PAPER WITH FULL DATA ATTACHED

Open Access

Global polygons for terrain classification divided into uniform slopes and basins

Junko Iwahashi[1*] and Dai Yamazaki[2]

Abstract

Global terrain classification data have been used for various issues related to topography such as the estimation of soil types and of ground vulnerability to earthquakes and the creation of seismic hazard maps. However, due to the resolution of digital elevation models (DEMs), the terrain classification data from previous studies could not discriminate small landforms such as plains at the bottom of narrow valleys and small rises in plains. Owing to the greater regional variation of small landforms, there is trade-off between DEMs of higher resolution and the creation of global geomorphological legends. To address this problem, we first merged regions with similar topographic characteristics using slope gradients and HAND (height above the nearest drainage) calculated by the 90-m-spatial-resolution DEMs interpolated from the multi-error-removed improved-terrain DEM (MERIT DEM), and united the polygons with the unit catchments of the MERIT-Basins dataset, so that the polygons contain calculated terrain measurements (slope gradient, HAND, surface texture, local convexity, sinks) and noise types as attributes, as well as the ID number of the unit catchment. In addition, we performed *k*-means clustering on the dataset using slope gradient, HAND, and surface texture, which can be combined with the dataset as a simple terrain classification. The clustering results were prepared in 15 and 40 global uniform clusters and 15 and 40 clusters for each basin to understand the global appearance of the terrain and provide zoning data for regional problem-solving. The 15 clusters were prepared to observe the outline of the terrain without any processing, whereas the 40 clusters were prepared to group and reclassify the polygons to create zoning data for each region. This dataset showed improvements in terms of capturing the small rises in plains compared to the authors' previous global terrain classification data. This dataset can be used as a proxy and is expected to contribute to modeling and estimation in various applications that are known to be related to topography. The datasets of this article are available at [https://gisstar.gsi.go.jp/terrain2021/].

Keywords: Geomorphological map, Terrain classification, MERIT DEM, Landform, DEM, Global, Geomorphometry, Vs30, Hydrology, HAND

1 Introduction

Terrestrial land comprises landforms of various ages and constituents that have undergone various slope processes such as fluvial, aeolian, glacial, and coastal, balanced by uplift and sea-level change (Summerfield 1991). Individual landforms involve many implications and vary with scale. Speight (1984) stated that units of 75–1000 m may

be considered as an appropriate size for mapping units to represent landform patterns, whereas mapping units of 15–75 m may be considered appropriate for the representation of landform elements.

Terrain classification data for landform patterns have been produced in many countries as geomorphological maps or geospatial information on topography and have been used in disaster management and regional development planning (Carrara et al. 2003; Del Monte 2016). Traditional geomorphological maps are created manually using aerial photographs and other methods to classify landforms and by considering differences in the

*Correspondence: iwahashi-j96pz@mlit.go.jp

[1] Geospatial Information Authority of Japan, 1 Kitasato, Tsukuba, Ibaraki 305-0811, Japan
Full list of author information is available at the end of the article

causes of formation, constituent materials, and time of formation through field surveys and data. However, such labor-intensive work is difficult to perform on wide areas. From around the 1990s, Dikau et al. (1991) and others proposed a method to mechanically classify terrain on a computer by using land surface features (hereinafter referred to as "terrain measurements") calculated from elevation data at grid points, known as a digital elevation model (DEM).

In practice, terrain classification data using DEMs have been used in areas such as disaster prevention, environmental assessment, and resource development to assess hazard vulnerability (Mihu-Pintilie and Nicu 2019), soil (Hengl et al. 2017), and seismic susceptibility (Kwork et al. 2018; Karimzadeh et al. 2019; Mori et al. 2020), and there is a strong need for data at the scale of landform patterns.

All of the global terrain classification maps produced since 2000 have been based on DEMs, and are based on the ideas of geomorphometry (Pike 1988; Hengl and Reuter 2008; Florinsky 2017). Furthermore, they are created by combining terrain measurements calculated from DEMs. Specifically, Meybeck et al. (2001), Iwashi and Pike (2007), Drăguț and Eisank (2012), Sayre et al. (2018), and Iwahashi et al. (2018). Meybeck et al. (2001) and Iwahashi and Pike (2007) represented large landforms using 1-km grid DEMs. Sayre et al. (2018) provided a terrain classification map product for mountainous areas. Drăguț and Eisank (2012) and Iwahashi et al. (2018) performed object-based classifications. GeoMorphons (Jasiewicz and Stepinski 2013), which is suitable for classifying the shape of mountain slopes, and TerraEX (Netzel et al. 2016), a tool that automatically searches for similar terrain and generates a similarity map, have also been developed to meet global needs.

A description of the research activities of the corresponding author's group is presented as follows. It was discovered that terrain classification data similar to man-made geomorphological maps of Japan could be created by combining terrain measurements calculated from DEMs (Iwahashi 1994; Iwahashi and Kamiya 1995). At that time, it was noticed that the combination of slope gradient and a valley density parameter (later organized as the surface texture; Iwahashi and Pike 2007) could be used to classify broad terrain types such as mountains, volcanic areas, hills, plateaus, and plains, and that the local convexity could be used to classify intermediate terrain types such as terraces and fans. From this context, two types of global terrain classification maps, cell-based and polygonal, were produced from 1-km spatial resolution Shuttle Radar Topography Mission 30 arc-second DEM (SRTM30) (Iwahashi and Pike 2007) and 280-m MERIT DEM (Iwahashi et al. 2018). The terrain

classification data given, or the creation method have been used or recommended as materials to estimate soil types (European Soil Data Centre, Joint Research Centre, European Commission 2008), to estimate Vs30 (average shear wave velocity for the top 30 m; a proxy of ground vulnerability to earthquakes) (Yong et al. 2012; Mori et al. 2020, etc.), and for seismic hazard mapping (Irsyam et al. 2017). However, due to the resolution of the DEMs used, the terrain classification data from these previous studies could not discriminate the bottom plains of narrow valleys or small rises within the plains. The use of higher-resolution DEMs was necessary to improve on this performance, but it was thought that automatic classification would be more difficult with higher-resolution DEMs because artificial steep cliffs in heavily man-made altered plains or unevenness of accuracy would become more apparent. To accommodate the increasing resolution of DEMs and to produce more practical data, Iwahashi et al. (2021) introduced HAND (Height Above the Nearest Drainage; Rennó et al. 2008; Nobre et al. 2011) as an additional parameter for the extraction of small rises in plains, as well as methods of topographic measurement in which DEM unevenness or noise is not amplified. Iwahashi et al. (2021) conducted a terrain classification for the whole of Japan using a 30-m DEM, aiming to achieve a terrain classification that reflected the vulnerability of the ground for various slopes from alluvial plains to mountainous areas, and that had both geomorphological and geotechnical classifications without major contradictions. However, there were still some issues left, including the issue of segmentation which showed a few unexpected connections of polygons of uniform slopes. In extending this achievement to the entire globe, a trade-off arises between higher-resolution DEMs and the creation of a global geomorphological legend. Although unsupervised terrain classification would be possible at any resolution, small landforms involve much more regional variation than large landforms in fitting a geomorphological legend to the classified data.

In recent years, geospatial dataset projects on geography and geology have been progressing in various fields (globally, for example, OneGeology administration 2020; Poggio et al. 2021). The creation of global datasets using satellite imagery as a resource is also underway, for example, the National Aeronautics and Space Administration (NASA)'s Socioeconomic Data and Applications Center (SEDAC) and NASA Earth Observations (NEO) have produced numerous global datasets on economical and natural geography, and a number of global datasets have been released. In the past few years, the rapid improvement in the specifications of GPUs has enabled general-purpose PCs to handle big data, which has led to rapid progress in the field of machine learning. Hence,

geospatial information is now required to be used as zoning data for machine learning, rather than simply being viewed directly like printed maps. One of the goals of geomorphometry is to use terrain measurements calculated from DEMs, i.e., physical quantities such as slope gradient, to model and estimate various issues that are known to be related to topography, such as landslide risk, the susceptibility of the ground to shaking, and soil types. Quantitatively zoned terrain data are useful for interpolating and mapping various measurements known to be correlated with topography. The approach of processing terrain classification using DEMs as one of the machine learning parameters for proxies of measured data of environmental analyses has been a theme of recent research (e.g., Keuper et al. 2020; Yamashita et al. 2022), and such needs must be addressed. On the other hand, maps that can represent an overview of the regional landforms are still needed for regions on which little geological and geomorphological information is available.

In this study, we produced global data to perform terrain classification based on the method of Iwahashi et al. (2021), which is better at classifying plains than those of previous studies (Iwahashi and Pike 2007; Iwahashi et al. 2018) and withstands the use of finer DEMs. The method was adjusted to fit the 90-m MERIT DEM.

2 Construction and content

This chapter describes the materials and methods used in the present work. The entire data construction process is shown as a flowchart in Fig. 1. The flowchart also shows where to find the explanation of each product in the subsections below.

2.1 Source data

2.1.1 DEM

The 3-arc-second grid MERIT DEM v1.0.3 (based on Yamazaki et al. 2017) was used as the main material. MERIT DEM was developed by merging several baseline DEMs, mainly the Shuttle Radar Topography Mission 3 arc-second DEM (SRTM3) ver. 2.1 provided by NASA and the ALOS World 3D (AW3D) 30-m DEM ver. 1 provided by the Japan Aerospace Exploration Agency (JAXA).

2.1.2 Region partitioning and interpolation of DEMs

The polygon data of delimitation was created and defined as regions for terrain measurements. We used the basin edges of MERIT-Basins (Lin et al. 2019) or the sea to divide the global area into 67 regions (Fig. 2). Although island regions of MERIT-Basins do not follow the watershed boundary, hereafter, we refer to the regions shown in Fig. 2 as, e.g., "Basin 41," for simplicity. This region partitioning was performed in consideration of the difficulty
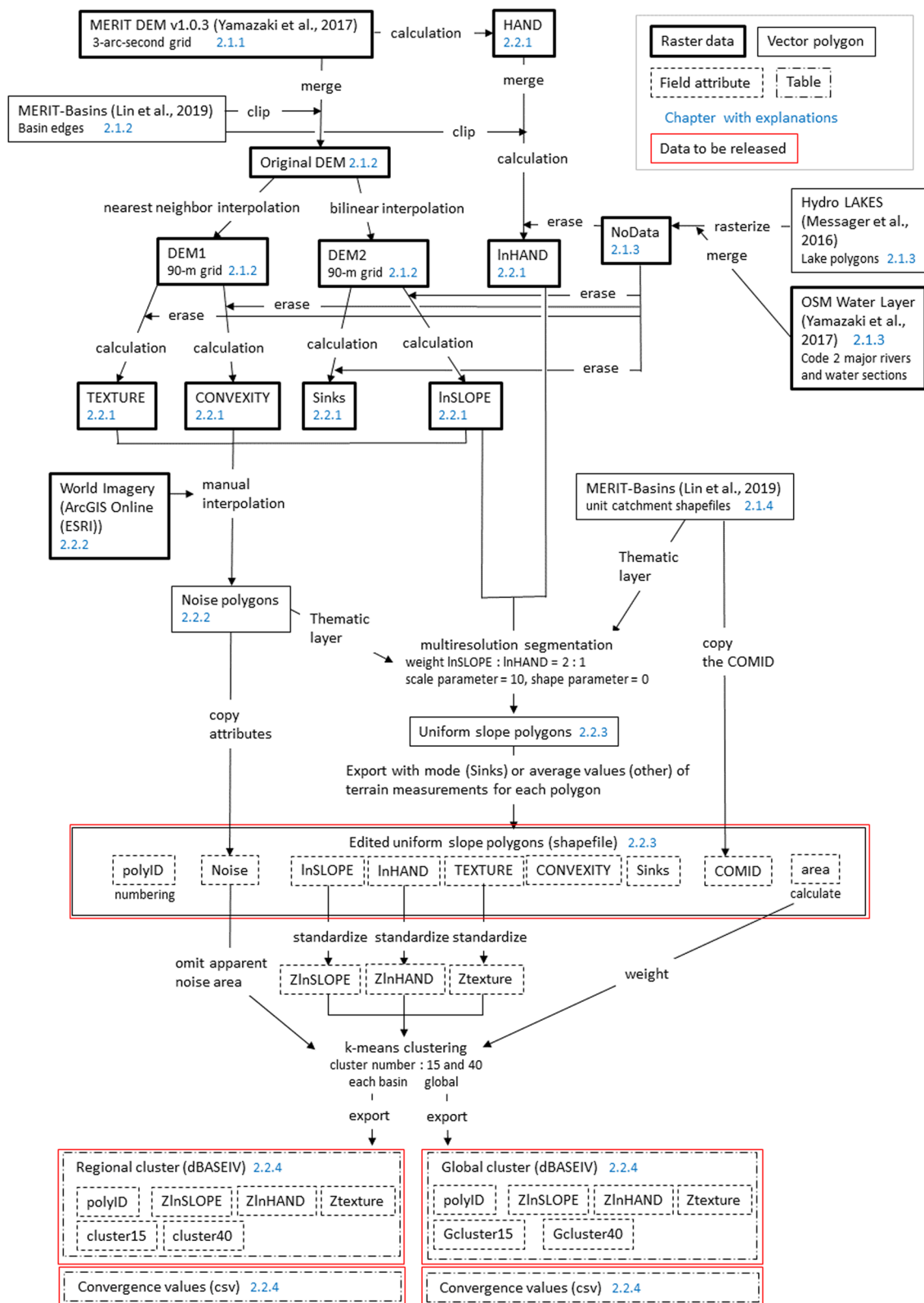
of processing the global DEM with a resolution of 3-arc-second grid data in a single calculation due to our system performance, and the usability of the resultant data was expected to be poor.

