Data Descriptor: Geomorpho90m - Global high-resolution geomorphometry layers: empirical evaluation and accuracy assessment

Giuseppe Amatulli[1,2*], Daniel McInerney [3], Tushar Sethi [4], Peter Strobl [5], Sami Domisch [6]

March 15, 2019

1. Yale University, School of Forestry & Environmental Studies, New Haven, CT 06511, USA.
2. Yale University, Center for Research Computing, New Haven, CT 06511, USA.
3. Coillte, Irish State Forestry Board, Plassey Road, Castletroy, Limerick, Ireland.
4. Spatial-Ecology, Meaderville House, Wheal Buller, Redruth, Cornwall, TR16 6ST, UK.
5. European Commission, Joint Research Centre, Directorate for Sustainable Resources, Ispra, Italy,
6. Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department of Ecosystem Research, Berlin, Germany.

*corresponding author: Giuseppe Amatulli (giuseppe.amatulli@gmail.com)

Abstract

Topographical relief is composed of the vertical and horizontal variations of the Earth's terrain and drives processes in geography, climatology, hydrology, and ecology. Its assessment and characterisation is fundamental for various types of modelling and simulation analyses. In this regard, the Multi-Error-Removed Improved Terrain (MERIT) Digital Elevation Model (DEM) is the best global, high-resolution DEM currently available at a 3 arc-seconds (90 m) resolution. This is an improved product as multiple error components have been corrected from the underlying Shuttle Radar Topography Mission (SRTM3) and ALOS World 3D - 30 m (AW3D30) DEMs. To depict topographical variations worldwide, we developed the Geomorpho90m dataset comprising of different geomorphometry features derived from the MERIT-DEM. The fully standardised geomorphometry variables consist of layers that describe (i) the rate of change using the first and second order derivatives, (ii) the ruggedness, and (iii) the geomorphology landform. To assess how remaining artifacts in the MERIT-DEM could affect the derived topographic variables, we compared our results with the same variables generated using the 3D Elevation Program (3DEP) DEM, which is the highest quality DEM for the United States of America. We compared the two data sources by calculating the first order derivative (i.e., the rate of change through space measured in degrees) of the difference between a MERIT-derived vs. a 3DEP-derived topographic variable. All newly-created topographic variables are readily available at resolutions of 3 and 7.5 arc-seconds under the WGS84 geographic system, and at a spatial resolution of 100 m under the Equi7 projection. The newly-developed Geomorpho90m dataset provides a globally standardised dataset for environmental models and analyses in the field of geography, geology, hydrology, ecology and biogeography. All newly-developed variables are available for download at Data Citation 1 (7.5 arc-sec) and GoogleDrive `https://drive.google.com/drive/folders/1D4YHUycBBhNFVVsz4ohaJI7QXV9BEh94` (3 arc-sec and 100m) and for visualization at LandGIS web-gis portal `https://landgis.opengeohub.org`.

Background & Summary

Geomorphometry is the science of quantitative analysis of the Earth's surface [1]. 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 to better understand patterns and processes in geography, geology, climatology, hydrology or biodiversity science.

Generally speaking, a DEM is a combination of a Digital Terrain Model (DTM) and a Digital Surface Model (DSM). The DSM is a 3D elevation model of the Earth surface that includes objects, such as tree and buildings, whereas the DTM represents the surface without any objects, returning the elevation at the bare ground. While DEMs provide the elevation above see levels, a wide array of geomorphometry metrics (also know as topographic, or geomorphometry variables) can be derived to yield information regarding topographical variation and land surface parameters. This spatial information

improves our understanding of the geographical, geomorphological and environmental properties of a given study area [2]. From the sole use of the DEMs, it is possible to describe the topographical complexity, which can shape the macro and micro climate of a given area [3]. The topographical variation contributes significantly to the environmental complexity of a region, and also defines the biotic and abiotic features at a sub-regional level [4].

Land surface parameters are quantitative measures of various morphometric properties of a surface. The most common parameters, slope or aspect, can be used to further derive more complex features or curvature profiles of a terrain at any given location. Such measures are central to hydrological parameters shaping flow and erosion processes within the landscape, and to delineate catchment and stream features [5]. Furthermore, mapping and assessing landform variability such as concavity and convexity is essential to obtain a better understanding of land erosion and landscape denudation dynamics in mountainous environments [6]. Beside, morphometric properties of a surface are important also to predict other phenomena such as wildfire risk modeling [7], mountain/alpine snow cover [8] and landslide formation[9].

While it is possible to obtain geomorphometric information from the latest DEMs on a case-by-case basis, standardisation is necessary to enable spatially comparative analyses between regions and continents. To this end, [4] established a suite of 15 geomorphometric variables based on 7.5 arc-second (∼250 m) Global Multi-resolution Terrain Elevation (GMTED) data. Building upon this work, we present the Geomorpho90m dataset (Data Citation 1). Here we extend this concept and calculated a suite of 26 DEM-derived geomorphometric variables based on the Multi-Error-Removed Improved Terrain (MERIT) DEM at a 3 arc-second (∼90 m) spatial resolution [10], which to-date is considered the best-effort in global DEMs at ∼90 m resolution [11, 12]. We perform the computation of these geomorphometric variables under the Equi7 projection [13], which minimises the pixel-level distortions that otherwise often occur when unprojected data (i.e. latitude-longitude in the World Geodetic System (WGS)) are being treated as if they were a rectilinear raster. We compare and validate the newly created variables with those generated using the DEMs derived from 3D Elevation Program (3DEP) and Light Detection and Ranging (LiDAR).

The Geomorpho90m dataset (Data Citation 1) comprises the following 26 variables: 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, geomorphology landform. All layers are globally available at a spatial resolution of ∼90 m and ∼250 m in the WGS, and at 100 m in the Equi7 coordinate reference system, providing a unique opportunity for a variety of applications in environmental modelling on such a high spatial resolution. The Geomorpho90m dataset is available at low resolution (250 m) from Data Citation 1, whereas the high resolution (90 m) data is available from GoogleDrive `https://drive.google.com/drive/folders/1D4YHUycBBhNFVVsz4ohaJI7QXV9BEh94`. A visualization of the layers is provided at the LandGIS web-gis portal `https://landgis.opengeohub.org`. (for reviewer: LandGIS web-gis portal is ingesting the data and will be ready when publication will be accepted. Meanwhile the data can be download from Data citation 1 and GoogleDrive. A preview of the geomorphometry layers is also offered at `http://www.spatial-ecology.net/dokuwiki/doku.php?id=topovar90m`).

Methods

The principal dataset employed in this study was the Multi-Error-Removed Improved Terrain (MERIT) DEM 3 arc-seconds (∼90 m) [10] and it was used to extract geomorphological terrain features. In addition, ancillary DEMs were used for the purposes of comparison, which included Light Detection and Ranging (LiDAR) derived products and data from the 3D Elevation Program (3DEP). The resultant geomorphological variables were a suite of 26 layers calculated from the source layers using a moving window analysis.

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 spheroidal reference surface: the datum. The datum for the WGS was established in 1984 and last revised in 2004. This WGS is therefore referred to as WGS84 or EPSG:4326.

Any kind of quasi-spherical surface represented in 2D (map format) can lead to three types of potential distortion: (i) length distortion; (ii) angular distortion; and (iii) areal distortion. These distortions are caused when the spherical coordinates of WGS84 are considered as a rectilinear grid (i.e. Plate Carrée or Universal Transverse Mercator projections) and increase with a latitudinal gradient, with very high areal/length distortion values in the Sub Arctic/Antarctic zones. Therefore,

any environmental values biased by the distortion, influence geographic analyses especially in the northern hemisphere [14]. These distortions result in local data oversampling when one projects generic satellite images to a regular raster grid [14]. The Grid Oversampling Factor (GOF) metric was therefore formulated to estimate local data oversampling, and indirectly quantify the overall distortion caused due to reprojection [14]. The GOF was used as a baseline, against which multiple projections were compared to determine the effect of the distortions.

To minimise the overall areal and shape distortions in the WGS84, the Equi7 Equidistant Azimuthal was recently proposed [14] and used in global mapping studies [15]. The Equi7 grid causes the least geometric distortion globally, minimising the GOF to a global overall of ∼1, compared to global and hemispherical grids. The Equi7 grid 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 covers the entire Earth with no gaps, and has a 50 km overlap between land borders. It is optimised for the storage and processing of global high resolution satellite 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. 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) [14]. The predefined tiling systems are useful for the implementation of a multi-process computation that does not require the information of the full continental zone.

