



OPEN

DATA DESCRIPTOR

# Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers

Giuseppe Amatulli<sup>1,2</sup>✉, Daniel McInerney<sup>3</sup>, Tushar Sethi<sup>4</sup>, Peter Strobl<sup>5</sup> & Sami Domisch<sup>6</sup>✉

Topographical relief comprises the vertical and horizontal variations of the Earth's terrain and drives processes in geomorphology, biogeography, climatology, hydrology and ecology. Its characterisation and assessment, through geomorphometry and feature extraction, is fundamental to numerous environmental modelling and simulation analyses. We, therefore, developed the Geomorpho90m global dataset comprising of different geomorphometric features derived from the MERIT-Digital Elevation Model (DEM) - the best global, high-resolution DEM available. The fully-standardised 26 geomorphometric variables consist of layers that describe the (i) rate of change across the elevation gradient, using first and second derivatives, (ii) ruggedness, and (iii) geomorphological forms. The Geomorpho90m variables are available at 3 (~90 m) and 7.5 arc-second (~250 m) resolutions under the WGS84 geodetic datum, and 100 m spatial resolution under the Equi7 projection. They are useful for modelling applications in fields such as geomorphology, geology, hydrology, ecology and biogeography.

## Background & Summary

Geomorphometry is the science of quantitative analysis of the Earth's surface<sup>1</sup> and is amply used to address several multiscale geoscientific problems<sup>2</sup>. The primary inputs for such terrain analyses are remotely sensed Digital Elevation Models (DEMs), which provide an opportunity to derive a wide range of environmental variables and facilitates a better understanding of the patterns and processes in geomorphology, geology, climatology, hydrology or biodiversity science.

DEMs are usually categorised as being either a Digital Surface Model (DSM) or a Digital Terrain Model (DTM). The DSM is a 3D elevation model of the Earth's surface that includes objects, such as trees and buildings, whereas the DTM represents the elevation of the bare earth without any temporary objects. DEMs can also be used to characterise topographical complexity<sup>3</sup>, as they provide the elevation above sea level, and allow for a wide array of geomorphometric metrics to be generated (also known as topographic, or geomorphometric variables). These parameters are quantitative measures of surface properties, which improve our understanding of the geographical, geomorphological and environmental properties of a given area of study<sup>4</sup>.

Topographical variations also influence numerous environmental dynamics, contributing significantly to the environmental complexity of a region. For example, they define the biotic and abiotic features at a sub-regional level<sup>5</sup>, and can shape the macro and micro climate of a given area<sup>3</sup>. The most common of these morphometric parameters, slope and aspect, can be used to further derive more complex features or curvature profiles of a terrain at any given location<sup>6</sup>. Florinsky<sup>2,6</sup>, in his monographs, makes accurate analytical and mathematical descriptions of 29 geomorphometric variables. Besides these, Sofia<sup>7</sup> has written an extensive literature review

<sup>1</sup>Yale University, School of Forestry & Environmental Studies, New Haven, CT, 06511, USA. <sup>2</sup>Yale University, Centre for Research Computing, New Haven, CT, 06511, USA. <sup>3</sup>Coillte Teoranta, Irish State Forestry Board, Plassey Road, Castletroy, Limerick, V94-C780, Ireland. <sup>4</sup>Spatial Ecology, 35A, Hazlemere Road, Penn, Buckinghamshire, HP10 8AD, United Kingdom. <sup>5</sup>European Commission, Joint Research Centre, Directorate for Sustainable Resources, Ispra, VA-21027, Italy. <sup>6</sup>Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department of Ecosystem Research, Müggelseedamm 310, 12587, Berlin, Germany. ✉e-mail: [giuseppe.amatulli@gmail.com](mailto:giuseppe.amatulli@gmail.com); [domisch@igb-berlin.de](mailto:domisch@igb-berlin.de)

of geomorphometry to advance the theoretical and practical understanding of geomorphological phenomena. Amongst all of the described geomorphometric variables, we selected 26 of the most commonly used ones.

Such topographic measures are also central to hydrological parameters shaping flow and erosion processes within the landscape, and to delineate catchment and stream features<sup>8</sup>. Additionally, the mapping and assessment of landform variability such as concavity and convexity is essential to obtain a better understanding of land erosion and landscape denudation dynamics in mountainous environments<sup>9</sup>. The morphometric properties of a surface are also important for predicting other phenomena such as wildfires<sup>10</sup>, mountain/alpine snow cover<sup>11</sup> and landslide formation<sup>12</sup>. Understanding surface morphometric properties is also crucial beyond such geomorphological applications, since terrain features play a critical role in understanding contemporary biodiversity patterns given species occurrences (based on their habitat preferences) and potential species migration corridors where such detailed terrain information over large spatial extents is crucial.

Even with the latest DEMs, geomorphometric information can only be obtained on a case-by-case, location-specific basis. Therefore, standardisation is necessary to enable spatially comparative analyses between regions and continents. To this end, Amatulli *et al.*<sup>5</sup> established a suite of 15 geomorphometric variables based on 7.5 arc-second (~250 m) Global Multi-resolution Terrain Elevation (GMTED)<sup>13</sup> data. Building upon this work, we present in this paper the Geomorpho90m dataset<sup>14–16</sup>. Here, we extend the concept in Amatulli *et al.*<sup>5</sup>, by calculating a suite of 26 DEM-derived geomorphometric variables based on the Multi-Error-Removed Improved Terrain (MERIT) DEM at a spatial resolution of 3 arc-second (~90 m)<sup>17,18</sup>, which to-date is considered the best-effort in global DEMs<sup>19,20</sup>.

This newly-developed Geomorpho90m dataset<sup>14–16</sup> of 26 variables hence provides the foundation for globally seamless, high-resolution studies. It consists of the following raster layers with standardised spatial extent and pixel resolution: slope, aspect, aspect sine, aspect cosine, eastness, northness, convergence, compound topographic index (also known as topographic wetness index), stream power index, first directional derivatives, profile and tangential curvature, second directional derivatives, elevation standard deviation, terrain roughness index, roughness, vector ruggedness measure, topographic position index, multiscale deviation and roughness, and geomorphologic forms. Through the use of the latest geocomputational methods and underlying DEMs, Geomorpho90m offers a marked improvement over previous datasets of this nature.

The quality and appeal of Geomorpho90m lies in the rigorous scripting procedures and tiling system that allowed for multi-core processing in a super computer. Additionally, the computation was performed using the Equi7 projection<sup>21</sup>, which minimises the pixel-level distortions that otherwise often occur when unprojected data (i.e. latitude-longitude in the World Geodetic System (WGS)) are treated as if they were a square raster under a cartesian coordinate system. The scripting procedures allow both multi-core processing and Equi7 reprojection, and are thus complex tasks that require advanced geocomputation programming skills. Therefore, for the sake of end-user expediency, we have undertaken these steps at the outset to offer an all-inclusive data product. Finally, to assess the quality of our work, we have compared the newly created variables with those generated from DEMs derived from 3D Elevation Program (3DEP) and Light Detection and Ranging (LiDAR).

## Methods

The Methods section is divided into three subsections that includes: (i) Source Data, that describes the MERIT-DEM used to produce the global geomorphometric layers, as well as other ancillary DEMs used to compare the main product; (ii) Projection and Tiling system, which was chosen to minimise computational errors from surface distortions; (iii) Derived Geomorphometric Variables, which describes each geomorphometric layer and its principal use in environmental modelling. Additional information on these variables and related procedures (software and scripting routines) can be found in Amatulli *et al.*<sup>5</sup>.

**Source data.** This section describes the DEM used to compute the global geomorphometric variables and two additional DEMs used for comparison. The principal dataset employed in this study was the Multi-Error-Removed Improved Terrain (MERIT) DEM 3 arc-seconds (~90 m)<sup>17</sup>, which was used to extract geomorphometric terrain features. In addition, ancillary DEMs were used for the purposes of comparison, which included DEMs derived from Light Detection and Ranging (LiDAR) and data from the 3D Elevation Program (3DEP). The resulting geomorphometric variables consisted of a suite of 26 layers calculated from the source layers, which had been previously reprojected under the Equi7 projection.

*Multi-Error-Removed Improved Terrain (MERIT) - DEM.* Due to the lack of global high-resolution DEM obtained from a single data-sensor source, there have been several attempts to combine DEMs generated from multiple sensors<sup>13</sup>. One of the principal DEM sources is the Shuttle Radar Topography Mission (SRTM), which was acquired in February 2000 and provides a near-global coverage DEM of the Earth's surface from 56°S to 60°N during an 11-day period. SRTM used a C-band radar system on board the space shuttle and relied on interferometry to generate the DEMs.

One of the characteristics of SRTM is that the short wavelengths associated with C-band radar is unable to sufficiently penetrate vegetated areas, in particular in forested areas<sup>22,23</sup>. Indeed, SRTM has been used as a DSM to estimate the height of vegetation canopies<sup>24</sup>. However, this effectively inhibits the creation of a correct DTM. The attenuation rate over forests is also affected by both the forest density and moisture content within the forest (e.g. more penetration is expected in drier forest types). In addition, as with any SAR instruments, the resulting dataset contains both speckle noise and gaps on steep slopes due to degraded accuracy caused by radar foreshortening and layover.

In recent years, many research initiatives have sought to improve the quality of DEM datasets derived from spaceborne products such as the SRTM and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM<sup>25</sup>. These improvements have largely focused on the removal of height errors and artefacts

contained within them. One such dataset is the Multi-Error-Removed Improved Terrain (MERIT) - DEM 3 arc-seconds ( $\sim 90$  m)<sup>17</sup>, which sought to correct SRTM for absolute elevation biases using high-precision LiDAR elevation from ICESat and ancillary datasets as a reference.

The MERIT-DEM is based on the National Aeronautics and Space Administration (NASA) SRTM3 version 2.1, the Japan Aerospace Exploration Agency (JAXA) AW3D global high resolution 3D map (version 1) and the Viewfinder Panorama's DEM. It removed inherent features or errors found in these products that included: stripe noise, absolute bias, tree height bias and speckle noise. Stripe noise was removed from the detection of unrealistic regular terrain undulations using a 2-D Fourier filtering technique. Absolute bias was corrected by calculating the difference between the DEM and the ICESat elevations<sup>26</sup>. Tree-height bias was estimated from the combination of tree density<sup>27</sup> and tree height<sup>28</sup> using ancillary layers, and by comparing the obtained MERIT-DEM to ICESat, which can deliver DSM as well as DTM estimates. The speckle noise was removed using an adaptive-scale smoothing filter<sup>29</sup>. Yamazaki *et al.*<sup>17</sup> noted that after the error removal, areas mapped with  $\pm 2$  m or better vertical accuracy increased by 19% and that previous DEMs contained slope distortions in many of the major floodplains.

The resulting MERIT-DEM product covers the land area between 90°N and 60°S at a spatial resolution of approximately  $\sim 90$  m at the equator in the Plate Carrée projection on a WGS84 Ellipsoid, and is considered to be the optimal best-effort DEM that is currently available as free and open data on a global scale<sup>19,20</sup>. One of the unique aspects of our study is the reprojection of MERIT-DEM to Equi7 (based on a bilinear interpolation method and the projection parameters previously outlined), and the associated computation of geomorphometric variables with minimal distortions. This is a complex procedure, which has been completed and made available for the benefit of future research.

**3D elevation program - 3DEP.** The U.S. Geological Survey's National Map provides 3D Elevation Program (3DEP) products and services of standardised DEMs for the US, which were previously referred to as the National Elevation Dataset (NED)<sup>30</sup>. The 3DEP products are delivered using consistent datum, elevation units, coordinate reference system, and are distributed at different spatial resolutions (1/3, 1, and 2 arc-seconds), and mosaicked and edge-matched.

For the purposes of this analysis, the Seamless 1 arc-second ( $\sim 30$  m ground sampling distance) 3DEP DEM product was used (hereinafter 3DEP-1). It provides complete coverage of the US land area and partial coverage of Alaska. The source data products for 3DEP-1 include:

- LiDAR point cloud data collected in 2014, which meets 3DEP specifications for horizontal accuracy and pulse spacing, as well as the resulting bare earth DTMs.
- IfSAR Digital Surface Model (DSM), which is a 5-metre raster only available over Alaska.
- IfSAR ortho-rectified radar intensity image, which is radar reflectance imagery only available over Alaska.