The original DEMs within each region were merged and clipped, then transformed into the Lambert Azimuthal Equal Area (WGS84) to produce a 90-m DEM with the coordinates near the center of the land area as the central meridian. Two resampling methods, the nearest neighbor method and the bilinear method, were used to create interpolated DEMs (DEM1, DEM2). We used two different resampling methods to avoid the stripe noise caused by the filtering process used in the subsequent calculation of the terrain volume using Geographic Information System (GIS) software. The noise is attributed to the raster processing method of GIS, and this problem can be avoided by using Triangulated Irregular Network (TIN) as the resampling method of DEMs. However, we did not use TIN because of the computational cost, so we used two resampling methods.
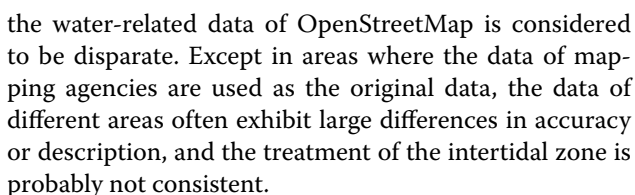
2.1.3 Masking data of NoData

To prevent underestimation of the amount of terrain measurements in areas with abrupt changes, such as the ground between water and mountains, and for product usability, it is necessary to set null (NoData) for land and water areas such as lakes, marshes, and large rivers. However, MERIT DEM does not include information for land and water areas. In a previous study (Iwahashi et al. 2018), the water part of the land cover classification of GlobCover 2009 (Bontemps et al. 2013) was used, but the resolution was coarse at about 300-m, and blank areas due to clouds were unavoidable in the independent peaks of remote islands because the data were sourced from satellite imagery. In this study, HydroLAKES (Messager et al. 2016) was used to mask lakes. HydroLAKES is a database of global polygon data of lakes larger than 10 ha. Some parts that required modification were manually modified and used, although only within Japan. HydroLAKES does not include brackish lakes or large rivers that should be NoData parts in this study, so supplementary data were necessary. Initially, we considered using land cover classification based on satellite imagery, but it was difficult to identify reclaimed land in shallow waters and islands in coral reefs. Therefore, we used the Code2 major rivers and water sections extracted from the OSM Water Layer (Yamazaki et al. 2017). The OSM Water Layer is a published dataset which was created by extracting water-related data from OpenStreetMap.

These masking data were used as NoData masks to minimize errors in areas with rapid changes in slope. These data do not necessarily reflect the actual extent of land water areas, because the criteria for acquisition of
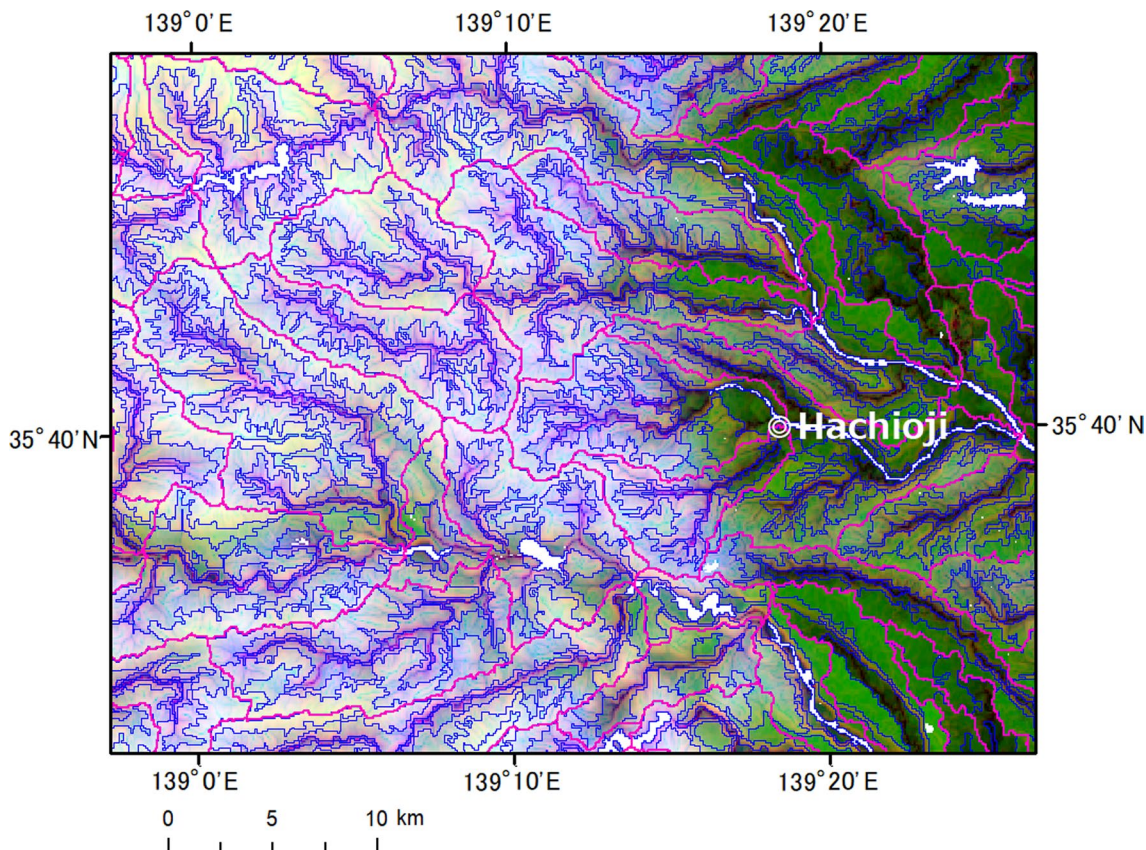
Fig. 1 Flowchart of the data construction process

Fig. 2 Index map of the region partitioning. Since the area including Australia has a large amount of data, the output data are divided into two parts using the boundaries of the unit catchments. This map also shows approximate locations of Figs. 5, 9, 10, and 14 (light blue lines)

the water-related data of OpenStreetMap is considered to be disparate. Except in areas where the data of mapping agencies are used as the original data, the data of different areas often exhibit large differences in accuracy or description, and the treatment of the intertidal zone is probably not consistent.

2.1.4 Supplementary data

MERIT-Basins' vector hydrography (based on MERIT Hydro v1.0) unit catchment shapefiles (Lin et al. 2019) was used as auxiliary thematic data for the segmentations. MERIT-Basins, which comprises unit catchment data calculated from MERIT DEM, was combined with polygon data which was segmented as uniform slopes and designed so that the uniform slopes are divided by the ridge lines of unit catchments. The unit catchments of MERIT-Basins, as shown in Fig. 3, are not detailed and are rather broad, which is an advantage for the purpose of this use. For the mountainous areas, they follow ridges well, while terraces and plains—whose geomorphologic boundaries do not follow watershed boundaries—are not sufficiently detailed to adversely affect their classification. Thus, this was performed to alleviate the problem of region segmentation in the previous studies (Iwahashi et al. 2018, 2021) and to allow users to combine the locations of upstream and downstream catchments in MERIT-Basins as polygon attributes. This process is expected to facilitate the extraction of data for

each catchment area and assessments of the connection between upstream and downstream.

2.2 Method

2.2.1 Terrain measurements

The terrain measurements used in this study basically followed the geometric signatures of Iwahashi et al. (2021). The 90-m DEMs (DEM1, DEM2) were used to calculate the terrain measurements as attributes shown in Table 1.

Following the method of Iwahashi et al. (2018), logarithmic values (ln) of the slope gradient (lnSLOPE) were calculated. This follows a common geomorphological map style in which mountainous steep slopes tend to be aggregated together, whereas plains are divided into details. We also calculated the logarithmic values (ln) of HAND to emphasize the topography of gentle slopes (lnHAND).

For the HAND calculation, Iwahashi et al. (2021) used TauDEM (Tarboton 1997, 2005). However, in this study, due to the system cost associated with the increase in data volume, we applied the HAND calculation method of MERIT Hydro (Yamazaki et al. 2019), a product that has already been hydrologically corrected, and created the data from MERIT DEM. MERIT Hydro uses a unique correction algorithm (Yamazaki et al. 2012) to perform hydrological elevation correction. This method calculates the flow direction from global elevation and water body data, and then minimizes the amount of topographic

Fig. 3 Region segmentation around the Kanto Mountains (Japan). The blue lines are the uniform slope segmentation lines by lnHAND and lnSLOPE, and the pink lines are the boundaries of unit catchments of MERIT-Basins. The color tones are R: lnSLOPE, G: lnHAND, and B: TEXTURE in this image

Table 1 Terrain attributes and their calculation methods

Attribute	Calculation method	Description
lnSLOPE	ln (SLOPE + 1) SLOPE: slope gradient (degrees) calculated from the 90 m DEM interpolated by the bilinear option	SLOPE was calculated within 3 by 3 cells windows, using QGIS 3.14, by the Horns method
lnHAND	ln (HAND + 1) HAND: Height Above the Nearest Drainage (m)	HAND was calculated by the method of Yamazaki et al. (2012) and Yamazaki et al. (2019)
TEXTURE	Density within a 10-cell radius of pits and peaks obtained by the difference between the Original DEM and the 3 × 3 median filtered DEM. Original DEM: the 90-m DEM interpolated by the nearest neighbor option	The surface texture of Iwahashi and Pike (2007). TEXTURE was calculated by the "terrain surface texture" tool of SAGA (Conrad 2012a) in QGIS 3.4 (Threshold: 5-m, radius: 10 cells, no distance weighting)
CONVEXITY	Density within a 10-cell radius of convex points obtained by processing the DEM with a 3 × 3 Laplacian filter. Original DEM: the 90 m DEM interpolated by the nearest neighbor option	The local convexity of Iwahashi and Pike (2007). CONVEXITY was calculated by the "terrain surface convexity" tool of SAGA (Conrad 2012b) in QGIS 3.4 (Threshold: 1-m, radius: 10 cells, no distance weighting)
Sinks	Sinks is the following region: ((Filled DEM)—(Original DEM)) > 0. Original DEM: the 90-m DEM interpolated by the bilinear option. Filled DEM: DEM with the surface depressions (sinks) filled in	Filled DEM was calculated by the "Fill Sinks (Wang & Liu)" (Wang and Liu 2006) tool of SAGA in QGIS 3.4

correction while maintaining the condition that the downstream elevation is not higher than the upstream elevation. Unlike HAND calculated by TauDEM as used in Iwahashi et al. (2021), these data did not undergo the fill surface depressions (sinks) process first, so it cannot extract lowlands by itself. The fill sinks tool of Wang and Liu (2006) was used separately to extract the number and areas of depressions, denoted as the variable Sinks (Table 1), and place attributes on the polygon data. The method of Wang and Liu (2006) can fill surface depressions relatively quickly, even for the large data of this study. Sinks were extracted by the difference between the filled DEM and the original DEM.