We decided to use the Equi7 projection, in view of the aforementioned attributes, as they allow the most efficient data storage while minimising area oversampling and distortion. We selected this projection following an assessment of the effect of slope values under WGS84 and Equi7. Our rationale for selecting the Equi7 is further discussed in the section "Geographic projection evaluation".

Source Data

Multi-Error-Removed Improved Terrain - MERIT

In recent years there have been many research initiatives to improve the DEM datasets derived from spaceborne products such as the Shuttle Radar Topography Mission (SRTM) products and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM. The improvements to these DEMs have largely involved the removal of artefacts contained within them. One example of the improved DEM products is the Multi-Error-Removed Improved Terrain (MERIT) DEM 3 arc-seconds (90 m) [10]. 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 removes inherent features or errors found in these products, such as tree height bias, absolute bias, stripe noise and speckle noise, through a combination of ancillary layers [16, 17] and filtering techniques, such as 2D-Fourier Transform filters and the adaptive-scale smoothing filter [18]. The product covers the land area between 90N and 60S at a spatial resolution of approximately 90 m at the equator, and is considered to be the optimal best-effort DEM that is currently available on a global scale [11, 12]. We reprojected the MERIT dataset from WGS84 to Equi7 using a bilinear interpolation method and the projection parameters previously described, and subsequently computed the geomorphometry variables.

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) [19]. The 3DEP products are delivered using consistent datum, elevation units and a coordinate reference system. The DEMs consist of two principal projects: the project-based DEMs produced from LiDAR for the full areal extent of projects, and the Seamless 3DEP DEM product. Both of these are generated from the combination of high-quality project data for the entire country. Each of these products are distributed at different spatial resolutions, and mosaicked and edge-matched.

For the purposes of this analysis, the Seamless 1 arc-second (∼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 collected in 2014, which meets 3DEP specifications for horizontal accuracy and pulse spacing, as well as the resulting bare earth DEMs.
- 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.

Since 3DEP-1 was created from a set of data sources with varying resolutions and projections, each dataset was converted to standardised units and a coordinate reference system. The elevation values were expressed in decimal units, while the NAD83 datum was used for the horizontal projection, and all of the data were recast in a Geographic Coordinate Reference System and delivered as 1x1 degree tiles (see additional metadata information provided by the USGS [1]). Prior to the computation of the geomorphometry variables, 3DEP-1 was reprojected to Equi7 using the projection parameters previously described.

Light Detection and Ranging - LiDAR

Light Detection and Ranging (LiDAR) is a remote sensing method that measures distances using a laser. The LiDAR sensor emits thousands of pulses of light 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 [20]. The DTM is a model of the bare earth, while the DSM is a model of the surface which includes the vegetation and tree canopy. The digital model datasets are generated using algorithms that convert irregular point cloud datasets to gridded rasters [21]. 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 assess the effect of tree height correction in MERIT, we compared to DTMs and DSMs derived from several LiDAR datasets. The LiDAR data (DOI: 10.5069/G9QJ7F74 and DOI: 10.5069/G9H41PB3) was downloaded from OpenTopography (`https://opentopography.org/`). Extensive forest cover was evident through a visual inspection in Google Earth, in conjunction with the forest cover map created by [16]. The presence of forest canopy allowed for the creation of DTM and DSM, which differed substantially. This in turn allowed for the comparison of the LiDAR-derived DTM with MERIT, as the latter excludes tree height.

Derived geomorphometry variables

Geomorphometry layers, also known as topographic variables or topographic indices, can be derived from DEMs. The MERIT-DEM (and secondary the 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 x 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 geomorphometric 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 geomorphology landforms (1 layer). We describe each single geomorphometry 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 acronym also correspond to the file name in Data Citation 1.

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. Slope can be computed in several ways, in accordance with vector direction. Details of each calculation are included below.
Slope. Terrain slope (*slope*) or rate of change of elevation is the measure of steepness 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 5-m of vertical displacement over 100-m. It is especially important for the quantification of soil erosion, water flow velocity, or agricultural suitability [18].
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 manipulation, including reprojection under the bilinear algorithm. The sine

[1] `https://www.usgs.gov/core-science-systems/ngp/3dep/3dep-product-metadata`

and cosine of the aspect, ranging from -1 to 1, can be used to emphasise differences in the north/south and east/west exposure [4].

Eastness and northness. Using aspect and slope, we calculated northness (*northness*) and eastness (*eastness*). The sine of the slope when multiplied by the cosine yields the northness, and when multiplied by the sine of the aspect provides the eastness [22, 4]. 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 [23] and also in the spatio-temporal estimation of snow depth [24].

Convergence. The convergence index (*convergence*) [25] 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 in tree species distribution and for down-scaling climate data over complex terrains [26].

Compound topographic index. The compound topographic index (*cti*) [13], also know 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 [27]. It is has been often used in applications that include species distribution modelling, species richness and composition, landslide susceptibility and soil carbon assessment [27, 28].

Stream power index. The stream power index (*spi*) [5] 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 [5]. 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 [29]. Moreover, directional slope has been used to detect artefacts (voids, pits, sinks, sensor stripes) in DEMs [11], 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 effect 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, if convex or its deceleration, if concave [6]. 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 [6, 18]. Concave curvature promotes convergent flow and therefore soil deposition, whereas convexity leads to dispersal, which implies increased soil erosion [6].

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 a 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 [30]. Described below are five indices computed using a classic 3x3 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 3x3 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 x 3 cells [31]. Flat areas have a value close to zero, while mountainous areas have positive values that can be greater than 2000 m, for example, in the Himalayan region.

Roughness. Roughness [32] (*roughness*) is expressed as the largest inter-cell absolute difference of a focal cell and its 8 surrounding cells. It is always positive, ranging from zero values in flat areas and large positive values in mountain areas.

Vector ruggedness measure. The vector ruggedness measure (*vrm*) [33] quantifies terrain ruggedness by measuring the variation in a 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 (3x3). The vector ruggedness measure quantifies local variation of slope in the terrain more independently than the topographic position index and terrain ruggedness index methods [33]. Values range from 0 to 1 in flat to rugged regions, respectively.

Topographic position index. The Topographic position index (*tpi*)[18, 34] is the difference between the elevation of a focal cell and the mean of its 8 surrounding cells. It ranges from positive and 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 [35]. 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 [36]. 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 [35]. To calculate the multiscale deviation over a range of spatial scales we vary the moving window dimensions ranging from 3x3 to 4001x4001 grid cells, by a constant increment of 3 grid cells. To our knowledge this is the first time that the multiscale deviation variables has been calculated at global scale.

Multiscale roughness. In general, roughness (*rough*) is the deviation of the normal vector from a hypothetical surface, and its units are degrees. In our case, we can quantify the angular deviations as the neighbourhood-averaged difference in the normal vectors of the original DEM and a smoothed DEM surface [30, 37]. Smoothed surfaces can be computed by applying a mean filter of the same size as the moving window. If these deviations are large, the surface is rough. On the other hand, if they are small, the surface is smooth. 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 x 3 to 4001x4001 grid cells, by a constant increment of 3 grid cells. To our knowledge this is the first time that the multiscale roughness variables has been calculated at global scale.

Geomorphology landforms

Geomorphon. The geomorphology landforms consist of 10 classes that can be extracted from DEMs using morphometry techniques [38]. This technique identifies geomorphological phonotypes also know 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, ridge, shoulder, spur, slope, hollow, footslope, valley, and pit (schematic representation in Figure 3, Figure 4 *b* of [4]). These geomorphon classes have been used in a wide range of studies such as landslide susceptibility mapping [39], human mobility [40], and ecosystem service assessments [41].

Table 1: The list of 26 MERIT-DEM derived geomorphometry variables (second column) is grouped based on their main mathematical operations (first column). The abbreviations in the third column correspond to the variable name of the file layer. Layer names are a combination of: variable abbreviation _ resolution - _ DEM source layers. For example, tri_250M_MERIT.tif, aspect-cosine_90M_MERIT.tif. All the data are stored on Data Citation 1.