Due to the nature of 3DEP, which is a LiDAR derived product for the conterminous United States, the elevation values can be considered the best available DTM at 1 arc-second. Prior to the computation of the geomorphometric variables, 3DEP-1 was reprojected to Equi7 using the projection parameters previously described. During the reprojection we set the pixel resolutions to 100 m to match and be comparable to the resolution of the Equi7 MERIT-DEM. However, in the manuscript we kept the nomenclature as 3DEP-1 and MERIT-DEM, which refer to the original DEM data sources.

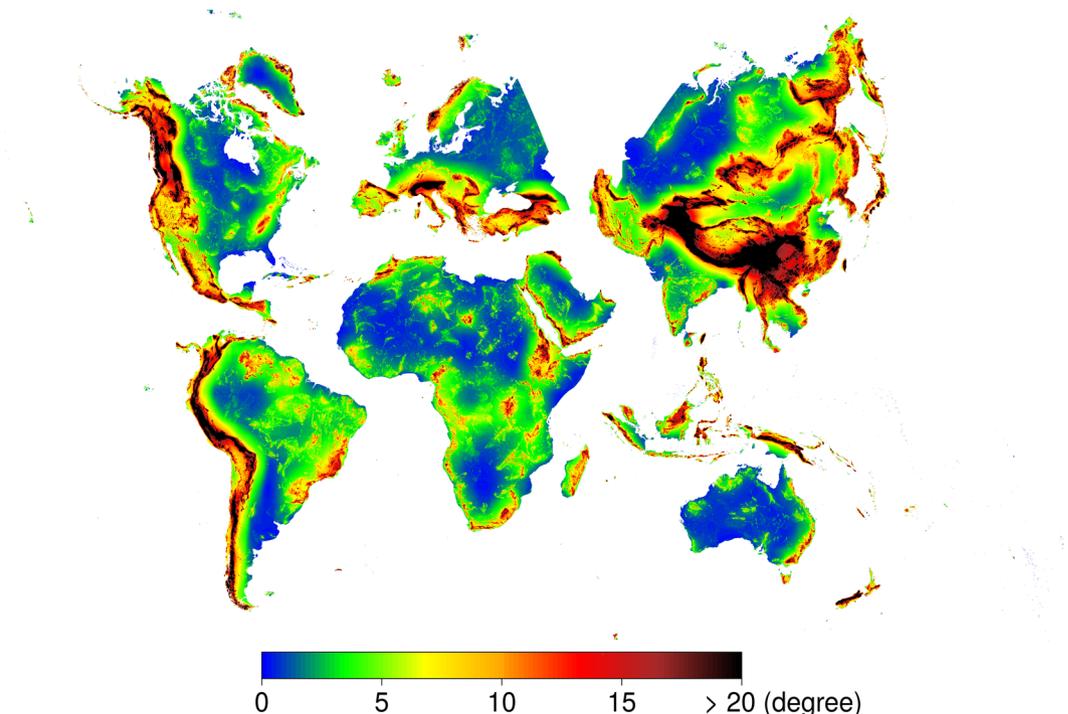
**Light detection and ranging (LiDAR).** Light Detection and Ranging (LiDAR) is an active remote sensing method that measures distances using a laser. The LiDAR sensor emits thousands of light pulses every second, and the scanner records the portion of the reflected light from the target. Given that the LiDAR sensors consist of a GPS and a precise timing unit, it is possible to calculate the exact 3-dimensional location of each point. LiDAR datasets are most commonly delivered in the LiDAR Aerial Survey format (LAS, or in its compressed format LAZ), and consist of 3-dimensional point clouds in a typical XYZ structure.

One of the main applications of LiDAR data is to generate gridded raster products that represent Digital Terrain Models (DTM) and DSM<sup>31</sup>. The digital model datasets are generated using algorithms that convert irregular point cloud datasets to gridded rasters<sup>32</sup>. In our study, we computed the DTM and DSM from LiDAR data, primarily to assess the quality of the tree-height correction applied to MERIT.

In order to evaluate the effect of tree height correction, we compared MERIT-DEM to DTMs and DSMs derived from two LiDAR datasets downloaded from the OpenTopography website<sup>33,34</sup>. Extensive forest cover was evident through a visual inspection in Google Earth, in conjunction with the forest cover map created by Hansen *et al.*<sup>27</sup>. The presence of forest canopy allowed for the creation of both DTM and DSM, which differed substantially in height. This in turn allowed for the comparison of the LiDAR-derived DTM with MERIT, as the latter excludes tree height.

**Projection and tiling system.** The World Geodetic System (WGS) is a global reference system for geospatial information used in geography and in several satellite derived products. Often used for global representation of the Earth, it consists of a spherical coordinate system expressed in degrees and a standard ellipsoidal reference surface fixed to a datum. The datum for the WGS was established in 1984 and last revised in 2004, and is referred to as WGS84 or EPSG code 4326<sup>35</sup>.

Any kind of quasi-spherical surface represented in 2D (map format) over large extents inevitably suffers from three types of potential distortion: (i) length distortions; (ii) angular distortions; and (iii) areal distortions. All of these distortions are caused when the spherical coordinates of WGS84 are regarded as square grids (i.e. Plate Carrée or Mercator projections) and increase with a latitudinal gradient, with very high areal/length distortion values in the subarctic and Antarctic zones. Therefore, any environmental values affected by the distortion,



**Fig. 1** Maximum multiscale roughness. Quasi-global representation of the MERIT-derived maximum multiscale roughness (*rough-magnitude*), computed under Equi7 with a pixel size of 100 m using the computation parameters reported in the Multiscale roughness description. For the purpose of plotting the graphs, the 100 m resolution was aggregated to 1 km ( $10 \times 10$  pixels) by calculating the mean. Projection distortion of each Equi7 zone does not allow the merging of zones for a global continuous surface. However, the zones have been positioned adjacent to each other as a mosaic of six projections with gaps in between, to present a quasi-global map. Global visualisation in WGS84 for the other geomorphometric variables can be seen on the Geomorpho90m webpage<sup>60</sup>.

degrade geographic analyses, especially in the northern hemisphere<sup>36</sup>. These distortions also result in local data oversampling when one projects generic satellite images to a regular raster grid<sup>36</sup>. The Grid Oversampling Factor (GOF) metric was devised to capture the local data oversampling effect, and indirectly quantify the overall distortion caused due to reprojection<sup>36</sup>. The GOF was then used as a baseline, against which multiple projections were compared to determine the effect of the distortions<sup>36</sup>.

To minimise the overall areal and shape distortions in WGS84, the Equi7 Equidistant Azimuthal set of projections was proposed<sup>36</sup> and is used in global mapping studies<sup>37</sup>. Bauer-Marschallinger *et al.*<sup>36</sup> reported figures and tables for angular distortion and GOF. Generally, the Equi7 grids minimise the GOF to a global overall of  $\sim 1$ , compared to global and hemispherical grids which have a minimum GOF of 1.3. The Equi7 system was developed based on computation and end-user requirements, and consequently, the following objectives were carefully considered: firstly, landmasses should form compact, contiguous areas; secondly, the oceans should be used as borders between the sub-grids and finally, countries should not be split. As a result, the Equi7 Grid system covers the entire global with no gaps, and has a 50 km overlap between land borders. It is optimised for the storage and processing of global high resolution Earth Observation image data and consists of 7 projected continental sub-grids based on the Equidistant Azimuthal projection, which includes: Europe, Asia, North America, South America, Africa, Oceania and Antarctica (see Fig. 1 for a graphical representation of six sub-grids - excluding Antarctica). Additional information such as projection centres and GOF for each continental sub-grid can be found in Table 3 of Bauer-Marschallinger *et al.*<sup>36</sup>. Each sub-grid has the projection centre close to the continent's barycentre and is divided into three nested tiling systems, which are T6 (600 km); T3 (300 km) and T1 (100 km)<sup>36</sup>. The predefined tiling systems are useful for the implementation of a multi-process workflow that does not require the information of the full continental zone. All Equi7 projection parameters, such as the seven continental central points, zone-continent borders, T6 tiles, etc., are stored in the Equi7 GitHub repository<sup>38</sup>.

A potential further reduction of projection distortions (linear, areal or angular), could be achieved by reprojecting the data into each of the 60 UTM grid zones and calculating the geomorphometric variables. Nonetheless, distortion increases in each UTM zone as the boundaries between the UTM zones are approached. Therefore, the enlargement of UTM tiles and consequently the weighted average as a function of UTM tile centre distance is required in order to avoid border effects. To achieve this solution requires complex programming and computation, especially considering the twice-computed pixels – once for each overlapped UTM tile, and the subsequently calculated weighted average. Such a complex exercise is only feasible when employing a high-resolution, e.g. a 30 m error corrected DTM, where it is important to preserve optimal accuracy.

Geomorphometric variable group	Geomorphometric variable name	Geomorphometric variable abbreviation	Software used
First order derivative	Slope	<i>slope</i>	GDAL: gdaldem <sup>66</sup>
	Aspect	<i>aspect</i>	GDAL: gdaldem <sup>66</sup>
	Aspect cosine	<i>aspect-cosine</i>	GDAL: gdaldem gdal_calc.py <sup>66</sup>
	Aspect sine	<i>aspect-sine</i>	GDAL: gdaldem gdal_calc.py <sup>66</sup>
	Eastness	<i>eastness</i>	GDAL: gdaldem gdal_calc.py <sup>66</sup>
	Northness	<i>northness</i>	GDAL: gdaldem gdal_calc.py <sup>66</sup>
	Convergence	<i>convergence</i>	GRASS GIS: r.convergence <sup>67</sup>
	Compound topographic index	<i>cti</i>	GRASS GIS: r.watershed <sup>67</sup>
	Stream power index	<i>spi</i>	GRASS GIS: r.watershed <sup>67</sup>
	East-West first order partial derivative	<i>dx</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
	North-South first order partial derivative	<i>dy</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
Second order derivative	Profile curvature	<i>pcurv</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
	Tangential curvature	<i>tcurv</i>	GRASS GIS: r.slope.aspect
	East-West second order partial derivative	<i>dxx</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
	North-South second order partial derivative	<i>dyy</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
	Second order partial derivative	<i>dxy</i>	GRASS GIS: r.slope.aspect <sup>67</sup>
Ruggedness	Elevation standard deviation	<i>elev-stdev</i>	PKTOOLS: pkfilter <sup>69</sup>
	Terrain ruggedness index	<i>tri</i>	GDAL: gdaldem <sup>66</sup>
	Roughness	<i>roughness</i>	GDAL: gdaldem <sup>66</sup>
	Vector ruggedness measure	<i>vrn</i>	GRASS GIS: r.vector.ruggedness.py <sup>67</sup>
	Topographic position index	<i>tpi</i>	GDAL: gdaldem <sup>66</sup>
	Maximum multiscale deviation	<i>dev-magnitude</i>	Whitebox: MaxElevationDeviation <sup>68</sup>
	Scale of the maximum multiscale deviation	<i>dev-scale</i>	Whitebox: MaxElevationDeviation <sup>68</sup>
	Maximum multiscale roughness	<i>rough-magnitude</i>	Whitebox: MultiscaleRoughness <sup>68</sup>
	Scale of the maximum multiscale roughness	<i>rough-scale</i>	Whitebox: MultiscaleRoughness <sup>68</sup>
geomorphological forms	Geomorphon	<i>geom</i>	GRASS: r.geomorphon <sup>67</sup>

**Table 1.** MERIT-DEM derived geomorphometric variables. The 26 geomorphometric variables (second column) are grouped based on their main mathematical operations (first column). The abbreviations in the third column correspond to the variable name of the file layer available for download.