Surface texture and local convexity are very noise-sensitive parameters. Although the MERIT DEM largely mitigates the unevenness of the source DEMs, the effect of the difference of baseline data or the stripe noise in MERIT DEM became apparent, especially in high-latitude regions and deserts (Iwahashi et al. 2018). This problem also remains in the 90-m-spatial-resolution MERIT DEM. The intensity with which such heterogeneity becomes apparent varies with the resolution of DEMs. The raster images from the terrain measurements in this study showed different patterns compared to the results from Iwahashi et al. (2018) at a 280 m resolution.

In the case of calculating TEXTURE (Table 1), by setting the threshold for extracting ridges and valleys during the calculation to 5-m, we were able to eliminate the effects of unevenness except in a few areas centered on deserts, while retaining useful information such as the distinction between volcanic and non-volcanic slopes. For local convexity, the noise could not be eliminated by increasing the threshold for the extraction of convex points, and expanding the window size for density calculation was an option (Iwahashi et al. 2021). However, in such a case, the overall disadvantage was high, considering the balance between the loss of information commensurate with the 90 m resolution and the fact that local convexity is an auxiliary parameter. Therefore, CONVEXITY (Table 1) was finally determined to calculate the density with the default radius of 10 cells.

2.2.2 Detection of areas containing significant noise

The areas that contain significant noise in the DEM must be identified. This is required to allow them to be removed from the analysis in advance, to prevent them from affecting the terrain classification results. As noted above, in some areas of CONVEXITY and a few areas of TEXTURE, the unevenness of the source DEMs were reflected. Many of these "noise" regions (hereafter Noise) are along orbital stripes, and the regions which are distributed in TEXTURE alone are in deserts. In addition, noise or low values in lnSLOPE (Table 1) were also

observed in the region of deserts or in parts of ice sheets. Within the CONVEXITY, TEXTURE, and lnSLOPE images used for classification in this study, Noise polygons were created as the portion of the image that can be significantly identified by visual inspection of the maximum–minimum stretch. The ice sheet areas were also identified as Noise using lnSLOPE images and World Imagery, which is provided as a base map in ArcGIS Online (ESRI). The forms of Noise that were omitted from the terrain classification is as follows

(1) Apparent noise on CONVEXITY.
(2) Apparent noise on TEXTURE and CONVEXITY.
(3) Apparent ice sheet.
(4) Apparent noise on TEXTURE.
(5) Apparent noise on lnSLOPE.
(6) Apparent noise on lnSLOPE, TEXTURE, and CONVEXITY.

Types 1 and 2 are found in a wide range of areas. Type 3 is found in the polar regions. Types 4 to 6 are found only in desert areas.

2.2.3 Segmentation

Using the raster data of terrain measurements and supplementary thematic data, the regions were divided into uniform slopes, and polygon data were created. The segmentation method followed Iwahashi et al. (2021). The multi-resolution segmentation (Baatz and Schäpe 2000) method implemented in eCognition (Trimble) was used to segment the data under the following conditions. TEXTURE and CONVEXITY were not used to perform segmentation.

Weight lnSLOPE : lnHAND = 2 : 1,
Scale parameter = 10, Shape parameter = 0.

The unit catchment of MERIT-Basins (Lin et al. 2019) and the Noise polygons (2.2.2) were used as thematic layers. Therefore, the uniform slope polygons were divided by the catchment areas as well as by terrain features (Fig. 3). In areas where there was noise, the type is noted in the field attributes. The uniform slope polygons were output to shapefile format data with the representative values for each terrain measurement contained in the polygons. For Australia, which has a very large amount of data, this work was divided into two portions due to the capacity of the system in use. The boundaries of unit catchments of MERIT-Basins were used for the division. The field attributes were edited to add the polygon number (PolyID) and polygon area (area), as well as to copy the ID (COMID) of the unit catchment from Lin et al. (2019).

2.2.4 Regional and global k-means classification

Edited uniform polygons (Fig. 1) were used to store the representative values of topographic quantities (Table 1). Using the field attributes of these data, we created a grouping of similarly shaped slopes using the same method for the entire globe, including remote islands and continents. We did so to publish rough, standardized classification data and to help users find unique applications in their own regions. The idea behind the classification is as follows. The method should be generic and not a black box. At least the seed of classification should be unsupervised. This is necessary for two reasons: in practice, generalize training data for a specific region to a wide area with different climatic and topographic formation processes at 90 m resolution is difficult, and we aimed to find unknown features in topographic data rather than simply imitate imitating existing maps imperfectly. Specifically, we used *k*-means clustering (Mac-Queen 1967), following Iwahashi et al. (2021). Previous studies (Iwahashi et al. 2018, 2021) have confirmed that *k*-means clustering can be used to classify landforms into mountainous areas, hills, terraces, fans, and plains, as well as to subdivide them into several categories related to ground strength and geological hazards. In addition, a general method such as clustering makes it easier for users to identify and classify unique regions independently.

Each of 67 regions was classified by *k*-means clustering (MacQueen 1967). Standardized values of the field attributes, lnSLOPE, lnHAND, and TEXTURE, were used to perform the calculations, following Iwahashi et al. (2021). Only the areas in which there were no noise regions in the raster data of these three terrain measurements were used to perform clustering. The area excluded from clustering (Noise codes 0 and 1) comprised about 1.8% of the global land area. The calculations were performed using SPSS v26 (IBM), with the number of clusters set to 15 and 40, the convergence criterion set to 0, and the maximum number of iterations set to 999. Although automatically setting the number of divisions (*k*) for clustering is difficult, in this study, we set *k* to 15 for a quick overview and then to 40 so that users can group data samples according to their needs. 40 clusters may be excessively fine for a domain with thousands of cases such as the Pacific countries, but later grouping is possible. The clustering was also performed in batches for the areas where polygons were created by dividing the area (Fig. 2, Basin 56–1 and 56–2). The data with 15 clusters are intended to be used as a simple terrain classification map of the region. The data with 40 clusters are intended to be grouped by users according to local conditions; therefore, the number of clusters was set to a high value.

In the same process as the regional data, we also created a batch of clustering data for the entire globe. This is supplementary data for the observation of maps connecting adjacent Basins and for use in problems that require a unified analysis of the entire globe, such as global environmental problems.

For the global data (43,965,174 cases) and some Basins with many cases (Fig. 2, Basins 62 and 26), the calculation did not converge at the maximum number of iterations (999) for the 40 clusters. However, the maximum absolute coordinate change for any given center was small, almost zero (7.434E-5) for the 40 global clusters, and 0.001 for the 40 clusters of Amazon River basin (Fig. 2, Basin 62), which was the largest.

3 Results and discussion

3.1 Dataset to be released

The datasets of this article can be downloaded from the Geospatial Information Authority of Japan (GSI Japan) website [https://gisstar.gsi.go.jp/terrain2021/]. As explained in the previous section, the polygon data are stored in the shape file format, and the attributes include the ID number (polyID) of the segmented polygons, the average values of terrain measurements in the polygon (lnSLOPE, lnHAND, TEXTURE, CONVEXITY), the mode value of Sinks, the code number of Noise, and the COMID number of MERIT-Basins. The regional and global clusters (40, 15) are provided as separate dBASE IV (DBF) files. The convergence values are stored in Comma-Separated Value (CSV) files. Details of the dataset are as follows

(1) Properties of the edited uniform slope polygons (shapefile dataset, Poly_XX (XX: the index number)).

polyID: ID number for each polygon.
lnSLOPE (calculated as in Table 1): average value.
lnHAND (calculated as in Table 1): average value.
TEXTURE (calculated as in Table 1): average value.
CONVEXITY (calculated as in Table 1): average value.
Sinks (calculated as in Table 1): (0) The majority of the area within a polygon does not comprise sinks. (1) The majority of the area within a polygon comprises sinks.
Noise (manual interpretation; datasets with apparent noise only): (0) No apparent noise or ice sheet. (1) Apparent noise on CONVEXITY. (2) Apparent noise on TEXTURE and CONVEXITY. (3) Apparent ice sheet. (4) Apparent noise on TEXTURE. (5)

Apparent noise on lnSLOPE. (6) Apparent noise on lnSLOPE, TEXTURE, and CONVEXITY.

COMID: ID number for each unit catchment of MERIT-Basin (Lin et al. 2019). Polygons with'-99999' are the regions which MERIT-Basin does not cover.

(2) Properties of the cluster file (regional or global cluster; dBASE IV, Cluster_XX or GlobalCluster_XX).

polyID: ID number for each polygon.

ZlnSLOPE: standardized lnSLOPE. Noise areas (2) to (6) are not included.

ZlnHAND: standardized lnHAND. Noise areas (2) to (6) are not included.

Ztexture: standardized TEXTURE. Noise areas (2) to (6) are not included.

Cluster15, Cluster40: *k*-means clustering of 15 and 40 categories using standardized lnSLOPE, lnHAND and TEXTURE of the Basin with each polygon area as the weight. The cluster number (1-15 or 1-40) is not common to other Basins. A map using these data cannot be merged with different Basins. Noise areas (2) to (6) are not included.

Gcluster15, Gcluster40: *k*-means clustering of 15 and 40 categories using standardized lnSLOPE, lnHAND and TEXTURE of the global data with each polygon area as the weight. The cluster number (1-15 or 1-40) is common to other Basins. A map using these data can be merged with different Basins. Noise areas (2) to (6) are not included.

(3) Properties of the cluster convergence values (CSV, XX_convergence_values_cluster15, XX_convergence_values_cluster40, Global_convergence_values_cluster15, Global_convergence_values_cluster40).

Convergence values for each clustering of ZlnSLOPE, ZlnHAND, Ztexture.

3.2 Usage of the dataset and notes

The clustering result (dBASE IV) can be combined with the shapefile (Poly_XX) using GIS software with polyID as the key. The result of the combination is itself a simple terrain classification map. As shown in Iwahashi et al. (2021), the 40 clusters can be grouped by comparing them with existing thematic data and scatter plots of clustering convergence values to construct a terrain classification map. CONVEXITY was not used for cluster

classification in this study and is provided only as an auxiliary parameter. In regions where CONVEXITY does not contain noise, it may be used as a supplement to classify valleys and hills in intermediate slopes such as hilly terrain. However, the necessary threshold may vary from region to region.

In continental regions, Sinks (Table 1) extracted by the "Fill Sinks (Wang and Liu)" tool using the 90-m DEM cover a wide area of plains. We provide users with both Sinks and lnHAND (Table 1). The lnHAND ($=\ln (\text{HAND}+1)$) can be converted to HAND.

Because the shapefile dataset contains COMID (Lin et al. 2019) as a field attribute, it is considered possible to combine attributes of information on the location of upstream and downstream of each catchment area contained in the MERIT-Basins' river data (Lin et al., 2019) using COMID as a key. The relationship between upstream and downstream in the same unit catchment can probably be determined by the value of lnHAND.

The notes for using the dataset are as follows. The DEM reflects the topography at the time of measurement. Therefore, the classification results are not likely to be the expected ground proxies for artificially altered areas such as reclaimed land or cut-and-fill areas. Because the Noise areas were designated visually, there may be some oversight or overestimation. Terrestrial lands described in OpenStreetMap that were formed after the measurement period of the source DEM of MERIT have an elevation of 0. Therefore, they were classified into the same cluster as low plains. We changed the cluster number to 0 where we noticed the issue, but some may still exist in the dataset.

In some areas of brackish lakes and large rivers, the centerline may not be NoData, but rather narrow polygons due to the OpenStreetMap used to extract these areas. Small lakes that are not included in Hydro LAKES (Messager et al. 2016) and OSM Water Layer (Yamazaki et al. 2017) are classified in the same cluster as low plains. These narrow polygons need to be removed before use. There are often small polygons at the confluences of valleys. These polygons coincide with the blank parts of MERIT-Basins and were caused by combining the segmented polygons and MERIT-Basins. Because the areas were very small, they did not exhibit obvious influence on the clustering results.