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

Code availability

Geomorphometry layer computation. Prior to computing the geomorphometry 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 [42]). To harmonize 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 [42], and employed the T6 tiling method to allow parallel and distributed processing of our work-flow. Tile size was 600 x 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. The so large tile size increments was need to avoid border effect especially for the multiscale deviation and multiscale roughness.
We used the following open source software packages to compute the geomorphometry layers (Table 1 reports the software and the specific commands used for each derived variable calculation):

- Geospatial Data Abstraction Library (GDAL, `http://www.gdal.org`, version number 2.1.2) [43]
- Geographic Resources Analysis Support System software (GRASS, `https://grass.osgeo.org`, version number 7.3.0) [6, 44]
- Whitebox Geospatial Analysis Tools (Whitebox GAT, `https://whiteboxgeospatial.wordpress.com` , version number 3.3.0) [45]
- Processing Kernel for geospatial data (pktools, `http://pktools.nongnu.org`, version number 2.6.3) [46]

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[47]. They also allow for the processing of very large datasets owing to efficient algorithms and optimised memory management. After computing all geomorphometry layers within the Equi7 projection, the layers were reprojected back to the WGS84 coordinate reference system (EPSG:4326 code) with a bilinear algorithm (or near for categorical variables) implemented in GDAL. We used a tiling system identical to MERIT DEM, in terms of dimension and nomenclature, i.e. 5 degree x 5 degree tiles with 6000x6000 cells each, to allow data integration and comparisons with the original MERIT elevation dataset. All calculations were processed in parallel using open-source software at the Center for Research Computing of Yale University.

LiDAR data processing. Two common products that can be extracted from LiDAR data are: DTM and DSM. The DTM is generated using the ground echoes from the LiDAR point cloud, in conjunction with an interpolation technique[48]. The approach employed in this paper for calculating the DTMs used the LiDAR processing tools within the pktools software [46]. This is a two stage approach; the first stage uses a minimum composite rule, which retains the pulse with the minimum height for each cell.

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

This output is then filtered using a progressive morphological filter [49], 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 (retain the pulse with the maximum height for 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 projection parameters (EPSG code) of the LiDAR were set to EPSG:26911 in accordance to the metadata associated with the dataset. The DSM and DTM cell size was set to 100 m to allow an easy reprojection to Equi7, which will enable a comparison with the other layers.

Data Records

The Geomorpho90m dataset is a set of gridded layers packed as GeoTIFF files freely available for download at `https://drive.google.com/drive/folders/1D4YHUycBBhNFVVsz4ohaJI7QXV9BEh94` (high resolution: 3 arc-second and 100

m) and from Data Citation 1 (low resolution: 250 m). A visualization of the layers is given online at `https://landgis.opengeohub.org` where users can browse a subset of the 26 geomorphometry layers. Moreover, at `http://www.spatial-ecology.net/dokuwiki/doku.php?id=topovar90m` we provide high resolution overview of each variable and description of the tiling system and downloading procedure.
We provide geomorphometry layers derived from the 90 m MERIT, at global extent (60S—85N latitude) including all continents except for Antarctica, in three different products:

- at 100 m resolution, under Equi7 projection [14].
- at 3 arc-second (∼90 m) resolution, under WGS84 Geographic Coordinate System.
- at 7.5 arc-second (∼250 m) resolution, under WGS84 Geographic Coordinate System.

For the 100 m and the 3 arc-second products the relative layer names are a combination of:

variable abbreviation_resolution_DEM source layers_tiling system

The layers under the Equi7 projection are considered to store the most reliable values due to the minimal geographic distortion. For our Geomorpho90m dataset we use the T6 tiling system (and tile nomenclature) proposed by [14]. We encourage users to use the Equi7 projection especially for studies carried out at the continental or global scale.

Below are two examples of the layer names under the Equi7 projection:

- tri_100M_MERIT_AF_006_066.tif: layer showing the terrain ruggedness index at a 100 m spatial resolution in the Equi7 projection 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 Geographic Coordinate System, and hence we reprojected the layers from the Equi7 to the WGS84 Geographic Coordinate System at a 3 arc-second (∼90 m) spatial grain. Here we used the identical tiling system as implemented in the MERIT dataset (more info at `http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/`). Each single tile covers a 5x5 degree (6000x6000 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 the WGS84 projection:

- slope_90M_MERIT_s30e125.tif: layer showing the slope at a 3 arc-second spatial resolution in the WGS84 Geographic Coordinate System 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.

For the online visualisation, we reprojected the layers from the Equi7 to the WGS84 Geographic Coordinate System (EPSG:4326) 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 in the web-gis portal. The layers at this spatial resolution are also available for download at Data Citation 1. For this product the relative layer names follow the convention of the `https://landgis.opengeohub.org` which are a combination of:

layer theme_variable abbreviation_source_value type_resolution_depth interval_year_version number

Below is an example of the layer name at a coarser spatial resolution of 250 m within the WGS84 projection:

- dtm_cti_merit.dem_m_250m_s0..0cm_2018_v1.0.tif: layer showing the compound topographic index slope at a 3 arc-second spatial resolution in the WGS84 Geographic Coordinate System.

In addition, each geomorphometry layer at a spatial resolution of 3 arc-second and 100 m were stored as 32-bit floating point (Float32 data type) for maximum precision, allowing for the computation of other custom variables (e.g., coefficient of variation) or for aggregating the layers coarser spatial resolutions for macro-scale environmental models at larger spatial scales. We also provide the projection parameters and tiling system reference as reported at [42] at `https://drive.google.com/drive/folders/1D4YHUycBBhNFVVsz4ohaJI7QXV9BEh94` .

Technical Validation

The quality of the underlying DEMs, in terms of the vertical and spatial accuracy, directly influence the quality of the newly-developed geomorphometry layers. In this section we assess the sensitivity of the geomorphometry layers with respect to the DEM accuracy by:

1. Evaluating the possible artifacts that could stem from the different geographic projections.
2. Comparing the possible differences and evaluate the correlations between the different geomorphometry layers derived from different DEMs to assess their similarity.

In particular we explored the:

- possible impact of the WGS84 and Equi7 projection on the cell values of each layer.
- divergence of the most common geomorphometry layers obtained from the MERIT and 3DEP-1 DEMs,
- assess the influence of removing the tree height bias in the MERIT DEM using the DTM and DSM obtained from LiDAR

These analyses highlight the quality of the MERIT-derived geomorphometry layers and provide a possibility to identify potential errors in DEMs.

Geographic projection evaluation

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 be included in the resulting map, but the selection of a suitable projection should in principle minimise the extent and type of distortions [14]. In general, map distortions diminish as the geographic area is reduced, for instance, when moving from from global to continental or regional scales. Moreover, distortions increase as you travel 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 those carried out on a large scale.

To assess the effect of the map distortions on the output of the geomorphometry variables, we analysed the slope variation under different conditions. The slope is defined as the rate of elevation change along the direction of the water flow line, and calculated using a 3x3 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 direction on the rate of change, the cell size must have the same dimension in the x and y direction. This is not the case when the Geographic Coordinate System is used and specifically where the longitudinal dimension (x) stays constant with respect to the latitudinal dimension (y) that decreases away from the equator. To quantify the influence of y variation, we compared slope values under the effect of the WGS84 Geographic Coordinate System with the Equi7 projection for two study areas of 500x500 grid cells, underlining the same MERIT-DEM. The first area is located in the sub-tropical zone in Costa Rica. To compare the same DEM under WGS84 and Equi7 at two distinct locations (Sub-Tropical and Sub-Arctic Zone) we transpose (coordinates shift - note: no-reprojection) the Costa Rica DEM to a Sub-Arctic Zone in Greenland, under the Equi7 projection. This produce a simple displacement along south-north axes without change of the elevation pixel value (no-interpolation). The graphs presented in Figure 2 show the difference in slope calculations as a direct result of using the WGS84 Coordinate System with those from the Equi7 projection. After reproject the DEM to WGS84, the slope was calculated and then re-projected back to the Equi7 projection in order to compare the results using the scatter plots (Figure 2i corresponds to the slope correlation of the Sub-Arctic zone rather Figure 2j corresponds to the Sub-Tropical Zone). In each scatter-plot (Figure 2i,j), the red line represents the 1:1 relationship, while the black line a fitted regression model between the variables. The variations in slope calculations between the two system are minimal within the sub-tropical zone (Figure 2j) as the study area is adjacent to the equator. However, in the Sub-Arctic zone, the variations are significantly different, with the slope calculated under WGS84 underestimating the slope when compared to Equi7.

