoczyk, "A global hybrid VS30 map with a topographic slope-based default and regional map insets," *Earthq. Spectra*, vol. 36, no. 3, pp. 1570–1584, Jun. 2020, doi: [10.1177/8755293020911137](https://doi.org/10.1177/8755293020911137).
- [70] S. J. Boulton and M. Stokes, "Which DEM is best for analyzing fluvial landscape development in mountainous terrains?" *Geomorphology*, vol. 310, pp. 168–187, Jun. 2018, doi: [10.1016/j.geomorph.2018.03.002](https://doi.org/10.1016/j.geomorph.2018.03.002).
- [71] A. Chrobak, J. Novotný, and P. Struý, "Geodiversity assessment as a first step in designating areas of geotourism potential. Case study: Western carpathians," *Frontiers Earth Sci.*, vol. 9, Nov. 2021, Art. no. 752669, doi: [10.3389/feart.2021.752669](https://doi.org/10.3389/feart.2021.752669).
- [72] G. J.-P. Schumann and P. D. Bates, "The need for a high-accuracy, open-access global DEM," *Frontiers Earth Sci.*, vol. 6, p. 225, Dec. 2018, doi: [10.3389/feart.2018.00225](https://doi.org/10.3389/feart.2018.00225).
- [73] S. A. Kulp and B. H. Strauss, "CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network," *Remote Sens. Environ.*, vol. 206, pp. 231–239, Mar. 2018, doi: [10.1016/j.rse.2017.12.026](https://doi.org/10.1016/j.rse.2017.12.026).
- [74] S. A. Kulp and B. H. Strauss, "New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding," *Nature Commun.*, vol. 10, 2019, Art. no. 4844, doi: [10.1038/s41467-019-12808-z](https://doi.org/10.1038/s41467-019-12808-z).
- [75] P. L. Guth, C. H. Grohmann, and S. Trevisani, "Subjective criterion for the DEMIX wine contest: Hillshade maps," in *Proc. Geomorphometry Conf.*, Iasi, Romania, 2023, pp. 115–119, doi: [10.5281/zenodo.8030735](https://doi.org/10.5281/zenodo.8030735).
- [76] O. Conrad et al., "System for automated geoscientific analyses (SAGA) v. 2.1.4," *Geoscientific Model Develop.*, vol. 8, no. 7, pp. 1991–2007, Jul. 2015, doi: [10.5194/gmd-8-1991-2015](https://doi.org/10.5194/gmd-8-1991-2015).
- [77] SAGA. (2024). *System for Automated Geoscientific Analyses, Version 9.3.0*. Accessed: Jan. 12, 2024. [Online]. Available: <https://saga-gis.sourceforge.io/en/index.html>
- [78] E. Stephens, G. Schumann, and P. Bates, "Problems with binary pattern measures for flood model evaluation," *Hydrological Processes*, vol. 28, no. 18, pp. 4928–4937, Aug. 2014, doi: [10.1002/hyp.9979](https://doi.org/10.1002/hyp.9979).
- [79] H. Zhang et al., "An integrated algorithm to evaluate flow direction and flow accumulation in flat regions of hydrologically corrected DEMs," *CATENA*, vol. 151, pp. 174–181, Apr. 2017, doi: [10.1016/j.catena.2016.12.009](https://doi.org/10.1016/j.catena.2016.12.009).
- [80] B. Smith and D. Sandwell, "Accuracy and resolution of shuttle radar topography mission data," *Geophys. Res. Lett.*, vol. 30, no. 9, p. 1467, May 2003, doi: [10.1029/2002gl016643](https://doi.org/10.1029/2002gl016643).
- [81] L. Pierce, J. Kelldorf, W. Walker, and O. Barros, "Evaluation of the horizontal resolution of SRTM elevation data," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 11, pp. 1235–1244, Nov. 2006, doi: [10.14358/pers.72.11.1235](https://doi.org/10.14358/pers.72.11.1235).



**Conrad Bielski** (Member, IEEE) received the Ph.D. degree in geography from the Université de Montréal, Quebec, QC, Canada, in 2002.