From the perspective of algorithm development, the optimal approach is to calculate the geomorphometric variables using latitude and longitude in a spherical domain rather than treating the raster as a square grid. Unfortunately, all geomorphometric algorithms implemented in Geographic Information Systems (GIS) and remote sensing software packages treat the latitudinal-longitudinal rasters as square grids. Therefore, the only current solution is to reproject the raster to a grid system that minimises distortions.

We decided to use the Equi7 projection having considered its positive attributes as well as minor drawbacks. This decision was primarily driven by Equi7's efficient data storage features, minimal area oversampling/distortion, and continental continuity<sup>36</sup>. In addition, we selected this scheme based on an assessment of the effect of slope values under WGS84 and Equi7. Our rationale for selecting the Equi7 is further discussed in the section "Evaluation of the geographic projection".

**Derived geomorphometric variables.** Geomorphometric layers, also known as topographic variables or topographic indices, can be derived from DEMs. The MERIT-DEM (as well as 3DEP-1 and LIDAR DEM) was used to calculate the 26 derived geomorphometric layers listed in Table 1. The derived layers were calculated based on a set of grid cells in the immediate vicinity of each focal cell, as defined by a moving window analysis. The size of the moving window was set to a  $3 \times 3$  cell grid for most of the cases. Nonetheless, two layers (Multiscale deviation and Multiscale roughness) were calculated based on moving window size variations (more information is available in the variable description sections). Overall, the 26 variables can be grouped as those that: (i) quantify the rate of change (first derivative - 11 layers; second derivative - 5 layers), (ii) describe the ruggedness (9 layers), and (iii) identify the geomorphological forms (1 layer). We describe each single geomorphometric variable below, which also points to the acronym labelled in italics in the text below and reported in column 2 of Table 1. The variable acronyms also correspond to the file names stored in the data repositories<sup>14–16</sup>.

**Terrain first derivatives.** In calculus, the first derivative is defined as the rate of change of a function, and in a geometric sense it is the slope of the tangent line of a function. When the first derivative is applied to a DEM, it represents the terrain slope of a relief. The terrain slope is the measure of steepness and is one of the most fundamental geomorphometric features that plays an important role in several natural phenomena such as soil accumulation, water infiltration, and snow depth, to mention a few. Slope can be computed in several ways in conjunction with vector direction. Details of each calculation are included below.

**Slope.** Terrain slope (*slope*), as mentioned above, is the rate of change of elevation in the direction of the water flow line. It is considered one of the most important terrain parameters and is often calculated first. It can be expressed in degrees or percentages, where for example, 5% means 5m of vertical displacement over 100m. It is especially important for the quantification of soil erosion, water flow velocity, or agricultural suitability<sup>29</sup>.

**Aspect.** Aspect (*aspect*) is the angular direction that a slope faces. It is expressed in degrees and therefore defined as a circular variable. We calculated the sine (*aspect-sine*) and cosine (*aspect-cosine*) of the aspect, changing a circular variable to a continuous variable, and allowing future processing, including reprojection under a bilinear algorithm or use them as continuous variables in regression analysis. The sine and cosine of the aspect, ranging from  $-1$  to  $1$ , can be used to emphasise differences in the north-south and east-west exposure<sup>5</sup>.

**Eastness and northness.** Using aspect and slope, we calculated northness (*northness*) and eastness (*eastness*). The sine of the slope when multiplied by the cosine of the aspect yields the northness, and when multiplied by the sine of the aspect provides the eastness<sup>5,39</sup>. Eastness and northness provide continuous measures describing the orientation in combination with the slope. In the northern hemisphere, a northness value close to  $1$  corresponds to a northern exposure on a vertical slope (i.e. a slope exposed to very low amount of solar radiation), while a value close to  $-1$  corresponds to a very steep southern slope exposed to a high amount of solar radiation. Eastness and northness has been often used in plant species distribution and forest mapping<sup>40</sup> and also in the spatio-temporal estimation of snow depth<sup>41</sup>.

**Convergence.** The convergence index (*convergence*)<sup>42</sup> is a terrain variable that highlights the convergent areas as channels and divergent areas as ridges. It ranges from  $-100$  for ridges to  $+100$  for sink areas and  $0$  for planar or flat areas. In combination with the curvature parameters it is useful for delineating different landforms. The convergence index has been used in several studies regarding tree species distribution analyses and for down-scaling climate data over complex terrains<sup>43</sup>.

**Compound topographic index.** The compound topographic index (*cti*)<sup>21</sup>, also known as topographic wetness index, is computed as the logarithm of the cumulative upstream catchment area divided by the tangent of the local slope angle. This index is a proxy of the long-term soil moisture availability<sup>44</sup>. It has been often used in applications that include species distribution modelling, species richness and composition, landslide susceptibility and soil carbon assessment<sup>44,45</sup>.

**Stream power index.** The stream power index (*spi*)<sup>8</sup> is computed as the product between the upstream catchment area and the tangent of the local slope angle. The stream power index reflects the erosive power associated with flow and the tendency of gravitational forces to move water downstream<sup>8</sup>. It is commonly used in soil erosion models, landslide susceptibility and groundwater estimation.

**First directional derivatives.** Directional derivative (*d*) is the rate of change of the elevation in a specific direction. In particular the East-West first order partial derivative (*dx*) is the slope in an East-West direction, while the North-South first order partial derivative (*dy*) is the slope in a North-South direction. The first directional derivatives (*dx* and *dy*) can be used to estimate overland water flow and sediment flow by means of the SIMWE model<sup>46</sup>. Moreover, directional slope has been used to detect artefacts (voids, pits, sinks, sensor stripes) in DEMs<sup>19</sup>, due to its sensitivity to systematic noise such as striping, or artefacts such as voids and pits, in the DEMs.

**Terrain second derivatives.** In calculus, the second derivative is the derivative of a derivative. In other words, it is the rate of change of the slope and represents the curvature or concavity of a function. When the second derivative is applied to a DEM it represents the rate change of slope or aspect in a particular direction. The unit of curvature is radians per metre, where positive and negative values indicate convex and concave surfaces, respectively. Terrain curvatures directly affect soil erosion and composition, water accumulation and infiltration, and therefore indirectly drive the presence and composition of flora and fauna. Terrain curvatures can also be used as input parameters for hydrological and soil erosion modelling.

**Profile and tangential curvature.** Profile curvature (*pcurv*) measures the rate of change of a slope along a flow line, and affects the acceleration of water flow along a surface<sup>29</sup>. The tangential curvature (*pcurv*) measures the rate of change perpendicular to the slope gradient and is related to the convergence and divergence of flow across a surface<sup>29</sup>. The analysis of curvatures allows one to understand how water and sediments move through the landscape and helps to quantify their accumulation or dispersal<sup>7,47</sup>.

**Second directional derivatives.** The second directional derivative is the rate of change of the slope in a predetermined direction: the East-West second order partial derivative (*dxx*) is the derivative of a slope in an East-West

direction, while a North-South second order partial derivative ( $dyy$ ) is the derivative of the slope in a North-South direction.

**Terrain ruggedness.** Surface roughness is a common topographic attribute and is frequently measured using DEMs. It describes the ruggedness and topographic complexity (elevation variability) of landscapes within an area. Roughness maps are derived by measuring topographic variability around each grid cell in a moving window approach. The roughness is scale-dependent as a function of the moving window size and will significantly influence the final roughness map<sup>48</sup>. Described below are five indices computed using a classic  $3 \times 3$  moving window approach, and two roughness indices obtained from a multiscale analysis, by progressively increasing the moving window size.

**Elevation standard deviation.** Standard deviation ( $elev-stdev$ ) is a measure of the amount of variation within a dataset. The standard deviation of elevation was calculated using a  $3 \times 3$  moving window. Values close to 0 indicate no variation, (i.e. flat areas), while areas with high standard deviation indicate areas with very steep terrain.

**Terrain ruggedness index.** The terrain ruggedness index ( $tri$ ) is a mean of the absolute differences in elevation between a focal cell and its 8 surrounding cells. It is a type of statistical variance of elevation change across the  $3 \times 3$  cells<sup>49</sup>. Flat areas have a value close to zero, while mountainous areas have positive values that can be greater than 500 m.

**Roughness.** Roughness ( $roughness$ )<sup>50</sup> is expressed as the largest inter-cell absolute difference of a focal cell and its 8 surrounding cells. It is expressed in unit length, in our case metres, and is always positive, ranging from zero values in flat areas to progressively larger positive values in mountain areas. This variable is a measure based on a maxima, therefore it is more sensitive to the artefacts that remain in the MERIT-DEM.

**Vector ruggedness measure.** The vector ruggedness measure ( $vrms$ )<sup>51</sup> quantifies terrain ruggedness by measuring the variation by means of sine and cosine of the slope in the three-dimensional orientation of grid cells, within a moving window. Slope and aspect are decomposed into 3-dimensional vector components (in the x, y, and z directions) using standard vector analysis in a user-specified moving window size ( $3 \times 3$ ). In other words, it captures variability of slope and aspect in a single measure. The vector ruggedness measure quantifies local variation of slope in the terrain more independently than the topographic position index and terrain ruggedness index methods<sup>51</sup>. It is dimensionless because of sine-cosine derivation, and values range from 0 to 1 in flat to rugged regions, respectively.

**Topographic position index.** The topographic position index ( $tpi$ )<sup>29,52</sup> is the difference between the elevation of a focal cell and the mean of its 8 surrounding cells. It ranges from positive to negative values and they correspond to ridges and valleys, respectively. Zero values correspond to flat areas.

**Multiscale deviation.** The deviation from mean elevation ( $dev$ ) is a unitless measure of topographic position, and the difference between the elevation of the centre cell and mean elevation divided by the standard deviation of the entire window<sup>53</sup>. The deviation from the mean elevation range is unbounded ( $-\infty, +\infty$ ), and a positive or negative sign indicates whether the central cell is above or below the surrounding mean elevation. Furthermore, the magnitude value indicates the relative spread of the elevation distribution in its surrounding area<sup>54</sup>. The multiscale analysis of the deviation consists of the estimation of spatial patterns using a range of window sizes. The maximum value of the multiscale deviation identifies the Maximum Elevation Deviation value ( $dev-magnitude$ ) and the window size ( $dev-scale$ ) where the maximum value is depicted<sup>53</sup>. To calculate the multiscale deviation over a range of spatial scales we vary the moving window dimensions ranging from  $3 \times 3$  to  $4001 \times 4001$  grid cells, by a constant increment of 3 grid cells. To our knowledge this is the first time that the multiscale deviation variables have been calculated at global scale.

**Multiscale roughness.** The multiscale roughness<sup>55</sup> ( $rough$ ) is the spherical standard deviation ( $\sigma_s$ ) of the sum of 3-dimensional vector components derived to calculate the  $vrms$ . Its units are degrees. It can be computed in moving windows of different sizes, and in case of a  $3 \times 3$  moving window, it corresponds to  $rough = \sigma_s(vrms)$ . Lindsay *et al.*<sup>55</sup> described the analytical aspect and computation parameters in detail. If these deviations are large, the surface is rough. On the other hand, if they are small, the surface is smooth<sup>48</sup>. The multiscale roughness variation follows a similar distribution to the multiscale deviation and its maximum values ( $rough-magnitude$ ) identify magnitude and scale ( $rough-scale$ ). To calculate multiscale roughness over a range of spatial scales, we vary the moving window dimensions ranging from  $3 \times 3$  to  $4001 \times 4001$  grid cells, by a constant increment of 3 grid cells (internal iterations). Likewise, to our knowledge this is the first time that the multiscale roughness variables have been calculated at global scale at 90 m resolution. A quasi-global representation of the MERIT-derived maximum multiscale roughness, is depicted in Fig. 1. From a computational perspective, the multiscale roughness and deviation variables have been the most difficult to calculate. This was due to the wide tile overlapping ( $>4001$  grid cells), useful to avoid border effects, and due to internal iterations that retain the maximum values. We decided to plot the  $rough-magnitude$  in Fig. 1 as a quasi-global visualisation, due to its novelty and practical utility as locally adaptable and scale-optimised analyses for mapping applications<sup>55</sup>.