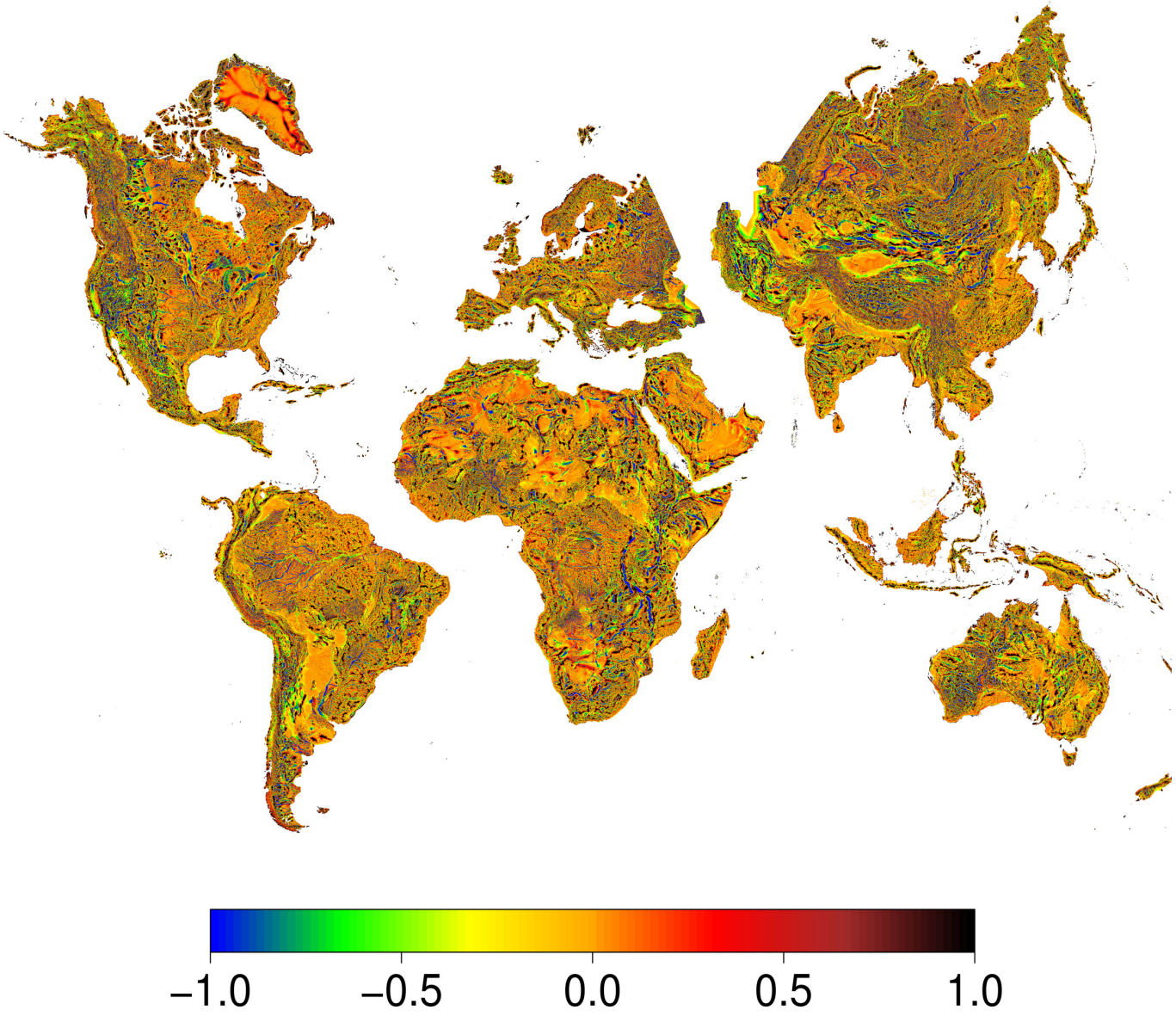


Figure 1: Quasi-global representation of maximum multiscale roughness MERIT-derived, computed under Equi7 with a pixel size of 100m and aggregate at 5km (50x50 pixels). Projection distortion of each Equi7 zone does not allow for the merging of zones for a global continuous surface. As a result, there is a gap between EU and Asia zones. However, the zones have been positioned to present a quasi-global map.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27595v1 | CC BY 4.0 Open Access | rec: 18 Mar 2019, publ: 18 Mar 2019

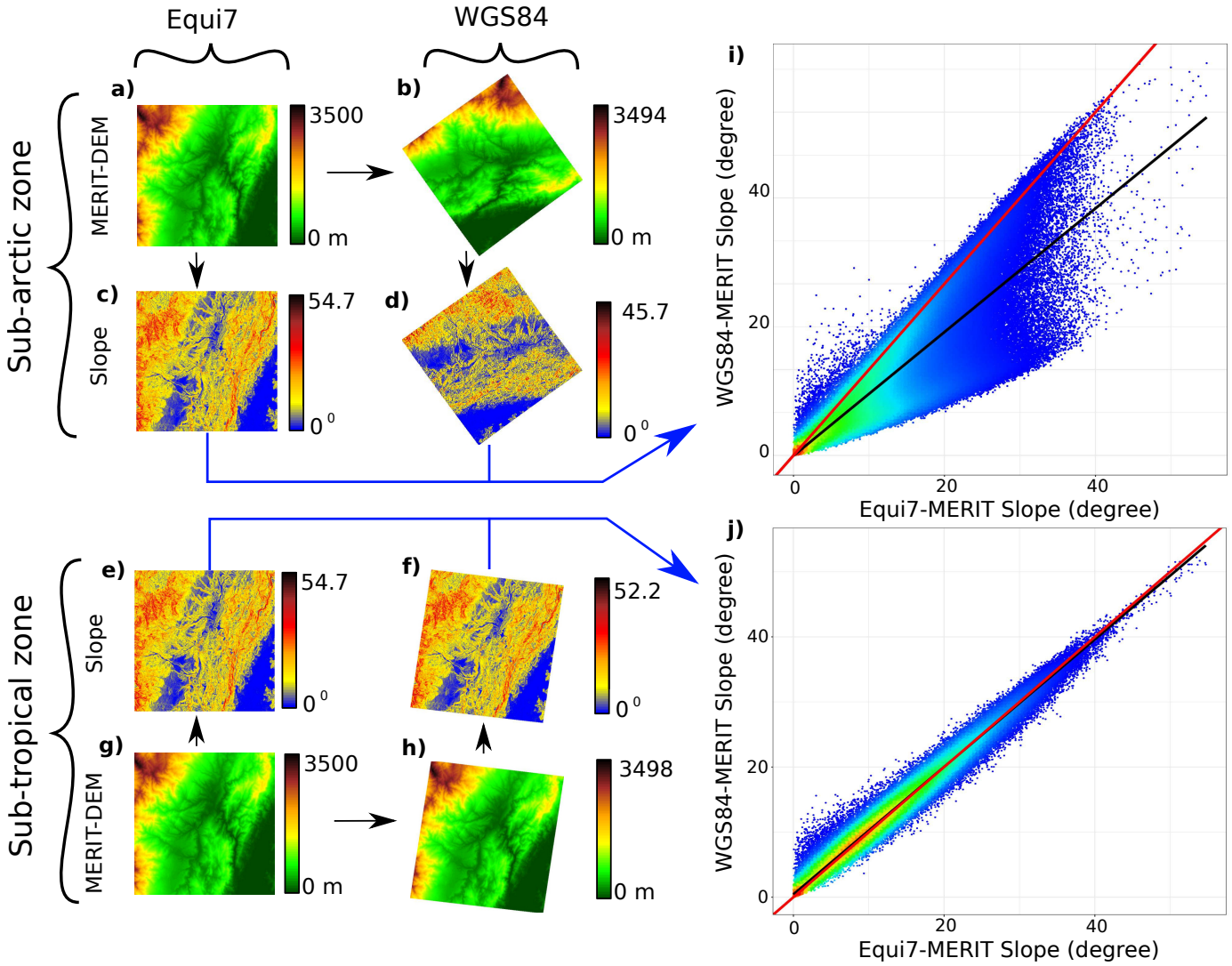


Figure 2: Graphical representation of the difference in terrain slope calculation due to the effect of World Geodetic System 1984 (WGS84) (raster panels right-hand side) compared to the Equi7 projection (raster panels left-hand side). A study area located in the Sub Tropical Zone (Costa Rica) was used to subset the MERIT-DEM for an area of 500x500 grid cells (g). This area has been transposed to a Sub Arctic Zone (Greenland) under Equi7 projection (a). After having been reprojected to WGS84 (b,h), the variable slope was calculated in the four conditions (c,d,e,f), and then reprojected back to Equi7 for comparisons (see blue line-arrows). The scatter plot on the right-hand side shows the WGS84-MERIT slope (d,f) vs Equi7-MERIT slope (c,e), for the Sub-arctic zone (i) and for the Sub-tropical zone (j). The red lines representing the 1:1 relationship and black lines representing a linear model between the two axes. The slope calculated under WGS84 in the Sub-arctic Zone is clearly underestimated with respect to the one calculated under the Equi7 projection.