3.3 General characteristics of the clusters
3.3.1 The global clusters
The data in this study were generated via unsupervised classification by topography from DEM only. The polygon boundaries indicate break lines where the lnSLOPE and lnHAND vary. The clustering results indicate groups of similar shape. In all Basins, slopes belonging to clusters with small HAND, TEXURE, and SLOPE form
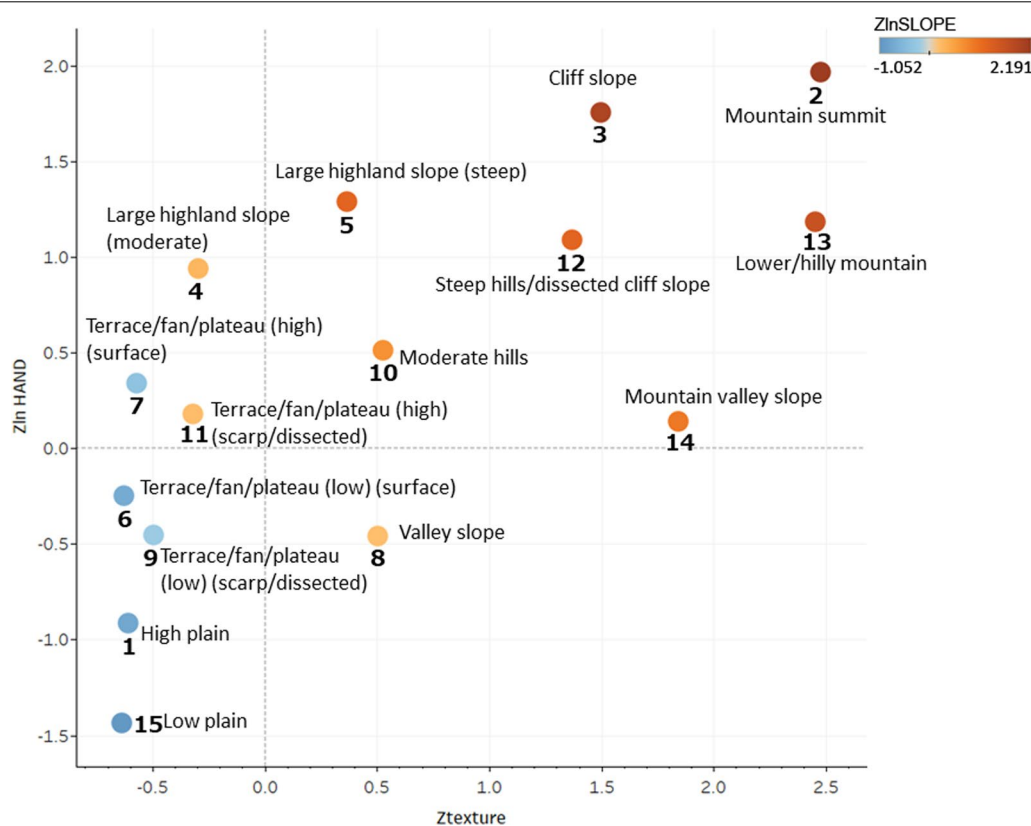
plains, while slopes belonging to clusters with large HAND and SLOPE form mountains and hills. Clusters with large TEXTURE and SLOPE but small HAND correspond to the valleys of mountains and hills, and some deeply eroded low hills. For mountains and hills, areas with relatively small SLOPE but neither summits nor valleys correspond to lower hilly mountains, hills, and dissected cliff slopes, and were considered as areas of active sediment production (Iwahashi et al. 2021). Clusters with small TEXTURE and large HAND correspond to plateaus, terraces, and sand dunes. Figure 4 shows a scatterplot of the convergence values of the 15 categories of global clusters, mainly showing ZlnHAND and Ztexture (the least collinear pair of the convergence values). Figure 4 also shows texts for interpretations on their scatter plots. Figure 5 shows color-coded maps of the 15 global clusters (Gcluster15) as polygon attributes around the Amazon River basin (a) located in a humid tropical plateau (Summerfield 1991). From the comparison with the shaded relief (c) of an enlarged view (b), an overview of the landform patterns was described. The mountains and hills were mainly divided into relatively fine textured terrains, large highland slopes, and their associated valleys (Fig. 5b). The nature of the large highland slopes showed

regional characteristics, with some regions capturing table mountains as shown in Fig. 5b, whereas others, e.g., in orogenic regions like Japan, are mainly Quaternary volcanic slopes, as will be shown below. In any case, the classification of mountain slopes would be useful for mountain assessment.
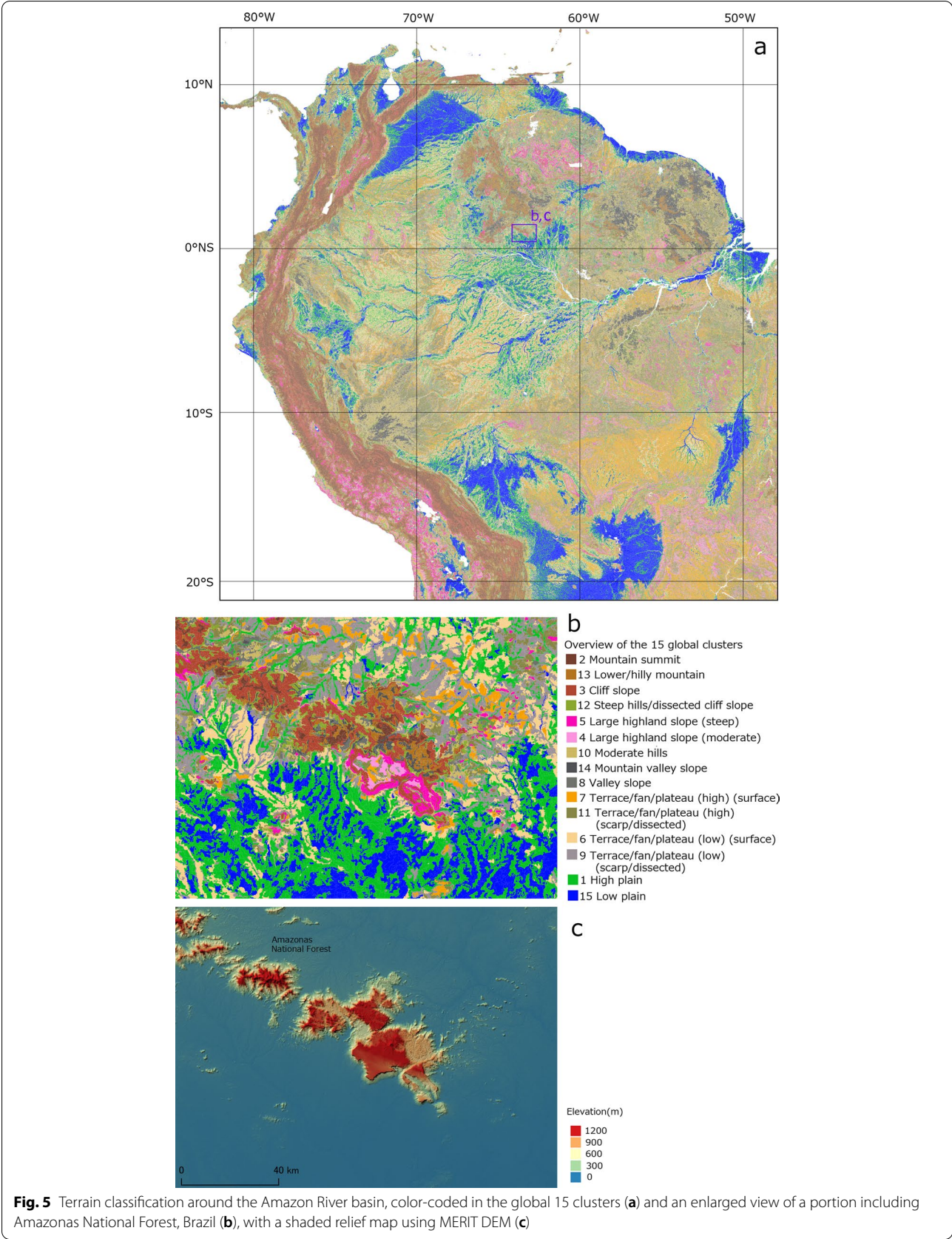
The data in this study cover not only major terrestrial areas but also remote islands, if the MERIT DEM is present. The 15 classification of global clusters is expected to prove useful for understanding the schematic topography of regions that are difficult to survey in the field, such as tropical rainforests, remote islands, arctic regions, and deserts. If users would like to obtain a more detailed categories of landforms for a wide area, the 40 categories of global clusters should be useful as material.

3.3.2 The regional clusters

It should be noted that the model with a 90 m grid resolution is appropriate to the boundary between the landform pattern and the landform element (Speight 1984). Therefore, the classification results presented in this study should capture even more landform elements than previous global terrain classification of the same type using 280-m DEMs (Iwahashi et al. 2018). This also

Fig. 4 Scatter plots of convergence values for the 15 classification global clusters and their interpretation
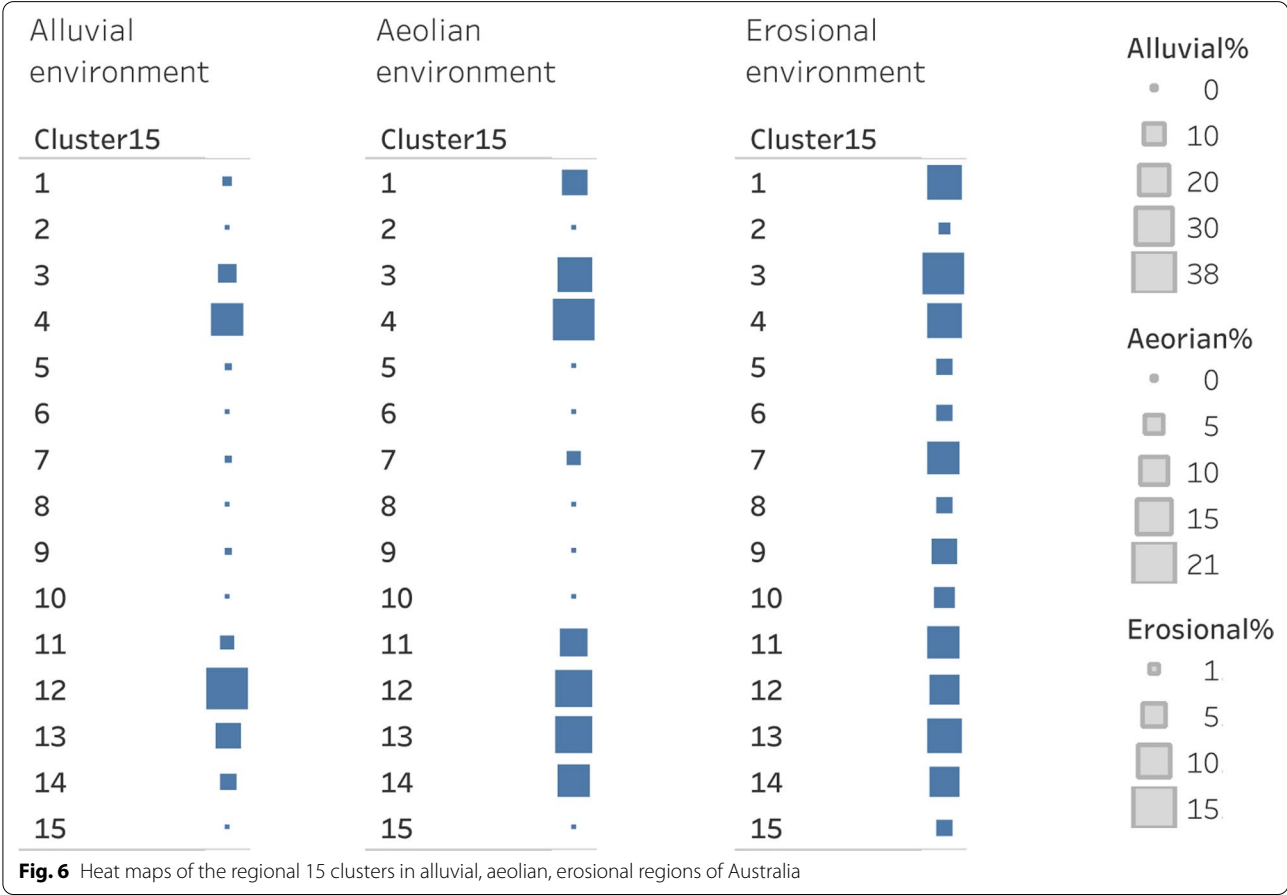
Fig. 5 Terrain classification around the Amazon River basin, color-coded in the global 15 clusters (**a**) and an enlarged view of a portion including Amazonas National Forest, Brazil (**b**), with a shaded relief map using MERIT DEM (**c**)

means that exogenic regional characteristics, such as differences in landform formation processes due to climate, are reflected more clearly. Therefore, in order to use these data to solve regional problems, regional analysis is considered necessary. This is the reason why we also provide regional clusters clustered in the Basins. In Basins where the number of cases was very large, the terrain classification (spatial pattern for each cluster) by the global and regional clusters did not different substantially. However, it is thought that more regionally specific clusters were generated.