**Geomorphological forms.** Geomorphon. The geomorphological forms ( $geom$ ) consist of 10 classes that can be extracted from DEMs using morphometry techniques<sup>56</sup>. This technique identifies geomorphological phenotypes

also known as geomorphons. It is based on pattern recognition rather than differential geometry and thus has high computational efficiency. It classifies the terrain in terms of the following features: flat, peak or summit, ridge, shoulder, spur, slope, hollow, footslope, valley, and pit or depression (class label-number for each geomorphon are respectively: 1,2,3,4,5,6,7,8,9,10; schematic representation in Figs. 3 and 4b of Amatulli *et al.*<sup>5</sup>). These geomorphon classes have been used in a wide range of studies such as landslide susceptibility mapping<sup>57</sup>, human mobility<sup>58</sup>, and ecosystem service assessments<sup>59</sup>. Other variables can be derived by the analysis of the geomorphon shapes: intensity, range, variance, extend, azimuth, elongation, width<sup>56</sup>. As they are considered experimental, we decided do not compute these variables.

A static global visualisation of the geomorphometric variables under WGS84 geodetic datum can be seen on the Geomorpho90m webpage<sup>60</sup>. Additionally, more detailed patterns are shown in Fig. 5 of Amatulli *et al.*<sup>5</sup>. Moreover, the correlation amongst all geomorphometric variables is shown in the correlation matrix reported in Figure 7 of Amatulli *et al.*<sup>5</sup>. Even if the underlying DEM has a coarse resolution of 250 m, the correlation and general pattern will be similar to the 90 m resolution variables. The matrix can be used to select variables in accordance with the concerned case study and as a function of the employed modelling technique.

## Data Records

**Data repository.** The Geomorpho90m dataset is a set of gridded layers stored as GeoTIFF files. It is derived from the 90 m MERIT-DEM, and is global in extent (60°–85°N latitude), including all continents except for Antarctica. Subsets of the 26 geomorphometric variables can be visualised on OpenLandMap WebGIS<sup>61</sup> under the thematic layers “Relief/Geology”. These layers can also be downloaded as three different products:

- at 100 m resolution, under Equi7, available to download at Spatial-Ecology repository<sup>14</sup>.
- at 3 arc-seconds (~90 m) resolution, under WGS84, available to download at OpenTopography repository<sup>15</sup>.
- at 7.5 arc-seconds (~250 m) resolution, under WGS84, available to download at PANGAEA repository<sup>16</sup>.

**File nomenclature.** The file name identifies: the geomorphometric variable abbreviation (see Table 1), the spatial resolution, the DEM source (MERIT) and the tiling system, within the following structure:

*variable abbreviation\_resolution\_DEM source layers\_tiling system.format*

The layers under the Equi7 projection are considered to contain the most reliable values owing to minimal geographic distortion. For our Geomorpho90m dataset we use the T6 tiling system and tile nomenclature (more info at the Equi7 github<sup>38</sup>) proposed by Bauer-Marschallinger *et al.*<sup>36</sup>. We encourage users to use the Equi7 projection especially for studies at a continental or global scale.

Below are two examples of the layer names under Equi7:

- *tri\_100M\_MERIT\_AF\_006\_066.tif*: layer showing the terrain ruggedness index at a 100 m spatial resolution in Equi7 stemming from the MERIT-DEM in Africa with the tile position 006\_066.
- *rough-magnitude\_100M\_MERIT\_AF\_006\_066.tif*: layer showing the maximum multiscale roughness at the identical location (tile position 006\_066).

The majority of users utilise the WGS84 geodetic datum, and hence we reprojected the layers from Equi7 to WGS84 at a 3 arc-second (~90 m) spatial grain. Here, we used the tiling system implemented in the MERIT-DEM dataset (more info at MERIT website<sup>18</sup>). Each tile covers a 5 × 5 degree (6000 × 6000 cell) extent, while the tile name describes the position of the lower left pixel of the layer.

Below are two examples of the layer names under WGS84:

- *slope\_90M\_MERIT\_s30e125.tif*: layer showing the slope at a 3 arc-second spatial resolution in the WGS84 stemming from the MERIT-DEM in Australia with the tile position s30e125.
- *aspect\_90M\_MERIT\_s30e125.tif*: layer showing the aspect at the identical location of the *slope\_90M\_MERIT\_s30e125.tif*.

In addition, each geomorphometric layer at a spatial resolution of 3 arc-seconds and 100 m was stored as 32-bit floating point (Float32 data type) for maximum precision, allowing for the computation of other customised variables (e.g., coefficient of variation) or for aggregating the layers at coarser spatial resolutions for macro-scale environmental models.

For the OpenLandMap WebGIS<sup>61</sup> visualisation, we reprojected the layers from Equi7 to WGS84 (EPSG:4326 code<sup>35</sup>) with a spatial resolution of 7.5 arc-second (~250 m), stored as 16-bit integer (Int16 or UInt16 data type; scale factor is reported in the GeoTIFF metadata) to enable a fast visualisation rendering.

## Dataset comparison

The quality of the underlying DEMs, in terms of the vertical and spatial accuracy, directly influences the quality of the newly-developed geomorphometric layers. In this section we assess the sensitivity of the geomorphometric layers with respect to DEM accuracy and we describe it in three subsections organised as follows: i) Evaluation of the geographic projection, where we show the possible artefacts that could stem from computing the geomorphometric variable under WGS84 geodetic datum and the Equi7 projection; ii) MERIT-DEM versus 3DEP-1-derived geomorphometric layers, where we describe the divergence of the most common geomorphometric layers obtained from MERIT-DEM and 3DEP-1 DEMs; iii) MERIT-DEM vs LiDAR elevation comparison, which

outlines the influence of removing the tree height bias in MERIT-DEM using the DTM and DSM obtained from LiDAR. Overall, these analyses highlight the quality of the MERIT-derived geomorphometric layers and make it possible to identify potential errors in DEMs.

**Evaluation of the geographic projection.** Cartographic projections are required to map the Earth's surface on a 2-D plane and this is particularly relevant for DEMs. Regardless of the projection used, some type of distortion will occur in the resulting map but the selection of a suitable projection should, in principle, minimise the extent and type of distortions<sup>36</sup>. In general, map distortions diminish as the geographic area is reduced, for instance, when moving from global to continental or regional scales. Moreover, distortions increase as one travels along a surface, away from the projection centre. This distortion is an unavoidable property of map projections, and it is important to assess its effect on any type of spatial analysis, particularly on those carried out on a large scale. To assess the effect of map distortions on the geomorphometric variables, we analysed the slope variations under two geographic locations having different surface distortions.

The slope is defined as the rate of elevation change along the direction of the water flow, and calculated using a  $3 \times 3$  cell moving window. The rate of change can be expressed as a percentage of elevation change over 100 metres. In order to have the same weight in the  $x$  and  $y$  directions on the rate of change, the cell size must have the same dimension in the  $x$  and  $y$  directions. This is not the case when the Geographic Coordinate System is used and specifically where the longitudinal gradient (in  $y$  dimension) stays constant with respect to the latitudinal gradient (in  $x$  dimension), which decreases away from the equator.

Since terrain slope is one of the most widely used geomorphometric variables, all GIS and remote sensing software have adopted algorithms to compute it. However, none of these algorithms employ a correction procedure to account for grid distortion in the  $x$  and  $y$  dimension. In other words they treat latitude, longitude data as a matrix on a square grid. Instead, it is established that meridians converge towards the Poles and parallel circles' distance varies only slightly. Thus, a square grid at the Equator becomes a rectangular grid at higher latitudes. Therefore, to quantify the influence of changes in latitudinal dimension of  $x$ , we compared slope values under WGS84 with those under Equi7 for two study areas of  $500 \times 500$  grid cells, with MERIT-DEM as a static base layer. This procedure can be shown using a simulated DEM under two distinct locations, one in the subtropical zone and the other in the subarctic zone. Nonetheless, to demonstrate a real potential effect, we select only one zone from MERIT-DEM.

The graphs presented in Fig. 2 show the difference in slope calculations as a direct result of using latitude and longitude on WGS84 with those from a square grid on Equi7. To compare the same MERIT-DEM under WGS84 and Equi7 at two distinct locations we transpose (i.e., Equi7 shift coordinates - note: no-reprojection) the subtropical zone MERIT-DEM (image centre: longitude -83.26, latitudes 9.05; in Costa Rica) (Fig. 2g) to a subarctic zone (image centre: longitude -38.19, latitudes 72.80; in Greenland) (Fig. 2a), under Equi7. This produces a simple displacement along the north-south axis without changes in the elevation pixel value (no-interpolation). After reprojecting the MERIT-DEM to WGS84 (Fig. 2b,h), we calculated slope (Fig. 2f,d) and subsequently reprojected it back to Equi7 (blue line-arrows in Fig. 2) to compare the results using the scatter plots. Figure 2i corresponds to the slope correlation of the subarctic zone, whereas Fig. 2j corresponds to the subtropical zone). In each scatter-plot (Figs. 2i,j), the red line represents a 1:1 relationship, while the black line is a fitted regression model between the variables. The variations in slope calculations between the two systems are minimal within the subtropical zone (Fig. 2j), as the study area is adjacent to the equator. However, in the subarctic zone, the variations are significantly different, as the slope calculated under WGS84 is underestimated compared to Equi7 slope. This is because the east or west-facing slopes will have their gradient significantly underestimated due to the stretching of  $x$  dimensions in the east-west direction (note all the points notably beneath the red line of Fig. 2i). On the other hand, the north- and south-facing slopes may be moderate correctly estimated (note all the points close to the red line of Fig. 2i).

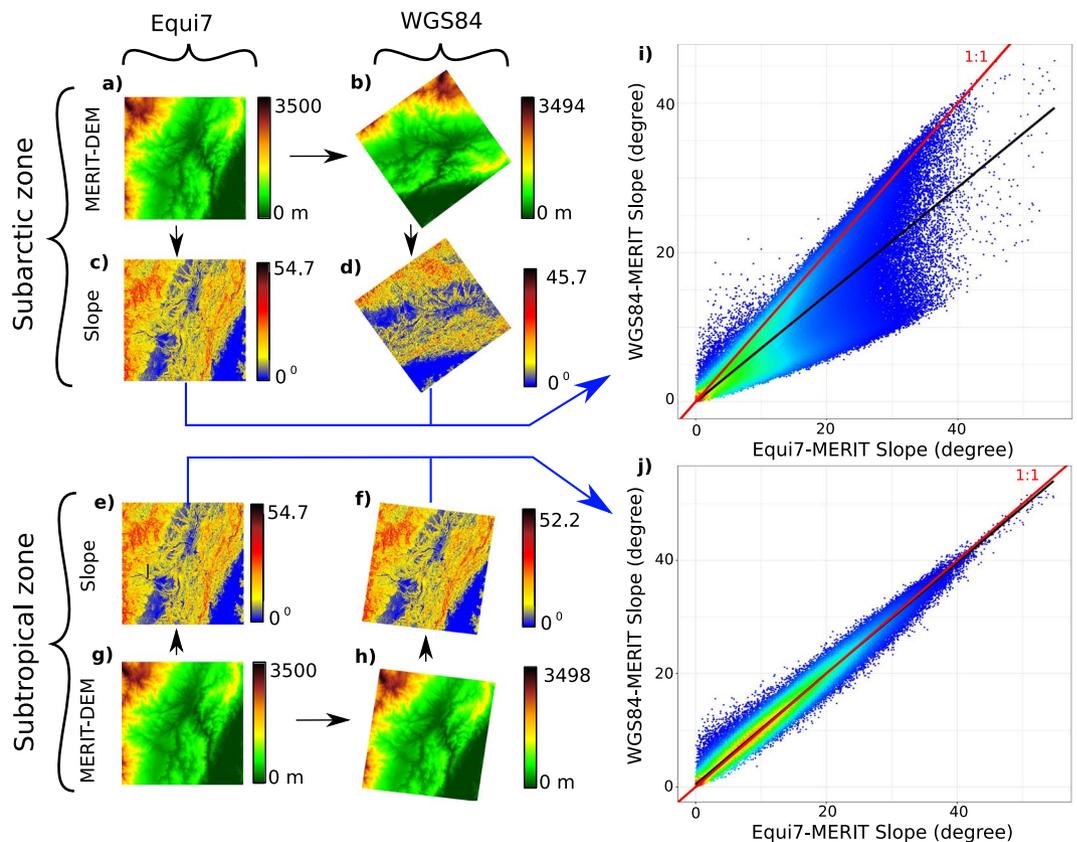
Similar to slope, all geomorphometric variables are influenced by underlying grid distortions. In particular, the slope is influenced by both length and angular distortions, as are all of the other geomorphometric variables listed under the first and second derivatives group. In contrast, the ruggedness geomorphometric variables are influenced more by areal distortions because of elevation differences at the pixel level. These results emphasise the importance of computing the geomorphometric variables under the Equi7 projection.

