Geomorphology 413 (2022) 108373

Contents lists available at ScienceDirect

Geomorphology

journal homepage: www.journals.elsevier.com/geomorphology

Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland)

Krzysztof Dyba, Jarosław Jasiewicz *

Adam Mickiewicz University, Applied Geoinformatics Research Unit, Bogumiła Krygowskiego 10, 61-680 Poznań, Poland

ARTICLE INFO

Keywords:
Geomorphometry of plains
Gaussian Mixture Model
Surface texture
Uncertainty
Poland

ABSTRACT

Low-relief areas are not fully the main subject of geomorphometric analyses. The development of the automatic classification of landforms mainly focuses on landforms related to the fluvial morphogenetic cycle. Thus, the morphogenetic diversity of the plains is not reflected in the existing classification systems. The area of Poland where the low relief area exceeds 80 % of the country's territory and results in various morphogenetic processes was selected for the analysis. The purpose of the analysis was recognition of the differentiation of surface types. The first step includes selecting appropriate morphogenetic variables, the second unsupervised classification using the Gaussian Mixture Model, and the third one encompassing the interpretation, namely the labeling process. Twenty Land Surface Types were distinguished, five belonging to uplands, and the remaining 15 types of plains were divided into four subgroups: rolling plains, dissection plains, smooth plains, and near-flat plains. Compared with other classification systems, terrain forms, morphogenetic strides, and physiographic division. The comparison showed a strong correspondence between the morphogenesis of the area and the inventory of surface types, and the high consistency of the Land Surface Types patterns within physiographic units.

1. Introduction

Plains of various origins cover over 30 % of the Earth's surface and, depending on the morphoclimatic zone, are shaped by various processes: glacial, periglacial, fluvial, aeolian, to indicate only the most important. In low-relief areas, methods offered by automatic landform classification poorly reflect existing cartographic studies provided by classic geomorphology. Geomorphological studies developed various classification systems that are a compilation of information from many sources, including topography, lithology, aerial or satellite imagery, and spatio-temporal context and follow the morphogenic and chronological principles (Fenneman, 1917; Klimaszewski, 1956; Rączkowska and Zwoliński, 2015), which are not recorded in the topography itself.

Early works of Hammond (1954, 1964) or Wood and Snell (1960) introduced methods of terrain classification through a spatial taxonomy, where decisions resulting from observation and personal knowledge have been replaced by rigorous rules applied over the entire study area in the same way. The increasing availability of Digital Elevation Models (DEMs) and growing computational power replaced manual calculations with much faster computer routines (Evans, 1972). Subsequent works led to the classification of landforms directly from DEM, based on neighboring relation between cells (Peucker and Douglas, 1975); first and second terrain derivatives (Dikau, 1989; Shary, 1995; Wood, 1996; Schmidt and Hewitt, 2004); the topographic position of a cell in the close and distant neighborhood (Weiss, 2001); or by employing a computer-vision system (Jasiewicz and Stepinski, 2013). All those methods allow recognizing the limited number of fundamental landforms, typical of normal (fluvial) morphogenetic cycle (Mark, 1975), leaving variability of plains on the margin of classification systems.

MacMillan et al. (2000) proposed a classification system addressed to both low- and high-relief areas based on the combination of compound terrain variables relative to local surface-specific points (Peucker and Douglas, 1975). Such adaptation was possible by confirmation of parameters to the input data; however, the form inventory proposed by MacMillan et al. (2000) did not go beyond recognizing different parts of the watershed profile. A terrain signature (Pike, 1988) including widely understood textural properties was the first step to landscape-oriented analysis introduced later by Iwahashi and Pike (2007) in the form of a self-adapting hierarchical classification scheme. An system of Iwahashi and Pike (2007) includes 16 origin-agnostic forms defined by description rather than by names and similar to that proposed by MacMillan et al. (2000), capturing the variability of both low- and high-relief areas.