One of the important findings for regional application is that the spatial pattern for each cluster in this study may reflect the boundaries of surficial geology or soil in some areas, but not in others. This applies equally to both the global clusters and the regional clusters. However, because the regional clusters are likely to be used more locally, we focus on these.

Figure 6 shows heat maps of the regional 15 clusters in the alluvial, aeolian, and erosional regions shown in the digital data of the Regolith Map of Australia (Craig 2013), which covers central and northeastern Australia. This region is in a largely arid environment, except along

the northern and eastern coasts. The cluster number is specific to Basin 56 (Fig. 1). Figure 7 is a scatter plot of cluster convergence values (ZlnHAND and Ztexture) of the regional 15 clusters. Figure 8 shows a dendrogram of the hierarchical clustering results using the convergence values of ZlnHAND, ZlnSLOPE, and Ztexture for the clusters. The cluster linkage method is the Ward method, and the distance is the Euclidean distance. Terrain groups grouped by Distance 2 are bracketed to the left of the cluster number in Fig. 8. Slopes of 4–12-13–14 and 1–3-11 branches are frequently distributed in Aeolian and Alluvial environments (Fig. 6). They are coarse in TEXTURE and have relatively low HAND and gentle SLOPE (Fig. 7), suggesting that they are probably poorly consolidated sediments. The slopes of branches 1–3-11 are steeper than those of 4–12-13–14, and are relatively less common on the alluvial and more common on the aeolian slopes, suggesting that they are formed of more coarse-grained sediments. Slopes of 7–9-5, 10–15, 6–8, and 2 branches are common on slopes in Erosional environments but rare in Alluvial and Aeolian, and they are plateaus, hills, and mountains.

Fig. 6 Heat maps of the regional 15 clusters in alluvial, aeolian, erosional regions of Australia
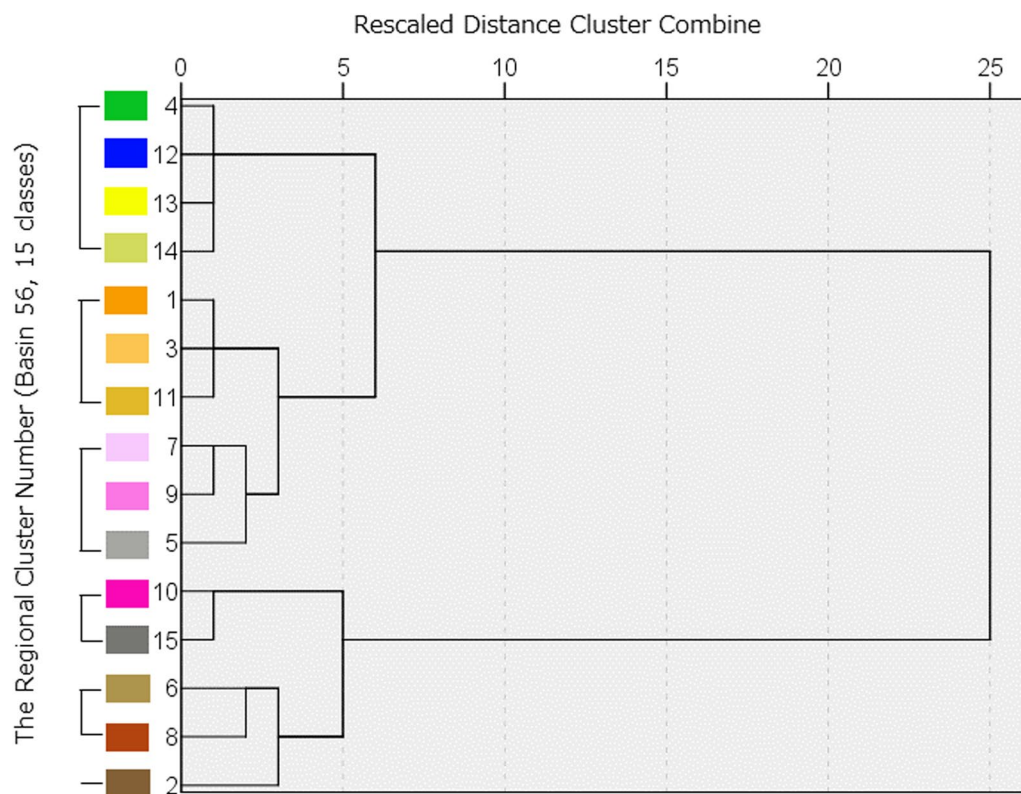
Fig. 7 Scatter plots of convergence values for the regional 15 clusters in Basin 56 (Australia)

Figure 9a shows a color-coded terrain classification map of the area near Rockhampton, which is in the Eastern Uplands Division (Pain et al. 2011) and has a relatively wet annual precipitation of approximately 1000 mm (Geoscience Australia 2022). The map is a superimposition of the 15 clusters of this study and the 1:5,000,000 Regolith Map (Craig 2013). In Fig. 9, there is a clear contrast between bedrock mountains (mainly formed by branches 10–15 and 6–8-2, including large slopes of 7 and 9 on the cliffs), regions where residual soil rests on the bedrock (mainly formed by branches 1–3-11 and 7–9), and regions of alluvial sediments (mainly formed by branches 4–12-13–14). The region of swamp sediments is not captured in the clustering results, but in the attached Sinks (Fig. 9b), polygons with predominant Sinks can be observed in similar locations. Thus, there is a possibility of meaningful reclassification by overlapping with clustering results. The distribution of Sinks is generally limited in mountainous areas such as this region, with a distribution that strongly suggests an association with low wetlands. The coastal zone sediments vary in slope type, from the 4–12-13–14 branch with Sinks, which is similar to the alluvial sediments, to the 7–9-5 and 10–15 branches, which are equivalent to the large

slope of the hills, suggesting the diversity of the coastal zone sediments.

Figure 10 shows a similar map of the region around Alice Springs, which is located between the Western Plateau Division and the Interior Lowlands Division (Pain et al. 2011) and has a dry annual precipitation of about 200 mm. Compared to the Regolith Map, the terrain classification of this study shows a similar relationship with the boundary of the erosional landform as in Fig. 9a, but less relationship with the dune fields formed by the aeolian sediments. In the dune fields, the relationship is less pronounced, and the slopes of the 4–12-13–14 and 1–3-11 branches are arranged according to the slopes of individual linear dunes. It is unclear to what extent Sinks are related to water resources due to differences in climate and geology in different regions, but when combined with areas where HAND is small (blue areas in Fig. 11), they may be useful for environmental assessment even in arid areas.

Thus, the terrain type to which each cluster corresponds differs from region to region. The use of these data to solve regional problems requires combined analysis and reclassification using regional information.

Fig. 8 Dendrogram of the hierarchical clustering using convergence values for the regional 15 clusters in Basin 56 (Australia)

It is probably necessary to separate the interpretation of clusters by rough physiography and climate zones (e.g., morphoclimatic zones; Summerfield 1991). Regions with abundant data can be interpreted and reclassified considering existing soil and geological information, while hills and mountains, where existing data are generally scarce, can be grouped using scatter plots of convergence values and hierarchical clustering results.

3.4 Small rises in the plains and Vs30
The expected benefit of higher-resolution DEMs and the use of HAND is the improved capture of small rises in the plains, such as terraces and sandy rises, which have different soil types from floodplains and coastal plains. In regions where the clusters reflect characteristics of surficial geology and soils, this benefit may be reasonable. However, quantitatively evaluating existing thematic maps is difficult as the correct answer for unsupervised clustering results such as those in this study. This is because the clustering results using DEMs are unsupervised classifications based on topography alone and they are based on a different concept from manual classification maps that consider geology and causation. The difference is even more clear in urban areas, where artificial changes have progressed and information on
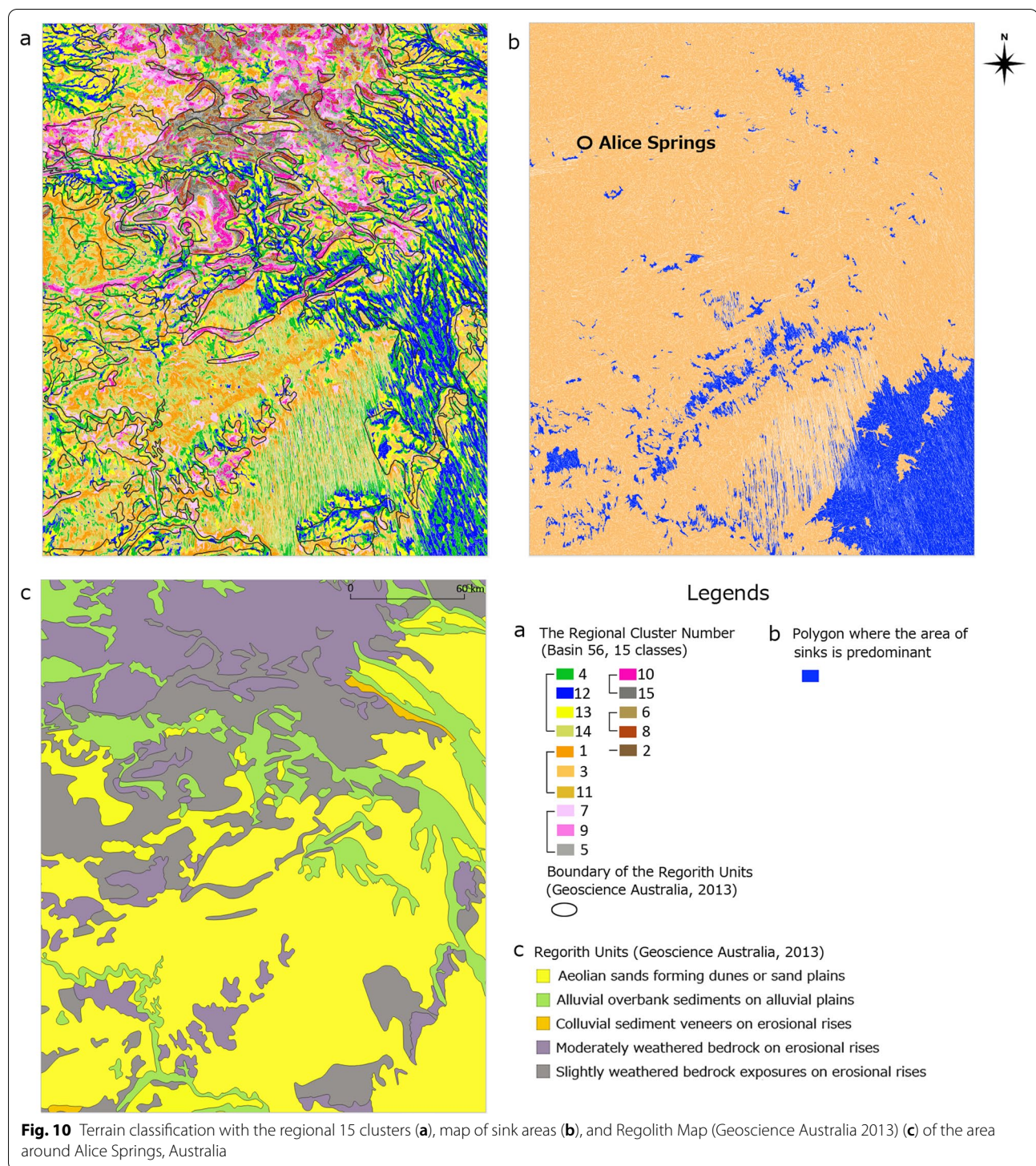
micro-elevations is often lost (but is still included in thematic maps) in DEMs. Despite these limitations, it is possible to measure the degree of improvement from the previous study by quantitatively examining "how well the target terrain species are separated as a cluster."

We compared the result of this study with those of Iwahashi et al. (2018), a global classification map using a 280-m MERIT DEM, for the Kanto region (Tokyo metropolitan area), which includes the world's largest metropolitan area in terms of population within the region and is mostly composed of alluvial plains and terraces. In addition, we compared the result with the terrain classification data of Japan (Iwahashi et al. 2021), which was classified using a Japanese 30-m DEM generated from accurate elevation data including measurements performed by airborne light detection and ranging (LiDAR) systems. Because terrain classification data have applications for estimating the shaking susceptibility proxy (Vs30), the results of this study are compared with Vs30 data (Senna et al. 2013, 2019) for the Kanto Plain to confirm their characteristics.