In conclusion, it is not that the WGS84 geodetic datum is wrong and distorted but its treatment of latitudinal and longitudinal grids as squares is erroneous, as in the Plate Carrée projection. Consequently, the calculation of any geomorphometric variables under WGS84 should be avoided.

**3DEP-1 versus MERIT-DEM comparison.** For geomorphometrical and hydrographical applications, the elevation difference between two DEMs is important since any application will be contingent on the values of the derived geomorphometric variables, for example, impacting on the delineation of streams and catchments. In the following sections, we analyse the difference in the elevation values between the 3DEP-1 and MERIT DEMs, as well as their derived geomorphometric variables. The 3DEP-1 is a LiDAR-based DEM and given its high accuracy, can be used as a reference elevation that has negligible errors. Initially, we compare the elevation difference between 3DEP-1 and the MERIT-DEM, and subsequently we analyse how the differences in DEMs influence the derived geomorphometric variables.

*Comparing DEMs using the Elevation Deviation Index (EDI).* The elevation difference, or deviation, at pixel level between two DEMs can be expressed as

$$\vec{\Delta}_i = x_i - y_i \quad (1)$$



**Fig. 2** Projection bias assessment. Graphical representation of the difference in terrain slope calculation due to the effect of using the World Geodetic System 1984 (WGS84) (raster panels right-hand side) compared to the Equi7 projection (raster panels left-hand side). A study area located in the subtropical zone (image centre: longitude  $-83.26$ , latitudes  $9.05$ ) was used to subset the MERIT-DEM for an area of  $500 \times 500$  grid cells (g). This area has been transposed to a subarctic zone (image centre: longitude  $-38.19$ , latitudes  $72.80$ ) under the Equi7 projection (a). After having been reprojected to WGS84 (b,h), the variable slope was calculated in the four conditions (c–f), and then reprojected back to Equi7 for comparisons (see blue line-arrows). The scatter plots on the right-hand side shows the WGS84-MERIT slope (d,f) vs. the MERIT slope under the Equi7 projection (c,e), respectively for the subarctic zone (i) and for the subtropical zone (j). The red lines represent the 1:1 relationship and black lines represent a linear model between the two axes. The slope calculated under WGS84 in the subarctic zone is clearly underestimated compared to the one calculated under the Equi7 projection.

where  $x$  and  $y$  are the elevation values in each single pixel  $i$ . The  $\vec{\Delta}_i$  value is equal to 0 if the two DEMs have the same elevation. The overall raster of  $\vec{\Delta}_i$  values represents the  $\Delta$  surface.

To identify areas where the deviation is stronger, the deviation at each pixel needs to be considered with the surrounding elevation pixel values ( $x_{i+1}$ ). In our case, we label 3DEP-1 as  $y$  and MERIT-DEM as  $x$ . Therefore, considering a circular window of  $23 \times 23$  pixels that slides across  $y$ , it is possible to obtain the standard deviation, which estimates the local elevation roughness. Mathematically, the standard deviation of  $y_i$  in a moving circular window is expressed as:

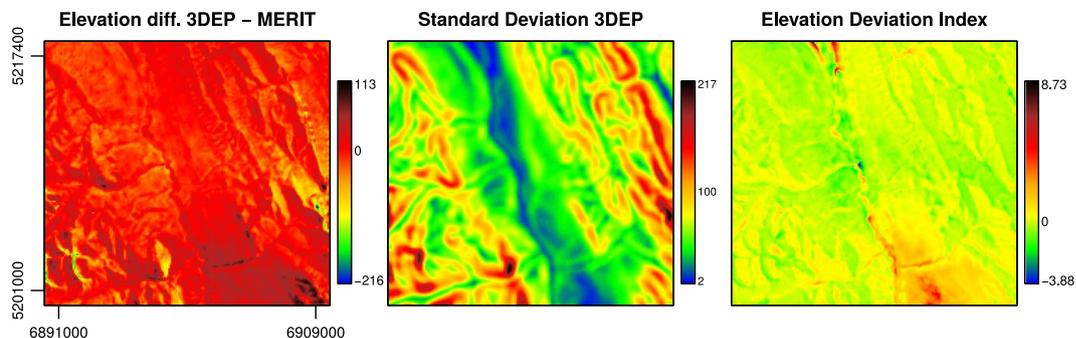
$$\sigma_i = \sqrt{\frac{1}{N-1} \sum_i^N (y_i - \bar{y})^2} \quad (2)$$

If we integrate the  $\vec{\Delta}_i$  and its surrounding standard deviation, we obtain the Elevation Deviation Index (EDI), which is defined as the ratio

$$EDI = \vec{\Delta}_i / (\sigma_i + k) \quad (3)$$

EDI represents the relative deviation over the surrounding elevation variability in the moving window. The component  $k$ , in above equation, is used to prevent the situation that areas completely flat, with  $\sigma_i = 0$ , will produce infinite values of the EDI. In our case, we set  $k = 0.1$ , which is a very small value compared to the calculated  $\sigma_i$ , even in quasi-flat areas.  $k$  does not influence the  $\sigma_i$  and consequently the overall performance of the EDI.

For example, a local elevation difference of 1 m will create a higher index in flat areas compared to mountain regions; and the index can be positive or negative with respect to the  $\vec{\Delta}_i$ . The EDI can be used to select zones



**Fig. 3** Elevation Deviation Index. Elevation Deviation Index  $EDI$  (c) obtained as the ratio of elevation difference (a: 3DEP-1 - MERIT) and elevation standard deviation calculated using a moving window of  $5 \times 5$  pixels of the 3DEP-1 (b). The coordinates reported in a are in Equi7 and expressed in metres. The study area refers to a zone of  $18.4 \times 20$  km, which has a high level of forest cover, and is located in Alberta, Canada, close to Jasper National Park - image centre  $118.25^\circ\text{W}$   $53.29^\circ\text{N}$ . The same area is used in the Fig. 4 to assess the geomorphometric variables.

where the elevation difference between the 3DEP-1 and the MERIT-DEM is substantial considering the roughness of the surrounding areas. Hence, flat areas will be more sensitive to the  $EDI$  compared to steep, mountainous areas. We expect that areas with extreme  $EDI$  will be more prone to deviating stream networks compared to zones with  $EDI$  close to 0 (see Fig. 3).

When comparing DEMs with unknown or significant errors, the standard deviation of the mean of the two DEMs can be calculated. Besides, this standard deviation is a measure of roughness, and the window size (Eq. 2) reflects neighbouring influences. A large window size will produce a larger standard deviation and thus lower  $EDI$ , on average. The moving window size can be adjusted with respect to the resolution of the DEMs or on the basis of the surrounding roughness. The  $EDI$  can be applied on a global scale by comparing different DEMs, and highlighting areas where the DEMs have discrepancies.

Figure 3 shows  $EDI$  and its components for an area of  $18.4 \times 20$  km. The extreme  $\Delta_i$  values (black and blue areas in a) do not necessarily produce extreme  $EDI$ . With respect to the  $EDI$ , in the largest part of the study area, the DEMs are in agreement (yellow - green colour) and located in zones with a high level of roughness. On the contrary, in flat areas (blue colour in b) the  $EDI$  can reach extreme values (blue and dark red colour in c).

**Comparing the continuous geomorphometric variables.** To compare the geomorphometric variables derived from the 3DEP-1 and MERIT-DEM under the same scale unit, we normalise the difference expressed as a  $\Delta$  surface. Hence, we deal with the difference (for example  $\text{pcurv-3DEP-1} - \text{pcurv-MERIT}$ ) by scaling all positive values to fall between 0 and +1, and negative values to fall between -1 and 0. As a result, the difference value at 0 remains at 0 when scaled (e.g. 0, 9 scaled to 0, 1; -3, 0 scaled to -1, 0).

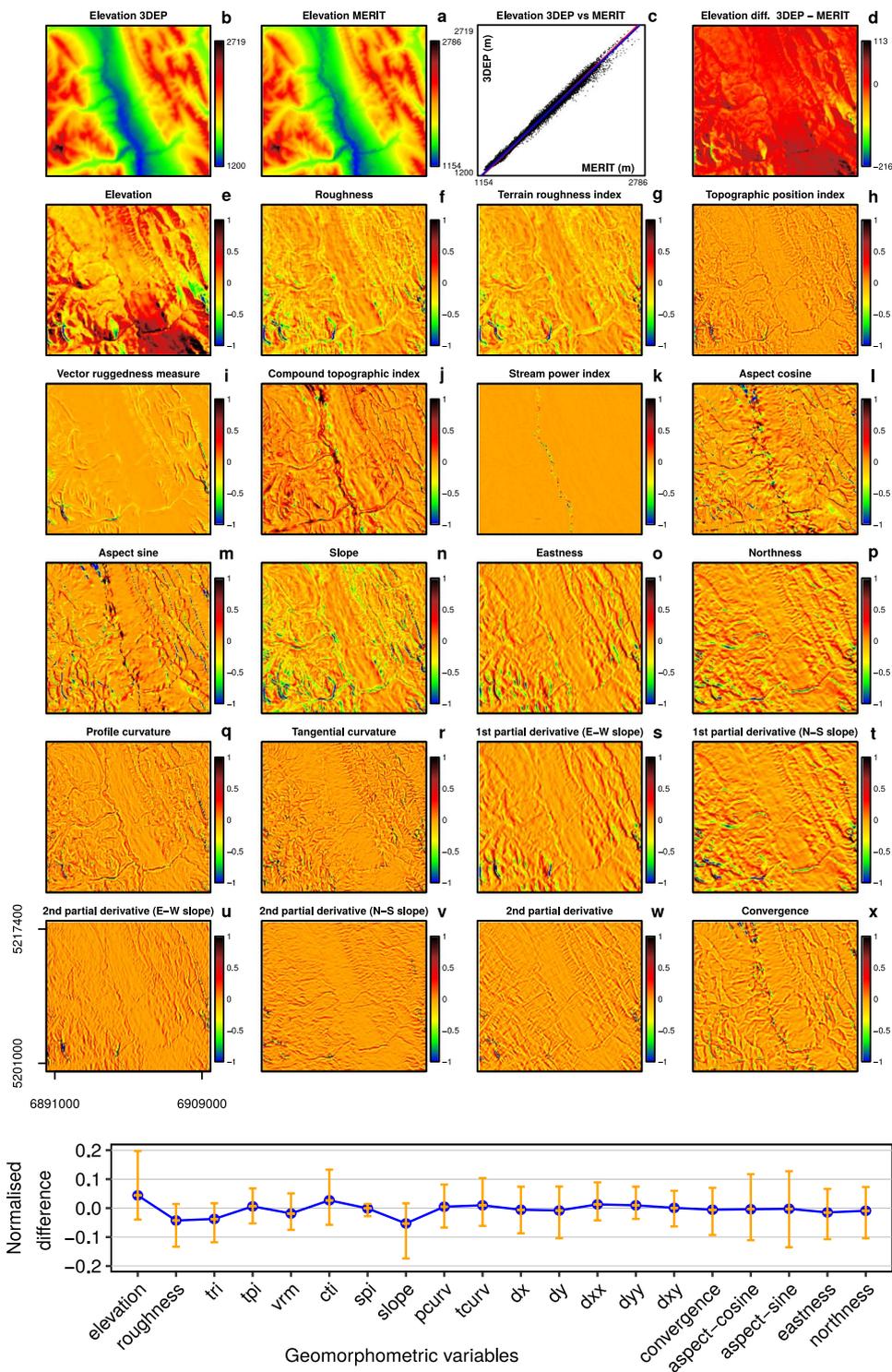
Consequently, the normalised  $\Delta$  surface derivative from each geomorphometric variable can be compared having the same unit and can be used to assess the sensitivity of the variables to the differences in DEM elevation. In fact, in instances where the  $\Delta$  surface has a value close to 0, this suggests that a geomorphometric variable is not strongly influenced by the difference between the two DEMs. In contrast, in instances where the  $\Delta$  surface has several pixels with negative or positive values, this means that they are influenced by the DEM's difference.