His post-doctoral research included working on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) at the Jet Propulsion Laboratory (JPL), Pasadena, CA, USA, and automatic information extraction algorithms at the Joint Research Centre (JRC), Ispra, Italy. In 2012, he founded EOXPLORE, Oberhaching, Germany, which provides research and development expertise in the field of

Earth observation and image processing for policy support and downstream services. His research focuses include natural capital accounting and environmental and social risk/responsibility.

Dr. Bielski is the Co-Chair of the CEOS Digital Elevation Model Inter-comparison eXercise (DEMIX) subgroup 2.



**Carlos López-Vázquez** (Senior Member, IEEE) received the Ph.D. degree in geomatics from the Royal Institute of Technology, Stockholm, Sweden, in 1997.

He is currently a Researcher with Universidad ORT Uruguay, Montevideo, Uruguay. His research interests include GIScience, spatial data infrastructures, and statistics and numerical analysis.



**Carlos H. Grohmann** (Member, IEEE) received the D.Sc. degree in geotectonics from the Universidade de São Paulo (USP), São Paulo, Brazil, in 2008.

He is currently an Associate Professor with the Institute of Energy and Environment (IEE), USP. His research interests include geomorphometry, 3-D modeling of topography with light detection and ranging (LiDAR) and structure from motion-multi view stereo (SM-MVS) photogrammetry, applications of machine learning to the study of mass movements, and environmental applications of remote

sensing.

Dr. Grohmann is a Board Member of the International Society for Geomorphometry. He was a recipient of the prize for the 2009 Best Brazilian Doctoral Thesis in Geosciences awarded by the Brazilian Agency for Higher Education Personnel Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). He is the Co-Chair of the CEOS Digital Elevation Model Inter-comparison eXercise (DEMIX) subgroup 2. He is the Editor-in-Chief of the *Brazilian Journal of Geology*.



**Peter L. Guth** received the Ph.D. degree in geology with a focus on structural mapping in the sheep range of Southern Nevada from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1980.

He retired after over 50 years with the U.S. Federal Government, including over 38 years teaching at service academies. He has been involved with manipulating digital elevation models (DEMs) and geomorphometry using Turbo Pascal and Delphi since 1984. His MICRODEM program figured heavily

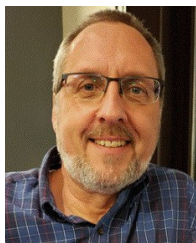
in the work for this article.

Dr. Guth is a fellow of the Geological Society of America. He is the Former President of the International Society for Geomorphometry and was a recipient of its Lifetime Achievement Award.



**Laurence Hawker** received the Ph.D. degree in geography from the University of Bristol, Bristol, U.K., in 2019.

He is currently a Senior Research Associate with the School of Geographical Sciences, University of Bristol, where he primarily works on the Natural Environment Research Council-funded Evoflood project (NE/S015590/1). His research interests are digital elevation models, hydrodynamic models, flood risk, remote sensing, and population projections.



**Dean Gesch** received the Ph.D. degree in geospatial science and engineering from South Dakota State University, Brookings, SD, USA, in November 2006.

He is currently a Research Physical Scientist with the U.S. Geological Survey, Sioux Falls, SD, USA. Since 1992, he has been conducting topographic science research at the Earth Resources Observation and Science Center (EROS), Sioux Falls, where he is the Principal Investigator leading the terrain monitoring and modeling focus area in the Integrated

Science and Applications Branch. His research topics include sea-level rise vulnerability assessment, topographic change analysis and monitoring, hurricane storm surge mapping, elevation data accuracy assessment, and topographic data development and applications.



**Sebastiano Trevisani** received the Ph.D. degree in earth sciences from the University of Padova, Padova, Italy, in March 2005.