MERIT versus 3DEP-1 derived geomorphometry layers

To compare the geomorphometry variables derived from the MERIT and 3DEP-1 products, and to be able to compare all of them under a common denominator; we calculated the first order derivative on the difference from the MERIT derived variables and 3DEP-1 derived variables. All of the obtained first order derivatives were measured in degrees, which allowed for comparisons. In other words, if we consider the roughness variable derived from MERIT and 3DEP-1, we can calculate the difference from which we calculated the first derivative. The first derivative is therefore the rate of change of the variables' difference and it is expressed in degrees. The first derivative of each geomorphometry variable is comparable and can be used to assess the sensitivity of the variables to the differences in DEM elevation. Figure 3 shows an overview of the first derivative of difference for each geomorphometry variables. Two elevation plots (Figure 3*a* and 3*b*) show the MERIT and 3DEP-1 DEMs and relative scatter plot (Figure 3*c*). In contrast, the other plots show the spatial variability of the geomorphometry difference, expressed in degrees. High derivative values mean high sensitivity to elevation, and conversely, low derivative values mean less sensitivity. The elevation difference (Figure 3*d*) show 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. In general the overall correlation between MERIT and 3DEP is very high, with the blue line representing a fitted regression model, which is very close to the 1:1 red line.

The roughness, terrain roughness index, the topographic index and the convergence are very sensitive to difference between DEMs, with derivative values being larger than 27 degrees (see Figure 3*f,g,h,x*). Similar behaviour can also be seen in Figure 4, where larger first order derivative mean values, and standard deviation, is reported for the convergence variable in both study areas (blue and orange lines).

To assess the sensitivity of the pattern delineation of the geomorphological forms derived from MERIT and 3DEP, we compare the geomorphological classification agreement for an area of 300x300 km (3000x3000 cell of 100 m) in South Dakota, USA. Figure 5 shows two raster plots (*a,b*) with a similar pattern at large scale but if we focus our attention to a smaller scale we can see differences in the classification at pixel level (see Figure 5*a,b* magnified circle). A common way to analyse the differences between two classifications is the calculation of a so-called confusion (or error) matrix [50]. 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 of another product. A confusion matrix can therefore be utilised to illustrate not only the degree to which the two classifications agree but also reveals how likely classes are mixed up with other ones.

In order to allow a numerical comparison of the geomorphological classifications we calculate two confusion matrices among the 10 classes in each product (see Figure 5). One is expressed as percentage of MERIT classes such that the sum within each row is equal to 100 (see Figure 5c). Considering 3DEP-1 as reference product this would give the 'user accuracy' of the MERIT geomorphometric classes. For instance, almost 70% of the 'flat' pixels in MERIT also 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 Figure 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 flat in MERIT. Additionally, the most likely confusion here is 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 of 'depression' and 'valley'. The reason for these results could be either due to an increased richness of detail (or actual resolution) offered by 3DEP-1 or a higher level of noise. In any case, this preliminary analysis shows that the underlying DEM data yield significantly different geomorphometric characteristics.

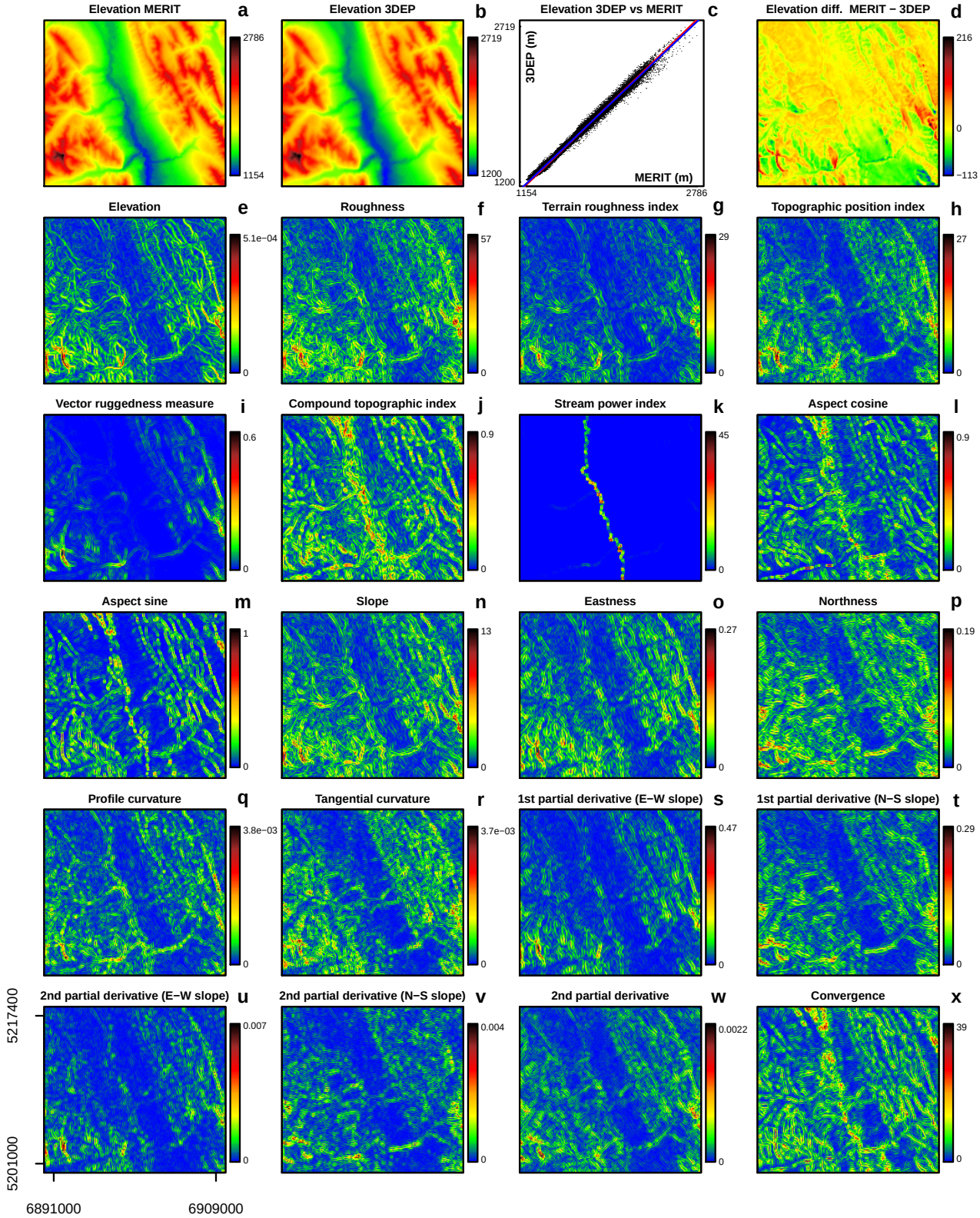


Figure 3: MERIT and 3DEP-1 DEMs absolute values and their correlation are depicted in plots *a,b,c*. The other plots show the first order derivative (i.e., rate of change through space measured in degrees) of the difference between MERIT-derived and 3DEP-derived geomorphometry variables. The colour legend shows the relative colour gradient of each single plot. The coordinates in plot *u* are expressed in metres under the Equi7 projection and they refer to a study area of 18.4 x 20 km located in Alberta, Canada, close to Jasper National Park.