The plains of the Kanto region, which is the metropolitan area of Japan, are surrounded by mountains and volcanos, and there are wide terraces, silty and clay floodplains, and coastal lowlands dotted with natural levees
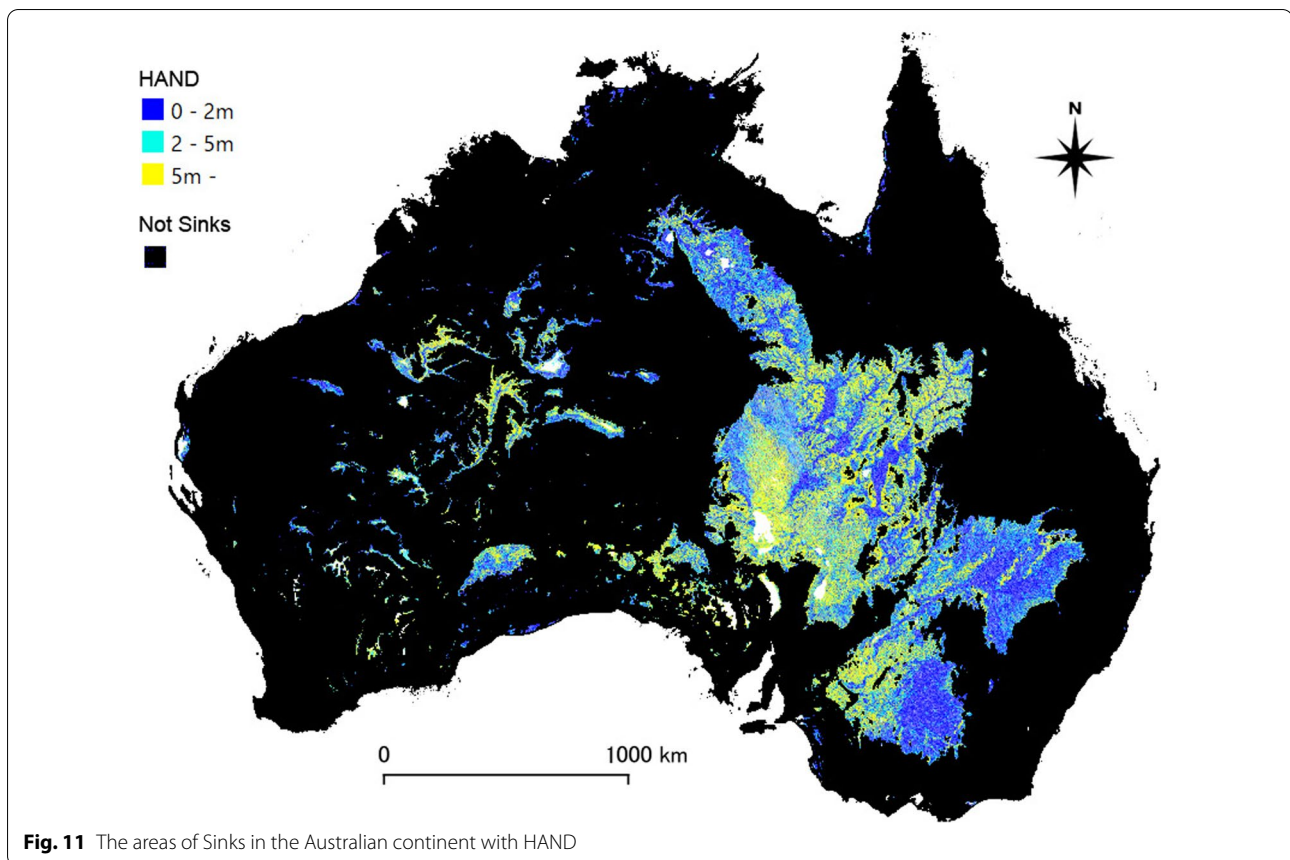
Fig. 9 Terrain classification with the regional 15 clusters (**a**), map of sink areas (**b**), and Regolith Map (Geoscience Australia 2013) (**c**) of the area around Rockhampton, Australia

Fig. 10 Terrain classification with the regional 15 clusters (**a**), map of sink areas (**b**), and Regolith Map (Geoscience Australia 2013) (**c**) of the area around Alice Springs, Australia

and elevated sand bar deposits deposited by sandy soil from past floods. There are also artificial embankments and sand bar deposits scattered throughout the floodplains and coastal lowlands. In the Kanto Plain, 1:25,000 to 1:50,000 scale geomorphological maps based on soil sampler data and aerial photographic interpretation have been produced since the 1960s, including the 1:50,000 Fundamental Land Classification Survey by National Land Agency, and the 1:25,000, Land Condition Map and Landform Classification Map for Flood Control by

Fig. 11 The areas of Sinks in the Australian continent with HAND

GSI (available on the GSI Maps; Geospatial Information Authority of Japan 2022), and many geological and geographic surveys have been conducted. Wakamatsu and Matsuoka (2013) compiled a large amount of map information and created an engineering geomorphologic classification map of Japan (here after JEGM), which is a 7.5 by 11.25 arc-second grid map of Japan. In this paper, the 2020 version (V4) included in the Japan Seismic Hazard Information Station (J-SHIS) Map (National Research Institute for Earth Science and Disaster Resilience (NIED) 2020) is used as a reference.

The grouping of the regional 40 clusters of Basin 41 (including Japan) is shown in Figs. 12, 13. In this case, we did not cross-tabulate the results with existing geological maps, but simply classified them based on the hierarchical clustering results in Fig. 12. The interpretation of the branches in the Kanto region, which was inferred from JEGM and the seamless digital geological map of Japan (Geological Survey of Japan, AIST 2019), and the color of the legend were added. In addition, the clusters with more than 10% Sinks (three clusters in the plains) were reclassified. This simple grouping using hierarchical clustering of cluster convergence values should be possible for other, less well-documented basins. Figure 13 shows

the grouping result on a scatterplot, and Fig. 14 shows JEGM (a) and the three terrain classifications using DEMs for the Kanto region. The terrain group which was created based on the regional clusters in this study (Fig. 14c) represents the major divisions of mountains, volcanos, hills, terraces, and plains well.

The terraces, i.e., the areas corresponding to the three legends of rocky strath terrace (rarely distributed in the Kanto area), gravelly terrace, and terrace covered with volcanic ash soil in the JEGM, are included in Legends 9, 10, and 11 in this study, Legends 6 and 8 in Iwahashi et al. (2018), and Legends 9, 10, 11a, and 11b in Iwahashi et al. (2021). The areas corresponding to natural levees, and marine sand and gravel bars in the JEGM are considered to be included in Legends 12 and 16 in this study and Legend 13 in Iwahashi et al. (2021). In Iwahashi et al. (2018), they are integrated with the floodplain/coastal plain and cannot be contrasted.

It is difficult to distinguish between the eroded part of a terrace (e.g., Legend 10 in Iwahashi et al. (2021)) vs. hills, low terraces and dissected fans, because the DEM terrain classification only classifies slopes with similar topography. In the global data (Fig. 14bc), the high fill along the coast is not classified as NoData, and therefore
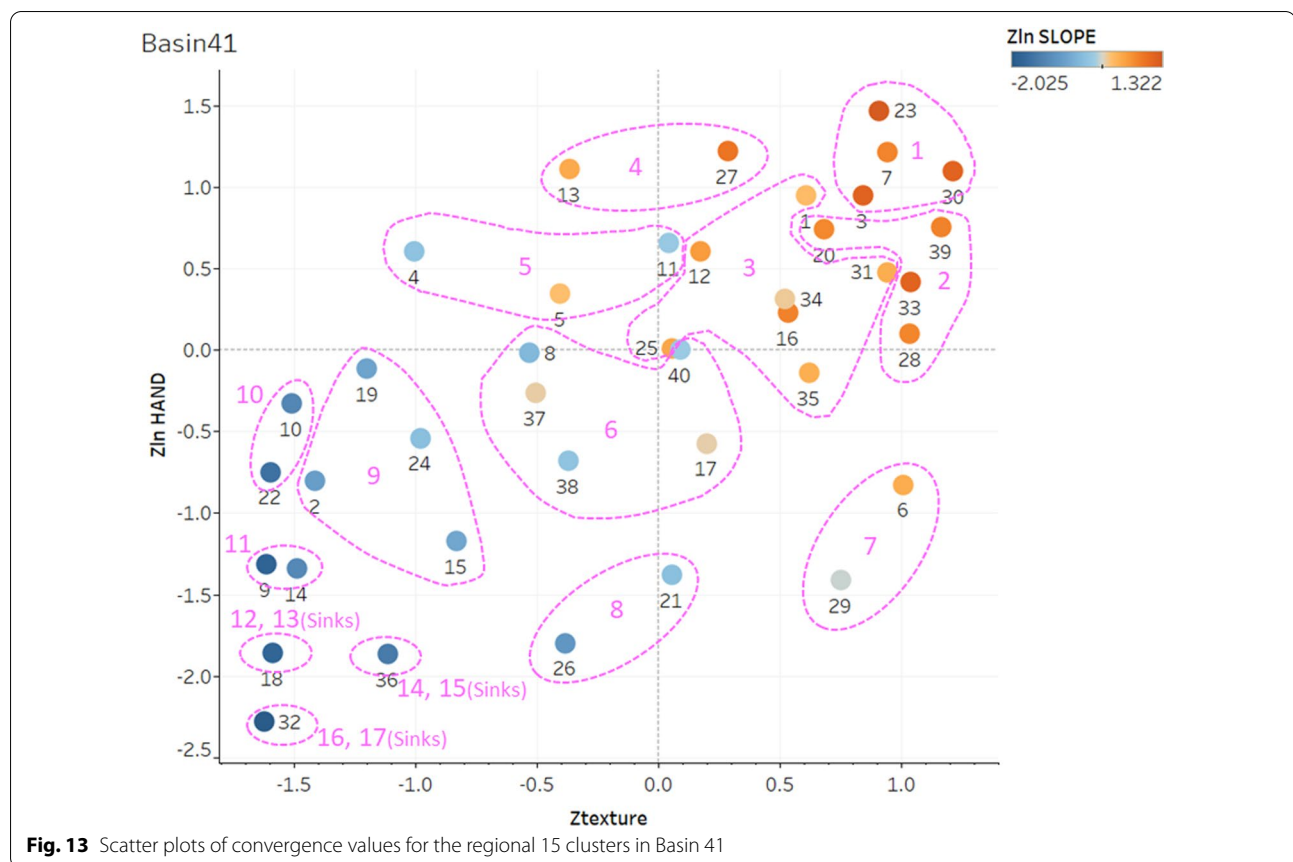
Fig. 12 Dendrogram of the hierarchical clustering using convergence values for the regional 40 clusters in Basin 41

is classified in the same category as terraces in the figure. At least in Japan, low plains such as floodplains and coastal plains have undergone urbanization and paddy field development and have been totally artificially modified for residential construction and flood control. And in many cases, micro-elevations have already been lost from the DEM. On the contrary, in some cases, the lowlands have been added with the small rises of fill due to land consolidation and housing construction.