From November 2010 to June 2018, he was a Researcher in applied geology with the Department of Architecture Construction and Conservation (DACC), Iuav University of Venice, Venice, Italy. He has also worked for various research institutes, including the Research Institute for the Hydrogeological Protection (IRPI), Padua, Italy, from 2007 to 2010; the National Council of Research

(CNR), Rome, Italy; and the National Institutes of Applied Oceanography and Geophysics of Trieste (OGS), Sgonico, Italy, from 2006 to 2007. Since July 2018, he has been an Associate Professor of applied geology with the Iuav University of Venice, where his teaching activity is related to the courses of "Environmental Geology" and "Applied Geology." His main research activities are related to geocomputational approaches for the analysis of geoenvironmental systems, with a special focus on geosphere–anthroposphere–interlinked dynamics (such as hydrogeology, natural hazards, geoengineering issues in urbanized contexts, geomorphometry, and sustainability). His expertise is in geostatistics and geomorphometry applied to geoenvironmental analysis.



**Virginia Herrera-Cruz** (Member, IEEE) completed her studies in forest technical engineering at the University of Huelva, Huelva, Spain, in 2004.

She explored the methodological improvement of automated processes for land cover classification in her M.Sc. thesis at the Department of Remote Sensing and Landscape Information Systems, University of Freiburg, Freiburg im Breisgau, Germany, in 2007. In 2007, she joined Airbus Defence and Space, Friedrichshafen, Germany, focusing her work on the integration of automated and semiautomated

image analysis processes within the mapping production chains. Since the launch of TerraSAR-X in 2007 and the TanDEM-X Mission in 2010, the development of synthetic-aperture radar (SAR)-related products and applications have become the main topic of her work. She is currently the Product Development Manager and the Product Owner of the global WorldDEM Neo Digital Terrain Model (DTM).



**Serge Riazanoff** received the Ph.D. degree in physical methods in remote sensing from the University of Paris 7, Paris, France, in 1989. His thesis focused on the automatic extraction of the hydrographic network by simulating rain on digital elevation models (DEMs).

He was an Associate Professor with Université Paris-Est Créteil, Créteil, France, as well as several other universities, institutes, and schools, particularly in the African continent. He is currently the Director of VisioTerra, Champs-sur-Marne, France, a company specializing in scientific consulting in Earth observation. He supervises software development projects for platforms facilitating end-user access to free Earth observations and models, providing geoservices, and producing cartography and studies for the European Space Agency, Rome, Italy.



**Axel Corseaux** received the M.S. degree in imaging science from Gustave Eiffel University, Champs-sur-Marne, France, in 2021.

He is currently an Earth Observation R&D Engineer with VisioTerra, Champs-sur-Marne. He is involved in building tools and carrying out studies for the visualization, analysis, and quality assessment of digital elevation models (DEMs). His work is carried out in the framework of Earthnet Data Assessment Pilot (EDAP), funded by the European Space Agency, Rome, Italy.

Mr. Corseaux has been an Active Member of DEM Inter-comparison eXercise (DEMIX) since the beginning of the project.



**Hannes I. Reuter** received the Ph.D. degree in soil science from the University of Hannover, Hannover, Germany, in 2004.

He has specialized in spatial research projects for natural resource management from global (SRTM Void filling, Crop Growth Modelling, and Landform Classification) and continental (Digital Soil Mapping, Wind Erosion Modelling, and Air Quality Monitoring) to local scales (Precision Agriculture and Water Erosion Modelling). One of his focal points is transforming valuable tacit knowledge into

statistical, uncertainty-related information, which can be used to run multiscale decision-making models for sound policy decision-making. In his current work, he tries to bridge technical feasibilities and functional requirements serving a wide community of European Commission colleagues. He is currently the President of the International Society of Geomorphometry.



**Peter Strobl** received the M.Sc. degree in geophysics from the University of Munich, Munich, Germany, in 1991, and the Ph.D. degree in remote sensing from the University of Potsdam, Potsdam, Germany, in 2000.

He is currently a Principal Scientist with the Joint Research Centre of the European Commission, Ispra, Italy. His work mainly concerns data architecture and interoperability in the frame of the Copernicus Earth Observation program, including the definition of standards, references, and quality metrics for

geospatial data.

Dr. Strobl is a member of various international working groups and advisory panels in the field of Earth observation and geospatial data and currently chairing the Terrain Mapping Sub-Group within the CEOS Working Group on Calibration and Validation.