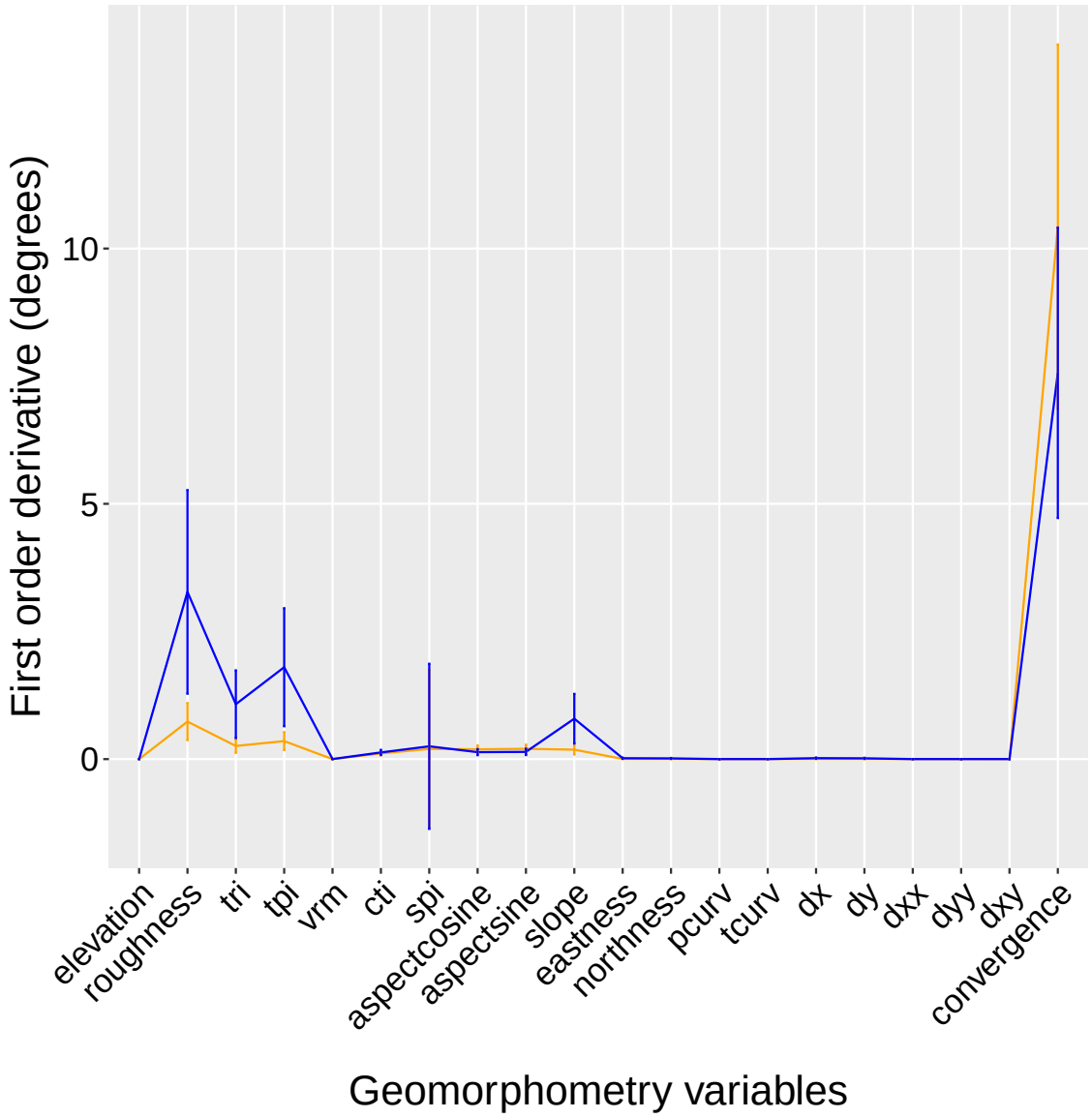


Figure 4: First order derivative mean values for each topographic variable. The first order derivative has been calculated as the difference between MERIT-derived and 3DEP-1-derived variables. The two lines represent the values for two study areas (tiles: 6000x6000 100 m cell) in USA. Vertical lines represent half of the standard deviation of first order derivatives of the geomorphometry variables.

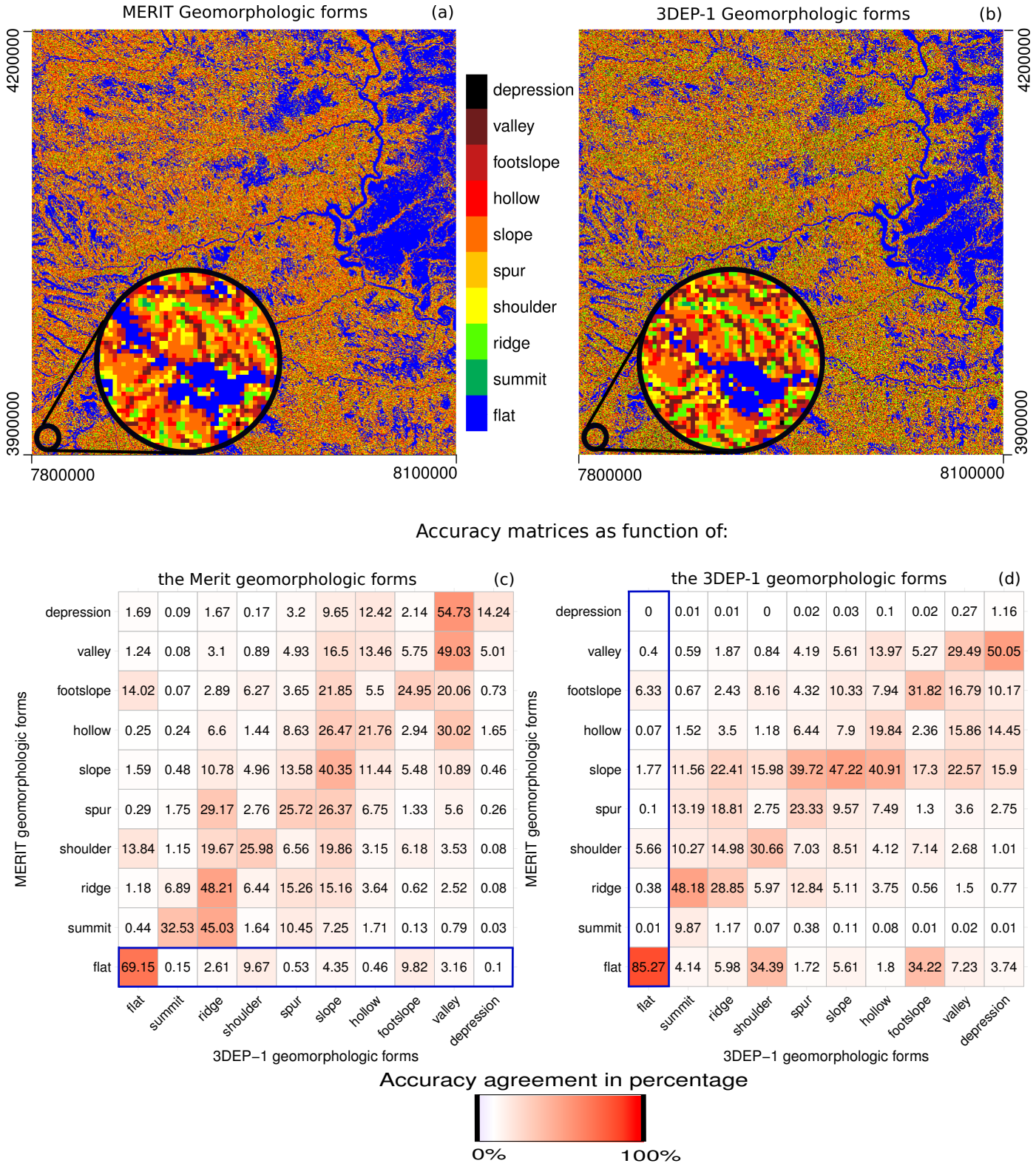
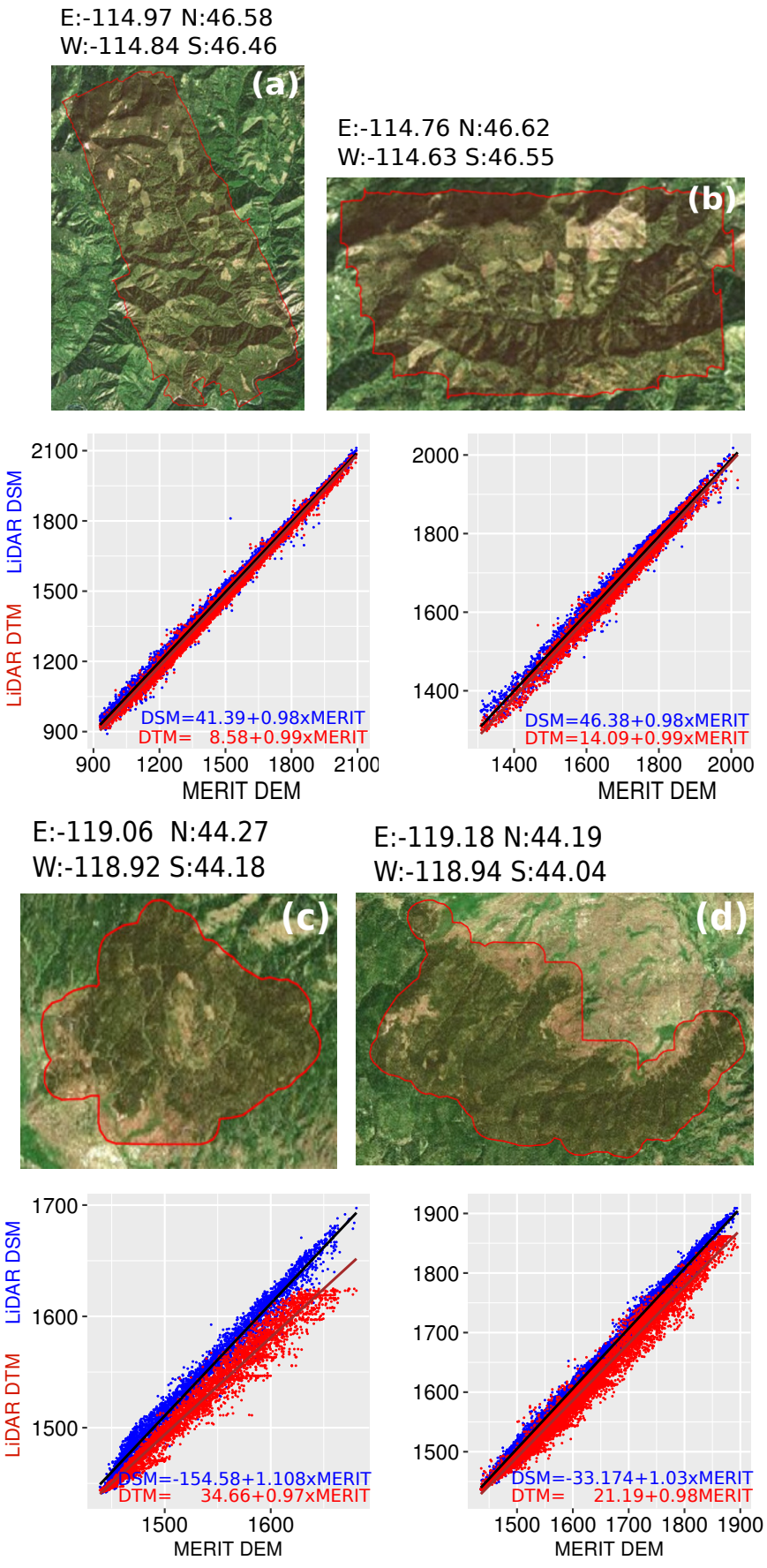