Figure 4 shows an overview of the normalised  $\Delta$  surface for each geomorphometric variables. Two elevation plots (Fig. 4a,b) show the 3DEP-1 and MERIT DEMs and relative scatter plot (Fig. 4c). The elevation difference (Fig. 4d) shows values ranging from -113 m to +216 m. The largest values of difference are located close to the peak areas, and the smallest values are concentrated in the valley areas. Figure 4e shows the normalised version. In contrast, the other plots show the spatial variability of the geomorphometric difference, expressed with normalised values. Values close to -1 and +1 mean high sensitivity to elevation difference, and conversely, values close to 0 mean less sensitivity. In general, the overall correlation between 3DEP-1 and MERIT-DEM is very high, with the blue line representing a fitted regression model, which is very close to the 1:1 red line. These results are in line with other studies that evaluate the accuracy of the MERIT-DEM<sup>17,19</sup>.

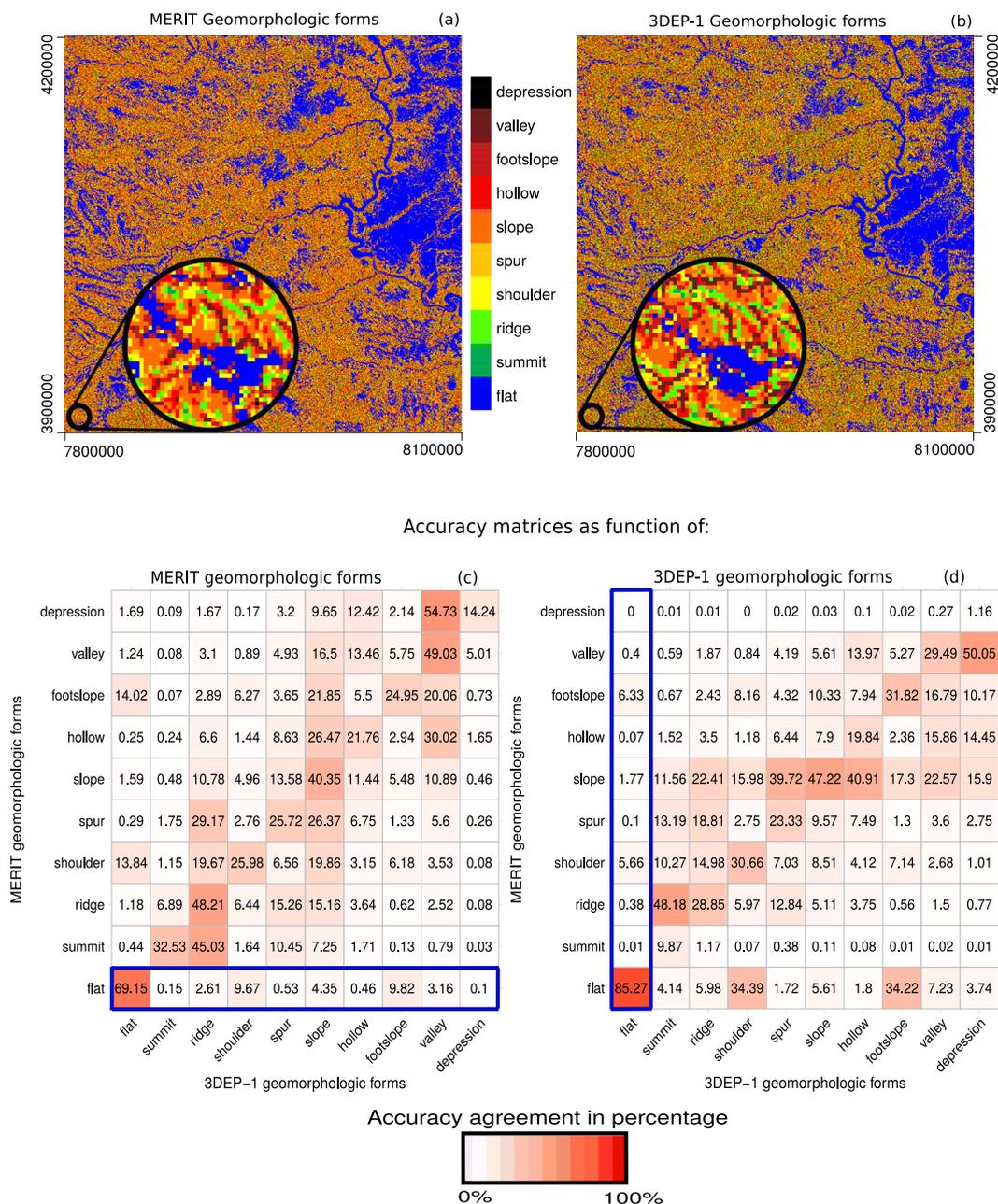
The differences in elevation are mainly due to the radar beam's shadow, which is usually evident in steep terrain. In fact, in Fig. 4d, the area with high elevation difference is located in the south-west corner with  $\Delta$  values larger than 200 m. The greatest relative deviations are in high-relief areas for most of the geomorphometric variables except for the compound topographic index, convergence, and sine and cosine of aspect (see Fig. 4j,x,l,m). For these exceptions, the relative deviations are greatest in low-relief areas, especially in the valleys. Indeed, flat areas are very sensitive to the DEMs accuracy and slight variations in the elevation can switch the aspect to the opposite direction.

The behaviour between the compound topographic index ( $cti$ ) and stream power index ( $spl$ ) differs due to the logarithmic scale used in the  $cti$ . Consequently, the deviation of the  $cti$  is visible when there is a small variation of the flow accumulation. On the other hand, the  $spl$  does not employ a logarithmic scale, and the deviation is only evident when there is a drastic change in the flow accumulation areas that are adjacent to stream locations.

Visually, the aspect-sine, the aspect-cosine, slope, eastness, northness and the convergence are geomorphometric features that are very sensitive to differences between the two DEMs (see Fig. 4f,g,h,x).



**Fig. 4** Normalised difference maps. Normalised difference maps represented as  $\Delta$  surface, for each geomorphometric variable derived from 3DEP-1 minus MERIT-DEM. To compare the geomorphometric variables under the same scale unit, the difference has been scaled from -1 to 1 (minimum and maximum stretching) keeping the 0 value (no difference) at the 0 position. The bottom plot reports the normalised difference as mean (blue line) and standard deviation values (orange vertical lines) of the maps. The mean and standard deviation plot helps to identify which geomorphometric variables are more sensitive to variation in the DEMs (e.g. high sensitivity for variables derived from slope and aspect). The coordinates  $u$  in Equi7 are expressed in metres and refer to a study area of  $18.4 \times 20$  km, which has a high level of forest cover, and is located in Alberta, Canada, close to Jasper National Park - image centre  $118.25^{\circ}\text{W}$   $53.29^{\circ}\text{N}$ .



**Fig. 5** Geomorphological forms maps and confusion matrices. The geomorphological forms have been computed for a study area of  $3000 \times 3000$ , 100 m pixels in South Dakota (USA) derived from MERIT (a–c) and 3DEP-1 (b–d), respectively. The confusion matrix values are expressed in percentages of the MERIT-DEM classes, with the sum of vertical values equal to 100 (d); and of the 3DEP-1 classes, with the sum of the horizontal values equal to 100 (c). The sum of the values in the blue boxes is equal to 100, and so on for each row (c) and column (d).

To support the visual assessment of the maps in Fig. 4 with numerical values, we analysed the normalised  $\Delta$  surface by plotting the mean values and standard deviation for the positive and negative values (see Fig. 4). In fact, similar patterns can also be seen for any of the aforementioned variables with high standard deviation (see vertical lines). The aspect is very sensitive to DEM differences in both steep terrain and flat areas. Even the derived sine and cosine  $\Delta$  surface show black and blue areas (+1 and -1, respectively) in the central valley (see Fig. 4i,m). Note that these areas are not apparent in the other variables. Just as 1st partial derivatives have been used to detect artefacts in DEMs<sup>19</sup>, the  $\Delta$  surface of the aspect-sine and aspect-cosine can be used to highlight areas where the two DEMs show differences in elevation.

*Comparing the categorical geomorphometric variables.* To assess the sensitivity of pattern delineation of the geomorphological forms derived from MERIT-DEM and 3DEP, we compare the geomorphological classification agreement for an area of  $300 \times 300$  km ( $3000 \times 3000$  pixels at 100 m spatial resolution) in South Dakota, USA.

Figure 5 shows two raster plots (*a, b*) with a similar pattern at large scale but when observed in finer detail, there are differences in the classification at the pixel level (see Fig. 5a,b magnified circle). A common way to analyse the differences between two classifications is the calculation of a so-called confusion (or error) matrix<sup>62</sup>. The confusion matrix displays the probabilities with which pixels belonging to a certain class in one product appear in the same or a different class in the compared product. A confusion matrix can therefore be used to illustrate not only the degree to which the two classifications agree but also reveal how likely a class is misclassified.

In order to allow a numerical comparison of the geomorphological classifications, we calculate two confusion matrices among the 10 classes in each product (see Fig. 5). One is expressed as percentage of MERIT-DEM classes such that the sum within each row is equal to 100 (see Fig. 5c). Considering 3DEP-1 as a reference product, this would give the “user accuracy” of the MERIT-DEM geomorphometric classes. For instance, almost 70% of the *flat* pixels in MERIT-DEM also appear to be *flat* in 3DEP-1, while 10% are overlapping either with *foothill* or *shoulder* which are indeed likely spatial neighbours to *flat*. This is possibly an indication for some co-registration or interpolation issues affecting the two products.

The other matrix is calculated to display the percentages with respect to the 3DEP-1 classification, i.e. each column will sum up to 100 (see Fig. 5d). It shows the likelihood with which a 3DEP-1 class appears in the same or other classes of MERIT, which is called “producer accuracy”, e.g. 86% of *flat* pixels in 3DEP-1 are also correctly classified as *flat* in MERIT. Additionally, the most likely confusion here is again with *foothill* or *shoulder* classes (both around 6%), which support the above assumption of a co-registration issue. Another interesting finding is the widespread confusion between the *summit* and *ridge* class.