Table 2 shows the precision, recall, and F-measure of the legends of the terraces, natural levees, and marine sand and gravel bars in JEGM and the legends of the

Fig. 13 Scatter plots of convergence values for the regional 15 clusters in Basin 41
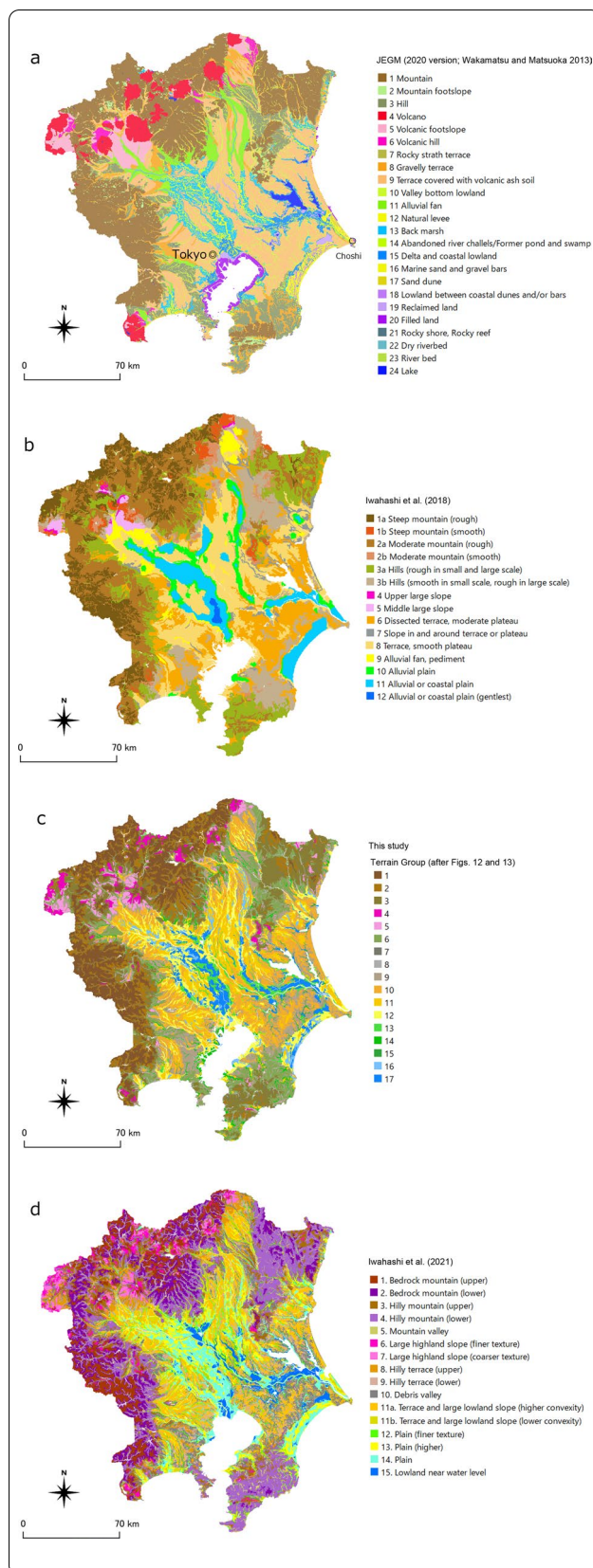
maps that can be compared. The legend of the maximum area in the grid of JEGM was checked and calculated. High-fill areas along the coast were excluded from the calculation. The baseline value was calculated by randomly deriving the same number of positives as grids in each category of JEGM by SPSS v26 (IBM). The baseline of F-value for terraces was 0.27. The baseline of F-value for natural levees and marine sand and gravel bars was 0.05 because the data exhibited a large bias in the number of positives and negatives.

As may be observed in Table 2, the capture of terraces in this study was significantly improved over that of Iwahashi et al. (2018), which used a MERIT DEM coarser by a factor of three, and did not use HAND. In contrast, the F-measure was not much different from that of Iwahashi et al. (2021), which used an accurate DEM that was three times finer and included LiDAR, even though the accuracy of the DEM was significantly different. This may be because this study used unit catchment polygons for the segmentations and there were differences in the calculation method of HAND. Recall was high in this study and Iwahashi et al. (2021), which means that the use of HAND reduces the oversight of terraces. In contrast, a phenomenon of confusion between sand dunes (for

example, along the northwest coast of Choshi) and terraces was evident. However, the overall distribution of the terraces is clear, and the data of this study can be considered to provide practical results for the terraces.

As for natural levees and marine sand and gravel bars, they are integrated with the floodplains and coastal plains in Iwahashi et al. (2018) and not detected at all. In this study and in Iwahashi et al. (2021), they were partially detected, indicating an improvement due to the use of HAND. Relatively clear natural levees in the upper reaches and marine sand and gravel bars were detected to some extent. However, the capture status of natural levees in downstream urban centers where man-made alterations are predominant was not good. Although improvements have been made from scratch, additional data on local conditions are considered necessary for urban areas.

Figure 15 shows a comparison with Vs30 data (in J-SHIS Map, NIED 2020), which is a site amplification factor for seismic shaking in the Kanto region. Vs30 data are derived from borehole core data by spatial interpolation using geological and geomorphological divisions as proxies (Senna et al. 2013, 2019). In the Kanto region, high-density miniature microtremor array surveys were

conducted in terraces and plains, and accurate data were produced by correction using a deep ground model. A box-and-whisker diagram with averaged Vs30 (AVS30) was created using the terrain type with the largest area as the representative type for each grid (Fig. 15). The points corresponding to high-fill areas along the coast were excluded. In the box-and-whisker diagram, the range of inter-quantile range (IQR) for mountains and hills is similar for Groups 1, 2, and 7, Groups 3 and 4, and Groups 5, 6, and 8, respectively. Therefore, from the comparison with Fig. 13, it appears that Vs30 in mountainous or hilly areas is not related to HAND, i.e., height from erosion reference surface, but to TEXTURE or SLOPE. However, this interpretation should be considered with caution because borehole core investigations in mountains are usually conducted near the valley floor. From the comparison with Fig. 12, among mountains, there is a distinct difference in the IQR of Vs30 between bedrock mountains (Groups 1, 2) and fragile mountains (Groups 3, 4, 5). In the Kanto region, large highland slopes (Groups 4, 5) in the fragile mountains are clearly concerned with mafic Quaternary volcanos (Geological Survey of Japan, AIST 2019). For terraces and plains, the higher the HAND, the higher the AVS30 in most classes, with the exception of Group 14 (hilly valleys). We will wait for further research to determine whether this trend may be observed in other parts of the world.

4 Conclusions

In this study, the combination of terrain measurements calculated from the 90-m MERIT DEM interpolated from the 3-arc-second grid MERIT DEM v1.0.3 (based on Yamazaki et al. 2017) and MERIT-Basins (Lin et al. 2019) have been used to perform segmentations, and shapefiles containing attributes for each unit slope were created and published from [https://gisstar.gsi.go.jp/terrain2021/]. The shapefiles were created for each of the 67 Basins, and the attributes were polygon ID, lnSLOPE, lnHAND, TEXTURE, CONVEXITY, Sinks, Noise, and COMID of MERIT-Basins. Data discretized by *k*-means clustering using several terrain measurements was also presented. The clustering results can be linked to the shapefiles by polygon ID, and two types were provided, including one for the entire globe, and another for each Basin.

The 15 clusters can be used for observing an overview of topography such as mountainous areas, plateaus,

Table 2 Precision, recall, and F-measure of the legends of the terraces (a), natural levees and marine sand and gravel bars (b) in JEGM and the legends of the terrain classification maps that can be compared

(a)

This study (Fig. 14c)		JEGM		JEGM			precision=	0.63
		0		1			recall=	0.77
other	0	TN=	303257	FN=	31410		F-measure=	0.69
Legends 9,10,11	1	FP=	59950	TP=	103928			

Iwahashi2021		JEGM		JEGM			precision=	0.68
		0		1			recall=	0.75
other	0	TN=	316370	FN=	34439		F-measure=	0.71
Legends 9,10,11ab	1	FP=	46837	TP=	100899			

Iwahashi2018		JEGM		JEGM			precision=	0.57
		0		1			recall=	0.52
other	0	TN=	310587	FN=	65540		F-measure=	0.54
Legends 6,8	1	FP=	52620	TP=	69798			

BL(random)		JEGM		JEGM			precision=	0.27
		0		1			recall=	0.27
BL	0	TN=	264891	FN=	98316		F-measure=	0.27
BL	1	FP=	98316	TP=	37022			

(b)

This study (Fig. 14c)		JEGM		JEGM			precision=	0.23
		0		1			recall=	0.08
other	0	TN=	327840	FN=	125025		F-measure=	0.11
Legends 12,16	1	FP=	35367	TP=	10313			

Iwahashi2021		JEGM		JEGM			precision=	0.29
		0		1			recall=	0.07
other	0	TN=	339435	FN=	125716		F-measure=	0.11
Legend 13	1	FP=	23772	TP=	9622			

BL(random)		JEGM		JEGM			precision=	0.05
		0		1			recall=	0.05
BL	0	TN=	451332	FN=	23008		F-measure=	0.05
BL	1	FP=	23008	TP=	1197			

Fig. 15 Box-and-whisker diagram of Vs30 for each of the terrain groups in this study (after Figs. 12 and 13) in the Kanto district, Japan (**a**). The Vs30 map (GIS data iare according to JSHIS v4: Senna et al. 2013, 2019) (**b**)

terraces, plains. The 15 classifications of global clusters will be useful for understanding the schematic topography of regions that are difficult to survey in the field. The 40 clusters can be grouped by statistics, such as the grouping with reference to the hierarchical clustering results of the convergence values of the clusters and the reclassification using Sinks (Fig. 14c), or machine learning for themes related to topography according to the needs of each region to construct meaningful maps. Based on the observations in Australia, the relationship between slope materials and the cluster is clear in some

areas and less so in others, so the usage is expected to vary from region to region.

In the experiment conducted for the Kanto region of Japan, the terraces in the map constructed from the 40 regional clusters were picked up clearly compared to the reference map, and the data of this study can be considered to provide practical results for the terraces. We were able to obtain the relationship between the Vs30 data and each terrain group on the map of the Kanto region, which shows a clear difference between bedrock mountains and fragile mountains described by the clusters. Of course,

in addition to using the clustering results, users can also make their own classification using the raw terrain measurements stored in the polygons.

The main point to be noted is that the results presented in this study only represent the topography at the time of measurement of the MERIT DEM source data, and any topography that has been rapidly changed since then, for example, by artificial or volcanic activities, is different from the current situation. When using the clustering results as a proxy for slope materials, it is important to note that the results are not as expected in areas that have been cut and filled by human modification. For example, in an analysis conducted in the Kanto region in Japan, natural levees and marine sand and gravel bars, which could not be captured at all in a previous study (Iwahashi et al. 2018), were captured reasonably well in areas with no artificial modifications, but notably less so in heavily modified urban areas. Additional data on local conditions are considered necessary for urban areas.

Recently, a free DEM with a 30 m resolution has been published (Hawker 2022), and the resolution of global-level DEMs will continue to increase in detail. However, due to scale issues, it is well known that DEM resolution affects the values and spatial patterns of terrain measurements (Zhang and Montgomery 1994; Deng et al. 2007; Mulder et al. 2011), and the relationship between scale, appropriate terrain measurement type, and calculation method is complex. Higher-resolution global data, if it is to go beyond unsupervised classification and to produce a robust geomorphological legend, will most likely need to be undertaken not by individuals, but by international teams with knowledge of geomorphological development in a variety of climatic and physiographic settings.

Abbreviations

AW3D: ALOS World 3D; CSV: Comma-Separated Values; DBF: DBASE; DEM: Digital elevation model; GIS: Geographic information system; GSI Japan: Geospatial Information Authority of Japan; HAND: Height above the nearest drainage; IQR: Interquantile Range; JAXA: Japan Aerospace Exploration Agency; JEGM: Japan engineering geomorphologic classification map; J-SHIS: Japan Seismic Hazard Information Station; LiDAR: Airborne light detection and ranging; MERIT: Multi-error-removed improved-terrain; NASA: National Aeronautics and Space Administration; NIED: National Research Institute for Earth Science and Disaster Resilience; OSM: OpenStreetMap; SEDAC: Socioeconomic Data and Applications Center; SRTM: Shuttle Radar Topography Mission; TIN: Triangulated Irregular Network; Vs30: Average shear wave velocity for the top 30 m.

Authors' information

JI is a member of the Geographic Information Analysis Research Division at the Geography and Crustal Dynamics Research Center of GSI Japan (https://www.gsi.go.jp/cais/geoinfo-index-e.html). DY is an associate professor at the

Institute of Industrial Science, University of Tokyo (http://hydro.iis.u-tokyo.ac.jp/~yamadai/).

Availability of data and materials

The datasets supporting the conclusions of this article are available in the GSI Japan website, [https://gisstar.gsi.go.jp/terrain2021/].

Declarations

Competing interests

The authors declare that they have no competing interest.

Author details

[1]Geospatial Information Authority of Japan, 1 Kitasato, Tsukuba, Ibaraki 305-0811, Japan. [2]Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.