Figure 5: Geomorphology landforms and confusion matrices of a study area of 3000x3000 100 m cell in South Dakota (USA), respectively for MERIT-derived (a-c) and 3DEP-1-derived (b-d). The confusion matrix values are expressed in percentages of the MERIT classes, with the sum of vertical values equal to 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).

MERIT vs LiDAR elevation assessment

LiDAR data that penetrates dense vegetation can be used to extract the DTM, whereas the pulse returns hitting the canopy cover can be 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 (red points) and the DSM vs. MERIT (blue points) of four study areas in USA with a high density of canopy forests. In the four scatter plots it is possible to distinguish the displacement of the DTM and DSM. The MERIT dataset has been corrected for the tree height bias [10], and consequently the MERIT 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 differences are analytically quantified by the linear model depicted in the scatter plot of Figure 6. In the situation where there is no vegetation or low vegetation (e.g. agriculture area in the valley (Figure 6*c,d*) or bare ground on top of the mountain (Figure 6*a,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 model (blue and red lines; the linear model functions are reported at the bottom of each scatter plot). The coefficients for the DTM vs. MERIT are slightly below one with an intercept value ranging from 8 to 34 m. On the opposite the DSM vs. MERIT have a coefficient larger than one and an intercept ranging from -154 to 46. The mean difference between DTM and DSM for each single study area range between 12 and 21 m. These values demonstrate the strong correlation between the MERIT and the LiDAR DTM.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27595v1 | CC BY 4.0 Open Access | rec: 18 Mar 2019, publ: 18 Mar 2019

Figure 6: Comparison between MERIT-DEM and LiDAR DSM and LiDAR DTM for 4 study areas, represented with their scatter-plots and their relative linear models.

Usage Notes

The Geomorpho90m dataset (Data Citation 1) is the first global scale geomorphometry layer product at a spatial resolution of 3 arc-seconds (∼90 m) and 100 m, and has the potential to broaden the research horizon for a variety of disciplines that require detailed geomorphometric and land surface information. For instance, the new layers (Data Citation 1) 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. In particular, regarding the hydrological application the Geomorpho90m dataset in combination with other environmental layers is creating the basis for computing freshwater-specific variables following the procedure described in [51]. Most importantly, the product's global coverage ensures that a standardised set of geomorphometry layers that enable comparative analyses across continents. Additionally, the newly-developed layers (Data Citation 1) provide an update for the previous GMTED2010-derived topographic variables, as described in [4]. We encourage users that downloaded data from `http://www.earthenv.org` to use dataset described in this work. Indeed, the new Geomorpho90m dataset is more accurate in terms of its spatial resolution (90 vs. 250 m) and reduces potential residual errors from the corrections applied in the underlying MERIT dataset [10]. Therefore, the Geomorpho90m dataset (Data Citation 1) provided deeper insights into the scale-dependency of geomorphometry characteristics across a wide variety of land surface-based studies.

Acknowledgements

This study was supported in part by the facilities and staff at the Yale Center for Research Computing (YCRC). We wish to thank Steve Weston for his help and advice regarding the cluster computation. Special thanks go to Dr. Pieter Kempeneers, the developer of the pktools software, available at `http://pktools.nongnu.org`. His tools were fundamental to implement a fast processing chain. G.A. also wishes to thank Dr. Anssi Pekkarinen who introduced him to the world of GeoComputationg under the Linux OS. G.A. was supported by the Yale Institute for Biospheric Studies (YIBS).

Competing financial interests

The author(s) declare no competing financial interests.

Data Citations

1. Amatulli, G. et al. PANGAEA https://doi.pangaea.de/10.1594/PANGAEA.899135

References

[1] R. J. Pike, "Geomorphometry-diversity in quantitative surface analysis," *Progress in physical geography*, vol. 24, no. 1, pp. 1–20, 2000.

[2] A. Stein and H. Kreft, "Terminology and quantification of environmental heterogeneity in species-richness research," *Biological Reviews*, vol. 90, no. 3, pp. 815–836, 2015.

[3] C. Alexander, B. Deák, and H. Heilmeier, "Micro-topography driven vegetation patterns in open mosaic landscapes," *Ecological indicators*, vol. 60, pp. 906–920, 2016.

[4] G. Amatulli, S. Domisch, M.-N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, and W. Jetz, "A suite of global, cross-scale topographic variables for environmental and biodiversity modeling," *Scientific data*, vol. 5, p. 180040, 2018.

[5] I. D. Moore, R. Grayson, and A. Ladson, "Digital terrain modelling: a review of hydrological, geomorphological, and biological applications," *Hydrological processes*, vol. 5, no. 1, pp. 3–30, 1991.

[6] M. Neteler and H. Mitasova, *Open source GIS: a GRASS GIS approach*, vol. 689. Springer Science & Business Media, 2013.

[7] G. Amatulli, M. J. Rodrigues, M. Trombetti, and R. Lovreglio, "Assessing long-term fire risk at local scale by means of decision tree technique," *Journal of Geophysical Research: Biogeosciences*, vol. 111, no. G4, 2006.

[8] T. Grunewald, J. Stotter, J. Pomeroy, R. Dadic, I. M. Banos, J. Marturià, M. Spross, C. Hopkinson, P. Burlando, and M. Lehnig, "Statistical modelling of the snow depth distribution in open alpine terrain," 2013.