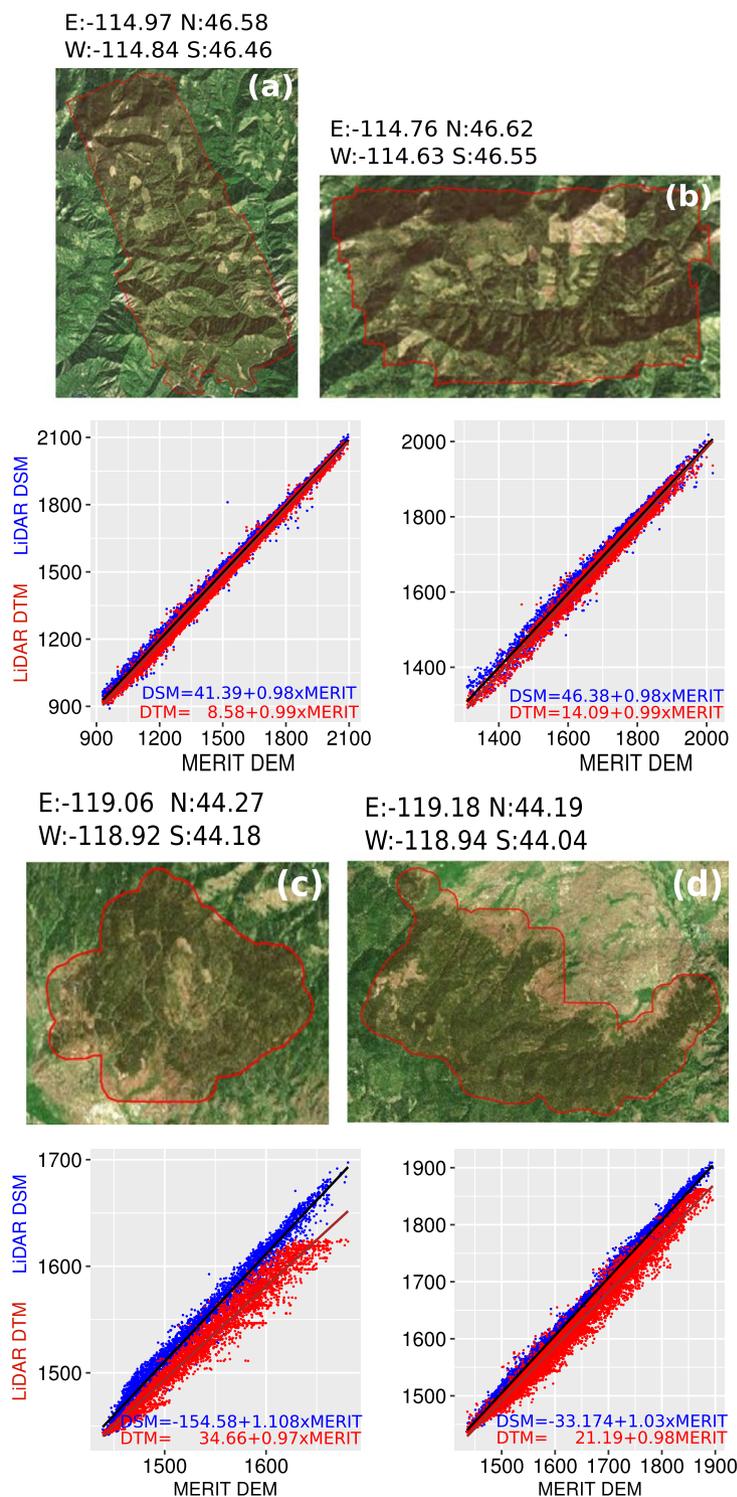
The congruence of *summit* pixels between the two products is in fact less likely than their respective confusion with *ridges* in the other. In addition, there are at least three times as many *summit* pixels detected in 3DEP-1 compared to MERIT. Similar anomalies occur for the morphologically inverse classes *depression* and *valley*. The reason for these results could either be due to an increased richness of detail (or actual resolution) offered by 3DEP-1 or a higher level of noise (though the latter being less likely given its high level of detail). Nevertheless, this preliminary analysis shows that the underlying DEM data yield significantly different geomorphometric characteristics and that the confusion matrix allows these differences to be numerically expressed.

**Comparing MERIT-DEM vs LiDAR elevation.** Last return points in LiDAR data, which penetrate dense vegetation, are used to extract the DTM, whereas the first returns that hit the canopy of vegetation are used to derive the DSM. In our case, the LiDAR DTM and DSM were used to assess the quality of the tree height removal procedure carried out for the MERIT-DEM. Figure 6 reports the DTM vs. MERIT-DEM (red points) and the DSM vs. MERIT-DEM (blue points) of four study areas in USA with a high forest cover. In the four scatter plots, it is possible to distinguish the height difference between the DTM and DSM. The MERIT-DEM dataset has been corrected for the tree height bias<sup>17</sup>, and consequently the MERIT-DEM elevation values are expected to be closer to the LiDAR DTM than those of the LiDAR DSM. The LiDAR DTM vs. MERIT, and the LiDAR DSM vs. MERIT-DEM differences are analytically quantified by the linear model depicted in the scatter plot of Fig. 6. Where there is no vegetation or low vegetation (e.g. agricultural plains - Fig. 6c,d; bare ground mountain tops - Fig. 6a,b), the DTM and DSM have almost identical values, denoted by an overlap of red and blue points. The presence of similar elevation values in the DTM and DSM causes a convergence of the linear models (blue and red lines; the linear model functions are reported at the bottom of each scatter plot). The convergence in the lower parts of the plot (lower elevations) corresponds with the landscape (flat areas) and the absence of forest cover, which contributes to very similar values in the DTM and DSM. Whereas, at higher elevations, where there is increased forest cover, there is a greater difference between the DTM and DSM. This phenomena is more evident in Fig. 6c,d. On the contrary, Fig. 6a,b show a more parallel trend of blue and red lines, which is due to forest cover that is more equally distributed along the relief. It is important to note that the scatter plots are plotted with different elevation ranges (x- and y-axis), and therefore the deviation of the red and blue line appears more evident in plots with lower elevation range (Fig. 6c). The regression coefficients for 3DEP-1 DTM vs. MERIT-DEM are slightly below 1 with an intercept value ranging from 8 to 34 m. On the other hand, the 3DEP-1 DSM vs. MERIT-DEM correlation has a regression coefficient larger than 1 and an intercept ranging from -154 to 46. The mean difference between DTM and DSM for Fig. 6a–d study areas ranges from 12 to 21 m. Overall, these values demonstrate the strong correlation between the MERIT-DEM and the LiDAR DTM.

**Identification of artefacts.** It is important to note that the use and application of MERIT is based on the understanding that while it currently represents the best quality DEM available, it still contains errors and artefacts, which have not been corrected within the context of this research, as this was considered beyond the scope of the specific research objective. Consequently these errors cascade into the new Geomorpho90m dataset. The effect of these errors are mainly due to stripes that were recurrent in the AW3D product. These are visible in flat areas, where the artefact error is larger than the delta between pixels (e.g. slope). For instance, stripe artefacts can be found at the following locations in western Russia<sup>63</sup> as well as in central Russia<sup>64</sup>.

### Usage Notes

The Geomorpho90m dataset<sup>14–16</sup> is the first global scale geomorphometric layer product at a spatial resolution of 3 arc-seconds (~90 m) and 100 m, and has the potential to open new research avenues for a variety of research disciplines that require detailed geomorphometric and land surface information. For instance, the new layers can provide essential input data for analysing and modelling patterns and processes in physical geography, hydrological and climate science, land-use and land cover change, ecology, biogeography, conservation and biodiversity science. For hydrological applications in particular, the Geomorpho90m dataset creates, in combination with other environmental layers, the basis for computing freshwater-specific variables as per the procedure described in<sup>65</sup>. Most importantly, the product’s global coverage ensures that a standardised set of geomorphometric layers that



**Fig. 6** MERIT-DEM vs LiDAR-DEM. Comparison of MERIT-DEM with the LiDAR DSM and LiDAR DTM for four study areas, represented by their scatter-plots and their relative linear models.

enable comparative analyses across continents. Additionally, the newly-developed layers provide an update for the previous GMTED2010-derived topographic variables, as described in Amatulli *et al.*<sup>5</sup>. The new Geomorpho90m dataset can be considered more accurate in terms of its spatial resolution than the one by Amatulli *et al.*<sup>5</sup> (90 vs. 250 m) and reduces potential residual errors from the corrections applied in the underlying MERIT-DEM dataset<sup>17</sup>. Therefore, the Geomorpho90m dataset provides deeper insights into the scale-dependency of geomorphometric characteristics across a wide variety of land surface-based studies.

The Geomorpho90m dataset<sup>14–16</sup> provides an update of the “GMTED2010-derived topographic variables” described at Amatulli *et al.*<sup>5</sup>.

## Code availability

### Geomorphometric layer computation

Prior to computing the geomorphometric layers, we reprojected the DEMs (MERIT, 3DEP-1, LiDAR DMS and DTM) to the Equi7 projection with a cell size 100 m (the projection parameters are available from<sup>38</sup>). To harmonise the different spatial grains, we used a bilinear algorithm implemented in *gdalwarp* within the open-source Geospatial Data Abstraction Library (GDAL). We kept the seven projection zones as defined in<sup>36</sup>, and employed the T6 tiling method to allow parallel and distributed processing of our work-flow. Tile size was 600 × 600 km where we buffered the borders by 401 km to avoid border artefacts between tiles. These overlapping and duplicate grid cells were removed when merging all tiles to seamless, continental maps. These large tile size increments were needed to avoid border effects, especially for multiscale deviation and multiscale roughness.

We used the following open source software packages to compute the geomorphometric layers (Table 1 reports the software and the specific commands used for each derived variable calculation):

Geospatial Data Abstraction Library (GDAL), version number 2.1.2<sup>66</sup>.

Geographic Resources Analysis Support System software (GRASS), version number 7.3.0<sup>67</sup>.

Whitebox Geospatial Analysis Tools (Whitebox GAT), version number 3.3.0<sup>68</sup>.

Processing Kernel for geospatial data (Pkttools), version number 2.6.3<sup>69,70</sup>.

All of these tools provide fast and scalable computation features and functions for raster-based workflows that are easily automated using a scripting language, such as Bash or Python<sup>71</sup>. They also allow for the processing of very large datasets owing to efficient algorithms and optimised memory management. After computing all geomorphometric layers within the Equi7 projection, the layers were reprojected back to the WGS84 coordinate reference system (EPSG:4326 code<sup>35</sup>) with a bilinear algorithm (or near for categorical variables) implemented in GDAL. This reversion of Geomorpho90m to WGS84 allows for the data to be seamlessly integrated with a broad set of global datasets.

We used a tiling system identical to MERIT-DEM, in terms of dimension and nomenclature, i.e. 5 × 5 degree tiles with 6000 × 6000 cells each, to ensure data integration and comparisons with the original MERIT-DEM. All calculations were processed in parallel using open-source software at the Center for Research Computing, Yale University.

### LiDAR data processing

Two common products that can be extracted from LiDAR data are: DTM and DSM. The DTM is generated using ground echoes from the LiDAR point cloud, in conjunction with an interpolation technique<sup>72</sup>. The approach employed in this paper for calculating the DTMs used the LiDAR processing tools found in the pkttools software<sup>70</sup>. This is a two-stage approach, where the first stage uses a minimum composite rule, which retains the LiDAR pulse for the minimum height of each cell.

```
pklas2img -a_srs EPSG:26911 -dx 100 -dy 100 -comp min -i input.las -o dtm_min.tif
```

For the second stage, the output is then filtered using a progressive morphological filter<sup>73</sup>, which uses an iterative filter based on increasing kernel sizes to remove non-ground points from the final DTM.

```
pkfilterdem -f promorph -dim 3 -dim 11 -i dtm_min.tif -o dtm_min_promorph.tif
```

The calculation of the DSM is more straightforward and uses either a maximum composite rule (pulse retention with the maximum height of the cell) or a defined percentile composite rule. In our case, we use the rule to retain the pulse with the value corresponding to the 95th percentile of all pulses within the cell.

```
pklas2img -a_srs EPSG:26911 -dx 100 -dy 100 -comp percentile -percentile 95 -i input.las -o dsm2.tif
```

The LiDAR projection parameters (EPSG code) were set to EPSG:26911 in accordance with the associated metadata. The DSM and DTM cell size was set to 100 m to allow a simple reprojection to Equi7, which enables a comparison with the other discussed layers.