References

Baatz M, Schäpe A (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Proceedings of Angewandte Geographische Informationsverarbeitung, vol XII, pp 12–23

Bontemps S, Defourny P, Radoux J, Van Bogaert E, Lamarche C, Achard F, Mayaux P, Boettcher M, Brockmann C, Kirches G, Zülkhe M, Kalogirou V, Arino O (2013) Consistent global land cover maps for climate modeling communities: current achievements of the ESA's land cover CCI. In: ESA living planet symposium 9-13 September 2013, Edinburgh, United Kingdom

Carrara A, Crosta G, Frattini P (2003) Geomorphological and historical data in assessing landslide hazard. Earth Surf Process Landforms 28:1125–1142. https://doi.org/10.1002/esp.545

Conrad O (2012a) Module terrain surface texture. In: SAGA-GIS Module Library Documentation (v2.2.5). http://www.saga-gis.org/saga_tool_doc/2.2.5/ta_morphometry_20.html. Accessed 26 Feb 2022

Conrad O (2012b) Module terrain surface convexity. In: SAGA-GIS Module Library Documentation (v2.2.5). http://www.saga-gis.org/saga_tool_doc/2.2.5/ta_morphometry_20.html. Accessed 26 Feb 2022

Craig MA (2013) Regolith map of Australia Edition 1-digital. Geoscience Australia. http://pid.geoscience.gov.au/dataset/ga/76662. Accessed 26 Feb 2022

Del Monte M, D'Orefice M, Luberti GM, Marini R, Pica A, Vergari F (2016) Geomorphological classification of urban landscapes: the case study of Rome (Italy). Journal of Maps 12(supl.):178–189. https://doi.org/10.1080/17445647.2016.1187977

Deng Y, Wilson JP, Bauer BO (2007) DEM resolution dependencies of terrain attributes across a landscape. Int J Geogr Inf Sci 21:187–213. https://doi.org/10.1080/13658810600894364

Dikau R, Brabb EE, Mark RM (1991) Landform classification of New Mexico by computer. US Geol Surv Open-File Rep. https://doi.org/10.3133/ofr91634

Drăguţ L, Eisank C (2012) Automated object-based classification of topography from SRTM data. Geomorphology 141–142:21–33. https://doi.org/10.1016/j.geomorph.2011.12.001

European Soil Data Centre, Joint Research Centre, European Commission (2008) Global Landform classification. https://esdac.jrc.ec.europa.eu/content/global-landform-classification. Accessed 26 Feb 2022

Florinsky IV (2017) An illustrated introduction to general geomorphometry. Prog Phys Geogr 41(6):723–752. https://doi.org/10.1177/0309133317733667

Geological Survey of Japan, AIST (2019) Seamless digital geological map of Japan 1: 200,000V2. June 5, 2019 version. Geological Survey of Japan, Tsukuba, National Institute of Advanced Industrial Science and Technology. https://gbank.gsj.jp/seamless/index_en.html?. Accessed 26 Feb 2022

Geoscience Australia (2022) Climatic Extremes. https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/climatic-extremes. Accessed 26 Feb 2022

Geospatial Information Authority of Japan (2022) The GSI Maps. https://maps.gsi.go.jp. Accessed 26 Feb 2022 (**in Japanese**) (Regularly updated)

Hawker L, Uhe P, Paulo L, Sosa J, Savage H, Sampson C, Neal J (2022) A 30 m global map of elevation with forests and buildings removed. Environ Res Lett 17:024016. https://doi.org/10.1088/1748-9326/ac4d4f

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Wei S, Wright MN, Xiaoyuan G, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12:e0169748. https://doi.org/10.1371/journal.pone.0169748

Hengl T and Reuter HI (ed) (2008) Geomorphometry concepts, software, applications. Elsevier, Amsterdam, pp 772

Irsyam M, Asrurifak M, Mikhail R, Wahdiny II, Rustiani S, Munirwansyah M (2017) Development of nationwide Vs30 map and calibrated conversion table for Indonesia using automated topographical classification. J Eng Technol Sci 49:457–471. https://doi.org/10.5614/j.eng.technol.sci.2017.49.4.3

Iwahashi J (1994) Development of landform classification using the digital elevation model. Dis Prev Res Inst Annu 37B:141–156

Iwahashi J, Kamiya I (1995) Landform classification using digital elevation model by the skills of image processing—mainly using the Digital National Land Information. Geoinformatics 6:97–108. https://doi.org/10.6010/geoinformatics1990.6.2_97(inJapanesewithEnglishabstract)

Iwahashi J, Kamiya I, Matsuoka M, Yamazaki D (2018) Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. Prog Earth Planet Sci 5:1. https://doi.org/10.1186/s40645-017-0157-2

Iwahashi J, Pike RJ (2007) Automated classification of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86:409–440. https://doi.org/10.1016/j.geomorph.2006.09.012

Iwahashi J, Yamazaki D, Nakano T, Endo R (2021) Classification of topography for ground vulnerability assessment of alluvial plains and mountains of Japan using 30 m DEM. Prog Earth Planet Sci 8:3. https://doi.org/10.1186/s40645-020-00398-0

Jasiewicz J, Stepinski TF (2013) Geomorphons—a pattern recognition approach to classification and mapping of landforms. Geomorphology 182:147–156. https://doi.org/10.1016/j.geomorph.2012.11.005

Karimzadeh S, Feizizadeh B, Matsuoka M (2019) DEM-based Vs30 map and terrain surface classification in nationwide scale—a case study in Iran. ISPRS Int J Geo Inf 8(12):537. https://doi.org/10.3390/ijgi8120537

Keuper F, Wild B, Kummu M, Beer C, Blume-Werry G, Fontaine S, Gavazov K, Gentsch N, Guggenberger G, Hugelius G, Jalava M, Koven C, Krab EJ, Kuhry P, Monteux S, Richter A, Shahzad T, Weedon JT, Dorrepaal E (2020) Carbon loss from northern circumpolar permafrost soils amplified by rhizosphere priming. Nat Geosci 560(13):560–565. https://doi.org/10.1038/s41561-020-0607-0

Kwok OLA, Stewart JP, Kwak DY, Sun P-L (2018) Taiwan-specific model for VS30 prediction considering between-proxy correlations. Earthq Spectra 34(4):1973–1993. https://doi.org/10.1193/061217EQS113M

Lin P, Pan M, Beck HE, Yang Y, Yamazaki D, Frasson R, David CH, Durand M, Pavelsky TM, Allen GH, Gleason CJ, Wood EF (2019) Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resour Res 55(8):6499–6516. https://doi.org/10.1029/2019WR025287

MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley symposium on mathematical statistics and probability, Berkeley, University of California Press, vol 1, pp 281–297

Messager ML, Lehner B, Grill G, Nedeva I, Schmitt O (2016) Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat Commun. https://doi.org/10.1038/ncomms13603

Meybeck M, Green P, Vörösmarty C (2001) A new typology for mountains and other relief cases: an application to global continental water resources and population distribution. Mt Res Dev 21(1):34–45. https://doi.org/10.1659/0276-4741(2001)021[0034:ANTFMA]2.0.CO;2

Mihu-Pintilie A, Nicu IC (2019) GIS-based landform classification of eneolithic archaeological sites in the Plateau-plain Transition Zone (NE Romania): habitation practices vs. Flood Hazard Percept Remote Sens 11(8):915. https://doi.org/10.3390/rs11080915

Mori F, Mendicelli A, Moscatelli M, Romagnoli G, Peronace E, Naso G (2020) A new Vs30 map for Italy based on the seismic microzonation dataset. Eng Geol 275:105745. https://doi.org/10.1016/j.enggeo.2020.105745

Mulder VL, Bruin S, Schaepman ME, Mayr TR (2011) The use of remote sensing in soil and terrain mapping–a review. Geoderma 162:1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

National Research Institute for Earth Science and Disaster Resilience (NIED) (2020) J-SHIS Map 2020 version. https://www.j-shis.bosai.go.jp/map/?lang=en. Accessed 26 Febr 2022

Netzel P, Jasiewicz J, Stepinski T (2016) TerraEx-a GeoWeb app for world-wide content-based search and distribution of elevation and landforms data. Int Conf GISci Short Paper Proc. https://doi.org/10.21433/B3110dk1t0vc

Nobre AD, Cuartas LA, Hodnett MG, Rennó CD, Rodrigues G, Siveira A, Waterloo MJ, Saleska S (2011) Height above the nearest drainage–a hydrologically relevant new terrain model. J Hydrol 404(1–2):13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

OneGeology administration (2020) OneGeology. https://onegeology.org/home.html. Accessed 26 Feb 2022

Pain CF, Gregory L, Wilson P, McKenzie N (2011) The physiographic regions of Australia: explanatory notes. Australian Collaborative Land Evaluation Program and National Committee on Soil and Terrain, Canberra, pp 30. https://publications.csiro.au/rpr/download?pid=csiro:EP113843&dsid=DS4 Accessed 26 Feb 2022

Pike RJ (1988) The geometric signature: quantifying landslide-terrain types from digital elevation models. Math Geol 20:491–511. https://doi.org/10.1007/BF00890333

Poggio L, de Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, Rossiter D (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7:217–240. https://doi.org/10.5194/soil-7-217-2021

Rennó CD, Nobre AD, Cuartas LA, Soares JV, Hodnett MG, Tomasella J, Waterloo MJ (2008) HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in Amazonia. Remote Sens Environ 112:3469–3481. https://doi.org/10.1016/j.rse.2008.03.018

Sayre R, Frye C, Karagulle D, Krauer J, Breyer S, Aniello P, Wright DJ, Payne D, Adler C, Warner H, VanSistine DP, Cress J (2018) A new high-resolution map of world mountains and an online tool for visualizing and comparing characterizations of global mountain distributions. Mt Res Dev 38(3):240–249. https://doi.org/10.1659/MRD-JOURNAL-D-17-00107.1

Senna S, Maeda T, Inagaki Y, Suzuki H, Matsuyama H, Fujiwara H (2013) Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband strong motion evaluation. J Disaster Res 8(5):889–903

Senna S, Wakai A, Yatagai A, Jin K, Matsuyama H, Suzuki H, Fujiwara H (2019) Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband strong motion evaluation in Japan. In: Proceedings of 7th international conference of earthquake and geotechnical engineering, Malta

Speight JG (1984) Landform. In: McDonald RC, Isbell RF, Speight JG, Walker J, Hopkins MS (eds) Australian soil and land survey field handbook. Inkata Press, Melbourne, pp 8–43

Summerfield MA (1991) Global geomorphology: an introduction to the study of landforms. Wiley, New York, p 537

Tarboton DG (1997) A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour Res 33:309–319. https://doi.org/10.1029/96WR03137

Tarboton DG (2005) Terrain analysis using digital elevation models (TauDEM). http://www.faculty.umb.edu/david.tenenbaum/eeos383/taudem.html. Accessed 26 Feb 2022

Wakamatsu K, Matsuoka M (2013) Nationwide 7.5-arc-second Japan engineering geomorphologic classification map and Vs30 zoning. J Disaster Res 8:904–91. https://doi.org/10.20965/jdr.2013.p0904

Wang L, Liu H (2006) An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. Int J Geogr Inf Sci 20(2):193–213. https://doi.org/10.1080/13658810500433453

Yamashita N, Ishizuka S, Hashimoto S, Ugawa S, Nanko K, Osone Y, Iwahashi J, Sakai Y, Inotami M, Kawanishi A, Morisada K, Tana N (2022) National-scale 3D mapping of soil organic carbon in a Japanese forest considering microtopography and tephra deposition. Geoderma 406:115534. https://doi.org/10.1016/j.geoderma.2021.115534

Yamazaki D, Baugh CA, Bates PD, Kanae S, Alsdorf DE, Oki T (2012) Adjustment of a spaceborne DEM for use in floodplain hydrodynamic modeling. J Hydrol 436–437:81–91. https://doi.org/10.1016/j.jhydrol.2012.02.045

Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O'Loughlin F, Neal JC, Sampson CC, Kanae S, Bates PD (2017) A high accuracy map of global terrain elevations. Geophys Res Lett 44:5844–5853. https://doi.org/10.1002/2017GL072874

Yamazaki D, Ikeshima D, Sosa J, Bates PD, Allen GH, Pavelsky TM (2019) MERIT Hydro: a high-resolution global hydrography map based on latest topography dataset. Water Resour Res 55:5053–5073. https://doi.org/10.1029/2019WR024873

Yong A, Hough SE, Iwahashi J, Braverman A (2012) A terrain-based site-conditions map of California with implications for the contiguous United States. Bull Seismol Soc Am 102:114–128. https://doi.org/10.1785/0120100262

Zhang W, Montgomery DR (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resour Res 30:1019–1028. https://doi.org/10.1029/93WR03553

Publisher's Note