[9] A. Farahmand and A. AghaKouchak, "A satellite-based global landslide model," 2013.

[10] D. Yamazaki, D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J. C. Neal, C. C. Sampson, S. Kanae, and P. D. Bates, "A high-accuracy map of global terrain elevations," *Geophysical Research Letters*, vol. 44, no. 11, pp. 5844–5853, 2017.

[11] C. Hirt, "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*, vol. 207, pp. 27–41, 2018.

[12] V. Moudrỳ, V. Lecours, K. Gdulová, L. Gábor, L. Moudrá, J. Kropáček, and J. Wild, "On the use of global dems in ecological modelling and the accuracy of new bare-earth dems," *Ecological Modelling*, vol. 383, pp. 3–9, 2018.

[13] K. J. BEVEN and M. J. Kirkby, "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*, vol. 24, no. 1, pp. 43–69, 1979.

[14] B. Bauer-Marschallinger, D. Sabel, and W. Wagner, "Optimisation of global grids for high-resolution remote sensing data," *Computers & Geosciences*, vol. 72, pp. 84–93, 2014.

[15] T. Hengl, J. M. de Jesus, G. B. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, *et al.*, "Soilgrids250m: Global gridded soil information based on machine learning," *PLoS one*, vol. 12, no. 2, p. e0169748, 2017.

[16] M. C. Hansen, P. Potapov, R. Moore, M. Hancher, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, "High-resolution global maps of 21st-century forest cover change.," *Science*, vol. 342, pp. 850–853, 2013.

[17] M. A. Lefsky, "A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system," *Geophysical Research Letters*, vol. 37, no. 15, 2010.

[18] J. Gallant and J. Wilson, *Terrain analysis: principles and applications*. John Wiley & Sons, 2000.

[19] USGS, *The 3D Elevation Program Initiative – A Call for Action*, 2018.

[20] N. Glenn, D. R. Streutker, J. Chadwick, J. Thackray, and S. Dorschb, "Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity," *Geomorphology*, vol. 73, pp. 131–148, 2006.

[21] M. Roggero, "Airborne laser scanning: Clustering in raw data," *International Archives of Photogrammetry Remote Sensing*, vol. 34-3, pp. 227–232, 2001.

[22] S. Fassnacht, K. Dressler, and R. Bales, "Snow water equivalent interpolation for the colorado river basin from snow telemetry (snotel) data," *Water Resources Research*, vol. 39, no. 8, 2003.

[23] T. W. Crowther, H. Glick, K. Covey, C. Bettigole, D. Maynard, S. Thomas, J. Smith, G. Hintler, M. Duguid, G. Amatulli, *et al.*, "Mapping tree density at a global scale," *Nature*, vol. 525, no. 7568, p. 201, 2015.

[24] A.-J. Collados-Lara, E. Pardo-Igúzquiza, and D. Pulido-Velazquez, "Spatiotemporal estimation of snow depth using point data from snow stakes, digital terrain models, and satellite data," *Hydrological processes*, vol. 31, no. 10, pp. 1966–1982, 2017.

[25] P. Claps, M. Fiorentino, and G. Oliveto, "Informational entropy of fractal river networks," *Journal of Hydrology*, vol. 187, no. 1-2, pp. 145–156, 1996.

[26] J. D. Fridley, "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*, vol. 48, no. 5, pp. 1033–1049, 2009.

[27] M. W. Raduła, T. H. Szymura, and M. Szymura, "Topographic wetness index explains soil moisture better than bioindication with ellenberg's indicator values," *Ecological Indicators*, vol. 85, pp. 172–179, 2018.

[28] A. Román-Sánchez, T. Vanwalleghem, A. Peña, A. Laguna, and J. Giráldez, "Controls on soil carbon storage from topography and vegetation in a rocky, semi-arid landscapes," *Geoderma*, vol. 311, pp. 159–166, 2018.

[29] H. Mitasova, L. Mitas, and W. M. Brown, "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.

[30] J. B. Lindsay and D. R. Newman, "Hyper-scale analysis of surface roughness," *PeerJ Preprints*, vol. 6, p. e27110v1, 2018.

[31] S. J. Riley, "Index that quantifies topographic heterogeneity," *intermountain Journal of sciences*, vol. 5, no. 1-4, pp. 23–27, 1999.

[32] S. L. Beasom, E. P. Wiggers, and J. R. Giardino, "A technique for assessing land surface ruggedness," *The Journal of Wildlife Management*, vol. 47, no. 4, pp. 1163–1166, 1983.

[33] J. M. Sappington, K. M. Longshore, and D. B. Thompson, "Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the mojave desert," *The Journal of wildlife management*, vol. 71, no. 5, pp. 1419–1426, 2007.

[34] J. Jenness, "Topographic position index (tpi)," *Flagstaff, AZ: Jenness Enterprises*, 2006.

[35] J. Lindsay, J. Cockburn, and H. Russell, "An integral image approach to performing multi-scale topographic position analysis," *Geomorphology*, vol. 245, pp. 51–61, 2015.

[36] D. Newman, J. Lindsay, and J. Cockburn, "Evaluating metrics of local topographic position for multiscale geomorphometric analysis," *Geomorphology*, vol. 312, pp. 40–50, 2018.

[37] M. Ko, H. Kang, J. Kim, Y. Lee, and J. Hwang, "How to measure quality of affordable 3d printing: Cultivating quantitative index in the user community," in *International Conference on Human-Computer Interaction*, pp. 116–121, Springer, 2016.

[38] J. Jasiewicz and T. F. Stepinski, "Geomorphons—a pattern recognition approach to classification and mapping of landforms," *Geomorphology*, vol. 182, pp. 147–156, 2013.

[39] W. Luo and C.-C. Liu, "Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods," *Landslides*, vol. 15, no. 3, pp. 465–474, 2018.

[40] N. D. Conrad, L. Helfmann, J. Zonker, S. Winkelmann, and C. Schütte, "Human mobility and innovation spreading in ancient times: a stochastic agent-based simulation approach," *EPJ Data Science*, vol. 7, no. 1, p. 24, 2018.

[41] E. C. Underwood, A. D. Hollander, P. R. Huber, and C. Schrader-Patton, "Mapping the value of national forest landscapes for ecosystem service provision," in *Valuing Chaparral*, pp. 245–270, Springer, 2018.

[42] B. Bauer-Marschallinger, D. Sabel, and W. Wagner, "Equi7 grids projection parameters." `https://github.com/TUW-GEO/Equi7Grid/tree/master/equi7grid/grids`, 2014.

[43] F. Warmerdam, "The geospatial data abstraction library," in *Open source approaches in spatial data handling*, pp. 87–104, Springer, 2008.

[44] M. Neteler, H. Bowman, M. Landa, and M. Metz, "Grass gis: A multi-purpose open source gis," *Environmental Modelling & Software*, vol. 31, pp. 124–130, 2012.

[45] J. Lindsay, "The whitebox geospatial analysis tools project and open-access gis," in *Proceedings of the GIS Research UK 22nd Annual Conference, The University of Glasgow*, pp. 16–18, 2014.

[46] D. McInerney and P. Kempeneers, *Open Source Geospatial Tools - Applications in Earth Observation*. Springer Verlag, 2015.

[47] G. Amatulli, S. Casalegno, R. D'Annunzio, R. Haapanen, P. Kempeneers, E. Lindquist, A. Pekkarinen, A. M. Wilson, and R. Zurita-Milla, "Teaching spatiotemporal analysis and efficient data processing in open source environment," in *Proceedingsofthe3rdOpen Source GeospatialResearch& Education Symposium*, p. 13, 2014.

[48] J. Lindsay and I. Creed, "Sensitivity of digital landscapes to artifact depressions in remotely-sensed dems," *Photogrammetric Engineering Remote Sensing*, vol. 9, pp. 1029–1036, 2005.

[49] K. Zhang, S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne lidar data," *IEEE transactions on geoscience and remote sensing*, vol. 41, no. 4, pp. 872–882, 2003.

[50] S. Stehman, "Selecting and interpreting measures of thematic classification accuracy.," *Remote Sensing of Environment*, vol. 62, pp. 77–89, 1997.

[51] S. Domisch, G. Amatulli, and W. Jetz, "Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution," *Scientific data*, vol. 2, 2015.