* Corresponding author.
E-mail addresses: krzdyb@amu.edu.pl (K. Dyba), jarekj@amu.edu.pl (J. Jasiewicz).

https://doi.org/10.1016/j.geomorph.2022.108373
Received 8 April 2022; Received in revised form 10 July 2022; Accepted 10 July 2022
Available online 16 July 2022

Automatic landform classifications merely relate to the form's origin or age. Several attempts toward origin-oriented surface classification used a supervised approach (Brown et al., 1998; Hengl and Rossiter, 2003; Prima et al., 2006). Those classifications require training areas where the surface genesis is inevitable and then use those areas to train an intrinsically complex classifier. Brown et al. (1998) explored supervised classification of different types of post- and pro-glacial plains, but the performance of their method was not confirmed on a larger, over-regional scale. Jasiewicz et al. (2014) proposed a new method of landscape classification based on the category co-occurrence matrix (Haralick et al., 1973) and captured a regional diversity of post-glacial plains, obtaining a performance of 70 % over the entire territory of Poland. However, the most significant limitation of supervised methods is that they cannot reveal new patterns in the data (Iwahashi et al., 2021), and the measure of success is the degree of compliance with existing, usually manually created maps. For that reason, supervised methods, including deep learning (Torres et al., 2019; Shumack et al., 2020; Xie et al., 2020) focus on the search for most effective classification methods reconstructing the complicated ground-truth (Du et al., 2019; Li et al., 2020; Janowski et al., 2022) rather than searching for alternative designs.

Unlike rule-based systems and supervised learning, unsupervised methods do not presuppose a closed inventory of landforms (Brändli, 1996; Irvin et al., 1997; Burrough et al., 2000; Adediran et al., 2004). So clustering works best in searching for new patterns leading to better classification schemes. For this reason, unsupervised methods have gained some popularity primarily in the study of extraterrestrial morphometric systems (Stepinski et al., 2006; Bue and Stepinski, 2006; Stepinski et al., 2007; Dan Capitan and Van De Wiel, 2012; Wang et al., 2017) and soil science (De Bruin and Stein, 1998). The weakest point of clustering is the dependency of results on numerous algorithms, free parameters, and variable selection, which makes unsupervised methods unsuitable for creating target cartographic works (Minar and Evans, 2008). However, recent studies proved the usefulness of clustering in optimizing mapping procedures (Wieczorek and Migoń, 2014; Szypuła and Wieczorek, 2020) and spatial analysis of complex geomorphological processes (Szymanowski et al., 2019).

The overview of previous studies concludes that despite the wide variety of plains, especially in post-glacial areas, no general methods have been made to consider the diversity of low-relief areas. In this article, we aim to present the quantitative variation of the geomorphometric features of the plains including several morphogenetic zones as a step toward developing assumptions for the automatic classification of plains. There is no clear definition of plains; therefore, we decided to use an area containing different surface types, expecting the plains to be separated in the analysis process as distinct from non-plain areas. As a study area, we selected the entire area of Poland as a lowland country dominated by plains of different origins (glacial, periglacial, fluvial, and locally coastal and aeolian), but also includes highlands and mountainous areas, which makes Poland a key area for such study (Jasiewicz et al., 2014; Szuman et al., 2021).

We do not make any assumptions about the range of plains, morphogenetic zones, or genetic-stratigraphic separations in the analysis process. The extent of the morphogenetic zones will be used only to assess the compliance of the proposed system with the existing concepts of relief development in Poland. We selected the unsupervised, unconstrained approach and Gaussian Mixture Model (GMM) (Day, 1969; Dempster et al., 1977) as a central part of the analysis. The GMM is a clustering system that provides so-called soft clusters, dividing an area into discrete units and determining the probability of belonging to each cluster. The clustering results were subjected to heuristic categorization and allowed to propose an unsupervised, non-genetic system that relates to the main morphogenetic zones occurring in Poland. The results were compared with other automatic methods and with existing expert-driven geomorphological maps in selected areas.