Received: 12 March 2019; Accepted: 31 March 2020;

Published online: 28 May 2020

## References

- Pike, R. J. Geomorphometry-diversity in quantitative surface analysis. *Progress in physical geography* **24**, 1–20 (2000).
- Florinsky, I. V. An illustrated introduction to general geomorphometry. *Progress in Physical Geography* **41**, 723–752 (2017).
- Alexander, C., Deák, B. & Heilmeyer, H. Micro-topography driven vegetation patterns in open mosaic landscapes. *Ecological indicators* **60**, 906–920 (2016).
- Stein, A. & Kreft, H. Terminology and quantification of environmental heterogeneity in species-richness research. *Biological Reviews* **90**, 815–836 (2015).
- Amatulli, G. *et al.* A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific data* **5**, 180040 (2018).
- Florinsky, I. *Digital terrain analysis in soil science and geology* (Academic Press, 2016).
- Sofia, G. Combining geomorphometry, feature extraction techniques and earth-surface processes research: The way forward. *Geomorphology* **355**, 107055 (2020).
- Moore, I. D., Grayson, R. & Ladson, A. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes* **5**, 3–30 (1991).
- Neteler, M. & Mitasova, H. *Open source GIS: a GRASS GIS approach*, vol. 689 (Springer Science & Business Media, 2013).
- Amatulli, G., Rodrigues, M. J., Trombetti, M. & Lovreglio, R. Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences* **111**, 1–15 (2006).
- Grunewald, T. *et al.* Statistical modelling of the snow depth distribution in open alpine terrain. *Hydrology and Earth System Sciences* **17**, 3005–3005 (2013).

12. Farahmand, A. & AghaKouchak, A. A satellite-based global landslide model. *Natural Hazards and Earth System Sciences* **13**, 1259–1267 (2013).
13. Danielson, J. J. & Gesch, D. B. *Global multi-resolution terrain elevation data 2010 (GMTED2010)* US Department of the Interior, US Geological Survey (2011).
14. Amatulli, G. *et al.* Geomorpho90m: technical documentation. Spatial Ecology, <http://www.spatial-ecology.net/dokuwiki/doku.php?id=topovar90m> (2020).
15. Amatulli, G. *et al.* Geomorpho90m - Global High-Resolution Geomorphometry Layers. *OpenTopography* <https://doi.org/10.5069/G91R6NPX> (2020).
16. Amatulli, G. *et al.* Geomorpho90m - Global high-resolution geomorphometry layers. *PANGAEA* <https://doi.org/10.1594/PANGAEA.899135> (2020).
17. Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophysical Research Letters* **44**, 5844–5853 (2017).
18. Yamazaki, D. MERIT-DEM: Multi-error-removed improved-terrain dem, [http://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT\\_DEM/](http://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT_DEM/) (2019).
19. Hirt, C. Artefact detection in global digital elevation models (DEMs): The maximum slope approach and its application for complete screening of the srtm v4. 1 and MERIT DEMs. *Remote Sensing of Environment* **207**, 27–41 (2018).
20. Moudry, V. *et al.* On the use of global DEMs in ecological modelling and the accuracy of new bare-earth DEMs. *Ecological Modelling* **383**, 3–9 (2018).
21. Beven, K. J. & Kirkby, M. J. A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal* **24**, 43–69 (1979).
22. Kellndorfer, J. *et al.* Vegetation height estimation from shuttle radar topography mission and national elevation datasets. *Remote sensing of Environment* **93**, 339–358 (2004).
23. Farr, T. G. *et al.* The shuttle radar topography mission. *Reviews of geophysics* **45**, 1–33 (2007).
24. Walker, W. S., Kellndorfer, J. M. & Pierce, L. E. Quality assessment of SRTM c-and x-band interferometric data: Implications for the retrieval of vegetation canopy height. *Remote Sensing of Environment* **106**, 428–448 (2007).
25. Robinson, N., Regetz, J. & Guralnick, R. P. Earthenv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused aster and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing* **87**, 57–67 (2014).
26. Harding, M. & Carabajal, C. Icesat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters* **32** (2005).
27. Hansen, M. *et al.* High-resolution global maps of 21<sup>st</sup>-century forest cover change. *Science* **342**, 850–853 (2013).
28. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences* **116** (2011).
29. Gallant, J. & Wilson, J. *Terrain analysis: principles and applications* (John Wiley & Sons, 2000).
30. USGS Team. *The 3D Elevation Program Initiative – A Call for Action*. United States Geological Survey, <https://www.usgs.gov/core-science-systems/ngp/3dep> (2018).
31. Glenn, N., Streutker, D. R., Chadwick, J., Thackray, J. & Dorsch, S. Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology* **73**, 131–148 (2006).
32. Roggero, M. Airborne laser scanning: Clustering in raw data. *International Archives of Photogrammetry Remote Sensing* **34-3**, 227–232 (2001).
33. LiDAR Survey of the Malheur National Forest, Oregon. *OpenTopography* <https://doi.org/10.5069/G9QJ7F74> (2012).
34. Clearwater NF, ID: Effects of Watershed Restoration on Hillslope Stability. *OpenTopography*, <https://doi.org/10.5069/G9H41PB3> (2012).
35. Spatial Reference Team. World geodetic system. Spatial Reference, <https://spatialreference.org/ref/epsg/wgs-84/> (1984).
36. Bauer-Marschallinger, B., Sabel, D. & Wagner, W. Optimisation of global grids for high-resolution remote sensing data. *Computers & Geosciences* **72**, 84–93 (2014).
37. Hengl, T. *et al.* Soilgrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).
38. Bauer-Marschallinger, B., Sabel, D. & Wagner, W. Equi7 grids projection parameters, <https://github.com/TUW-GEO/Equi7Grid/tree/master/equi7grid/grids> (2014).
39. Fassnacht, S., Dressler, K. & Bales, R. Snow water equivalent interpolation for the colorado river basin from snow telemetry (snotel) data. *Water Resources Research* **39** (2003).
40. Crowther, T. W. *et al.* Mapping tree density at a global scale. *Nature* **525**, 201 (2015).
41. Collados-Lara, A.-J., Pardo-Igúzquiza, E. & Pulido-Velazquez, D. Spatiotemporal estimation of snow depth using point data from snow stakes, digital terrain models, and satellite data. *Hydrological processes* **31**, 1966–1982 (2017).
42. Claps, P., Fiorentino, M. & Oliveto, G. Informational entropy of fractal river networks. *Journal of Hydrology* **187**, 145–156 (1996).
43. Fridley, J. D. Downscaling climate over complex terrain: high finescale (<1000 m) spatial variation of near-ground temperatures in a montane forested landscape (great smoky mountains). *Journal of Applied Meteorology and Climatology* **48**, 1033–1049 (2009).
44. Radula, M. W., Szymura, T. H. & Szymura, M. Topographic wetness index explains soil moisture better than bioindication with ellenberg's indicator values. *Ecological Indicators* **85**, 172–179 (2018).
45. Román-Sánchez, A., Vanwalleghem, T., Peña, A., Laguna, A. & Giráldez, J. Controls on soil carbon storage from topography and vegetation in a rocky, semi-arid landscapes. *Geoderma* **311**, 159–166 (2018).
46. Mitasova, H., Mitas, L. & Brown, W. M. Multiscale simulation of land use impact on soil erosion and deposition patterns. In *Sustaining the Global Farm. Selected papers from the 10th international Soil Conservation Meeting*. Purdue University (2001).
47. Stefano, C. D., Ferro, V., Porto, P. & Tusa, G. Slope curvature influence on soil erosion and deposition processes. *Water resources research* **36**, 607–617 (2000).
48. Lindsay, J. B. & Newman, D. R. Hyper-scale analysis of surface roughness. *PeerJ Preprints* **6**, e27110v1 (2018).
49. Riley, S. J. Index that quantifies topographic heterogeneity. *Intermountain Journal of sciences* **5**, 23–27 (1999).
50. Beasom, S. L., Wiggers, E. P. & Giardino, J. R. A technique for assessing land surface ruggedness. *The Journal of Wildlife Management* **47**, 1163–1166 (1983).
51. Sappington, J. M., Longshore, K. M. & Thompson, D. B. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the mojave desert. *The Journal of wildlife management* **71**, 1419–1426 (2007).
52. Jenness, J. Topographic position index (tpi). *Flagstaff, AZ: Jenness Enterprises* (2006).
53. Lindsay, J., Cockburn, J. & Russell, H. An integral image approach to performing multi-scale topographic position analysis. *Geomorphology* **245**, 51–61 (2015).
54. Newman, D., Lindsay, J. & Cockburn, J. Evaluating metrics of local topographic position for multiscale geomorphometric analysis. *Geomorphology* **312**, 40–50 (2018).
55. Lindsay, J. B., Newman, D. R. & Francioni, A. Scale-optimized surface roughness for topographic analysis. *Geosciences* **9**, 322 (2019).
56. Jasiewicz, J. & Stepinski, T. F. Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology* **182**, 147–156 (2013).
57. Luo, W. & Liu, C.-C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* **15**, 465–474 (2018).
58. Conrad, N. D., Helfmann, L., Zonker, J., Winkelmann, S. & Schütte, C. Human mobility and innovation spreading in ancient times: a stochastic agent-based simulation approach. *EPJ Data Science* **7**, 24 (2018).

59. Underwood, E. C., Hollander, A. D., Huber, P. R. & Schrader-Patton, C. Mapping the value of national forest landscapes for ecosystem service provision. In *Valuing Chaparral*, 245–270 (Springer, 2018).
60. Amatulli, G. Geomorpho90m: Technical documentation. Spatial Ecology, <http://www.spatial-ecology.net/dokuwiki/doku.php?id=topovar90m> (2020).
61. OpenLandMap Team. OpenLandMap an open land data project. OpenGeoHub, <https://openlandmap.org> (2019).
62. Stehman, S. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* **62**, 77–89 (1997).
63. OpenLandMap Team. Western Russia artefacts. OpenGeoHub <https://openlandmap.org/#/?base=Stamen> (2019).
64. OpenLandMap Team. Central Russia artefacts. OpenGeoHub, <https://openlandmap.org/#/?base=Stamen> (2019).
65. Domisch, S., Amatulli, G. & Jetz, W. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scientific data* **2**, 1–13 (2015).
66. GDAL Development Team. *GDAL - Geospatial Data Abstraction Library, Version 2.2.3*. Open Source Geospatial Foundation, <https://gdal.org/> (2017).
67. GRASS Development Team. *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*. Open Source Geospatial Foundation, <http://grass.osgeo.org> (2017).
68. Lindsay, J. B. Whitebox GAT: A case study in geomorphometric analysis. *Computers & Geosciences* **95**, 75–84 (2016).
69. Kempeneers, P. *PKTOOLS - Processing Kernel for geospatial data, Version 2.6.7.6*, <http://pktools.nongnu.org/html/index.html> (2018).
70. McInerney, D. & Kempeneers, P. *Open Source Geospatial Tools - Applications in Earth Observation* (Springer Verlag, 2015).
71. Amatulli, G. *et al.* Teaching spatiotemporal analysis and efficient data processing in open source environment. In *Proceedings of the 3rd Open Source Geospatial Research & Education Symposium* **13** (2014).
72. Lindsay, J. & Creed, I. Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEMs. *Photogrammetric Engineering & Remote Sensing* **9**, 1029–1036 (2005).
73. Zhang, K. *et al.* A progressive morphological filter for removing nonground measurements from airborne lidar data. *IEEE transactions on geoscience and remote sensing* **41**, 872–882 (2003).

## Acknowledgements

This study was supported in part by the facilities and staff at the Yale centre for Research Computing (YCRC). We wish to thank the YCRC's staff for the help and advice regarding the cluster computation. Special thanks go to Dr. Pieter Kempeneers, the developer of the pktools software<sup>59</sup>. His tools were fundamental in executing a fast processing chain. G.A. was supported by Yale University in particular from Institute for Biospheric Studies, School of Forestry & Environmental Studies, Centre for Research Computing. S.D. was funded by the Leibniz Association through the Leibniz Competition (grant number J45/2018). The publication of this article was funded by the Open Access Fund of the Leibniz Association and by the Open Access Fund of the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB). This work benefited from discussions as part of the “Global freshwater biodiversity, biogeography and conservation” project (<https://glowabio.org/>).

## Author contributions

G.A. designed the study and developed/implemented the computational methodology and the processing chain in the Yale-HPC to derive all derived continuous/categorical geomorphometric variables, standardised and validated the dataset layers and wrote the first manuscript draft; D.M. designed and performed the LiDAR processing required for DEM accuracy assessment. D.M. T.S. P.S. and S.D. contributed to the writing of the manuscript and discussed the results.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.A. or S.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020