2. Materials and methods

2.1. Study area

Poland is a country in Central Europe, located between latitudes 49°–55°N and longitudes 14°–25°E (Fig. 1A). Over 80 % of the area does not exceed 200 m a.s.l. and areas above 600 m a.s.l. occupy no more than 4 % of the country's territory (Fig. 1B). The relief of Polish plains results from the repeatedly advancing the Pleistocene ice sheets and subsequent erosion and deposit processes (Galon, 1972; Starkel, 1980; Mojski, 2005) resulting in a continuous cover of Quaternary deposits (Fig. 1C). Morphogenetic zones of Poland (Gilewska, 1991; Kondracki, 2002; Solon et al., 2018) form latitudinal strips (Lencewicz, 1937). Starting from the north extends Coastland, including a narrow strip of coastal plains, shaped by the transformation of post-glacial areas as a result of changes in the extent of the southern Baltic coastline (Mojski, 1995). Northern Poland is occupied by Postglacial Lakelands, formed during the last ice advance (Kozarski, 1986; Szuman et al., 2021) built of glacial and fluvioglacial deposits (Fig. 1C), slightly transformed during the Holocene. The southern range of the lake districts marks the Last Glacial Maximum line (Fig. 1D), the most important geomorphological boundary in the Polish Lowlands (Galon, 1972). To the south of the Last Glacial Maximum extend post-glacial Denudated Plains; a surface where denudation processes have heavily transformed older glacial deposits and the original post-glacial relief has been partially or completely removed (Dylik, 1969). The three northern stripes are entirely classified as lowlands, but plains, covered by quaternary glacial deposits and loess, still constitute a significant part of the surface in the Upland and Foreland zones. Regardless of the latitudinal structure, plains areas (Fig. 1C) also include prevailing longitudinal river valleys shaped during the Pleistocene and Holocene, but often on older Tertiary assumptions (Starkel, 2001). Aeolian deposits include Holocene dunes that form narrow areas at the coastline. Older dune fields and single dunes appear in many places of Middle Poland formed at the end of the last glaciation outside the impact of the ice sheet (Nowaczyk, 1995). The loess covers in southern Poland, occurring in the uplands and foothills, are of a similar age (Badura et al., 2013). Regardless of inter-zonal differences, the relief inside the morphometric zones is also varied, resulting in various glacial denudation and fluvial processes.

The topographic dataset for analysis is 30 m floating-point DEM, created from original integer-based Digital Terrain Elevation Data (DTED L2) by adaptive smoothing and resampling (Jasiewicz et al., 2014). The DEM is projected to the PUWG92 coordinate system (EPSG:2180) and consists of 23,007 × 21,393 cells. The DEM is limited to the borders of Poland to avoid data inconsistencies caused by different techniques of DEM development and different original cartographic sources provided by other cartographic surveys.

2.2. Selection and description of morphometric variables

Even though several dozen geomorphometric variables have been proposed (Pike, 1988; Florinsky, 2017; Franklin, 2020) in the last fifty years, the complexity of the terrain surface is merely a composite of vertical and horizontal differentiation (Mark, 1975). The former is included in the term relief, the latter - texture, and the relation between vertical and horizontal irregularities are expressed in terms of relief distribution and inclination. The selection of variables is an essential step in geomorphometric analysis (Minar and Evans, 2008), but this step is not always analyzed in detail. In this study, variables were selected to represent the full spectrum of the morphometric diversity of plains dominating in Poland. The spatial pattern of variables was compared visually with our knowledge of the factors influencing the relief of the plains. We assumed that in low-relief areas, an important role plays the vertical and horizontal variability together with relief distribution. Based on the analysis of the occurrence frequency of landforms, we have designated two neighborhood scales applied in the variable calculation:

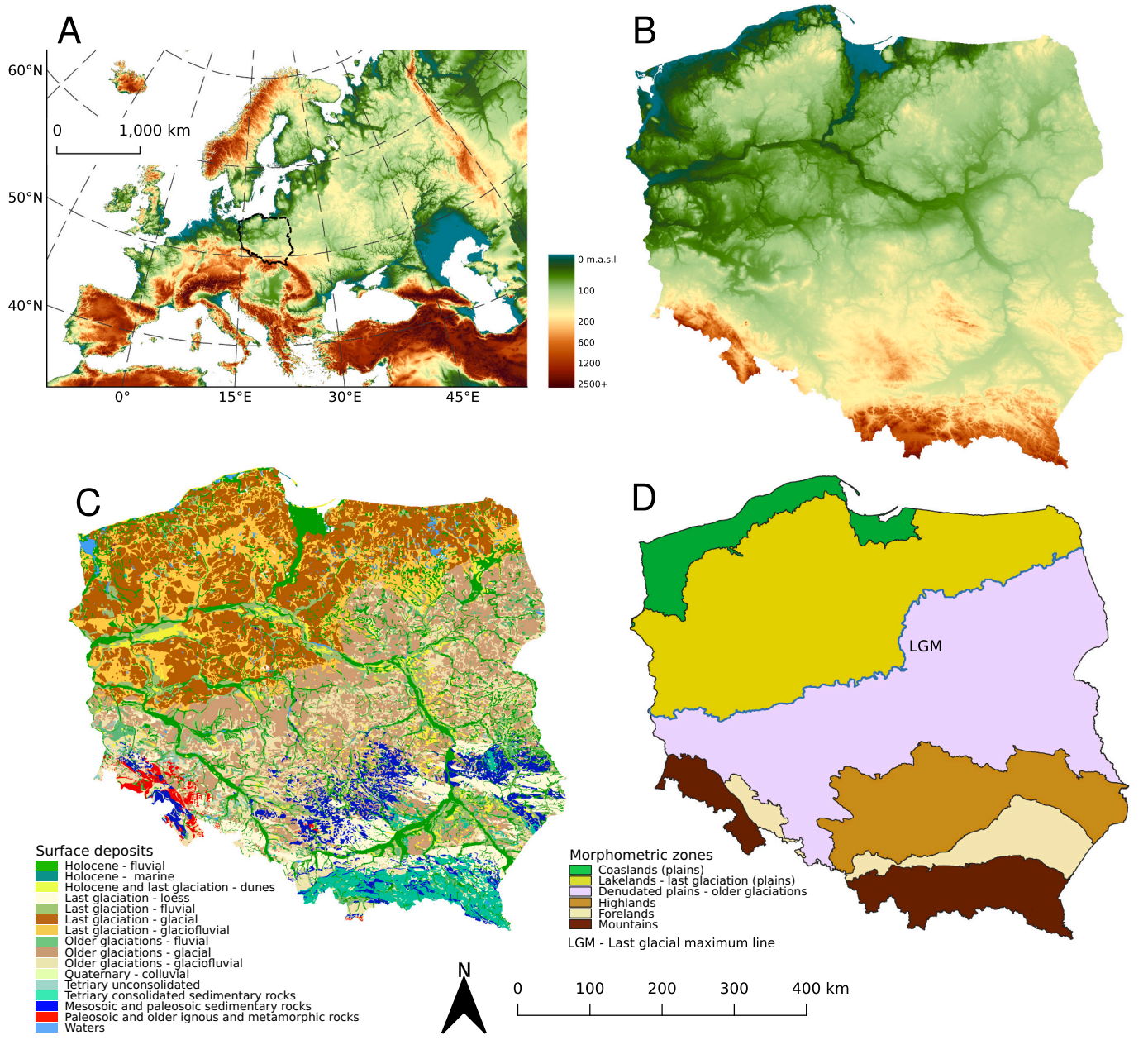

Fig. 1. A) Location of Polish borders superimposed on the Europe relief; B) Relief of Poland; C) Geology of Poland - origin of surface sediments, based on Geological Map of Poland 1:500,000 (Marks et al., 2006), simplified; D) Main morphogenetic zones (after Gilewska (1991)), boundaries adopted to physio-geographic boundaries (Solon et al., 2018).

a regional scale, with an extension of about 54 km, including the trend of plateaus and large valleys, and a local scale with a size of 900 m (31 pixels), adjusted to the width of medium valley forms. In the analysis, we omitted variables intended for modeling selected natural phenomena, such as solar radiation or hydrological processes, and variables that could not be correctly calculated for the entire study area. Finally, from more than 20 analyzed variables, we selected 10, presented in Table 1 and Fig. 2, that, in our opinion, represent best the variability of factors describing the relief of postglacial plains and carry different information.

The relative height of the terrain is expressed by Residual Elevation (RESE), Local Topographic Position (LTPI), Elevation above the erosional Base (EBAS), and Relief (RELF) supplemented by Absolute Elevation (ELEV). RESE and LTPI describe the position of a pixel against trends (Maxwell and Warner, 2015) calculated at the regional and local scale. The RELF variable is also calculated on the local scale and

Table 1
Morphometric variables used in the study. See details in the text.

Variable	Symbol	Range	Mean value	Unit
Absolute Elevation	ELEV	−0.3, 2483	170.9 ± 129.1	m.
Residual Elevation	RESE	−328.2, 1564.4	−0.24 ± 47.5	m.
Elevation above erosional base	EBAS	−71.6, 2168.1	85.2 ± 94.6	m.
Relief	RELF	0, 1091.8	32.4 ± 50.1	m.
Local Topographic Position	LTPI	−349.2, 458.6	−0.68 ± 11.5	m.
Ruggedness	RUGN	0, 321.2	6.3 ± 10.3	m.
Slope position	SPOS	−0.5, 0.47	0 ± 0.13	norm
Flatness	FLAT	0, 1	0.61 ± 0.36	norm
Surface noise	SNIS	0, 1	0.03 ± 0.02	norm
Mean convergence	MCON	0, 72.2	12.6 ± 3.67	deg.

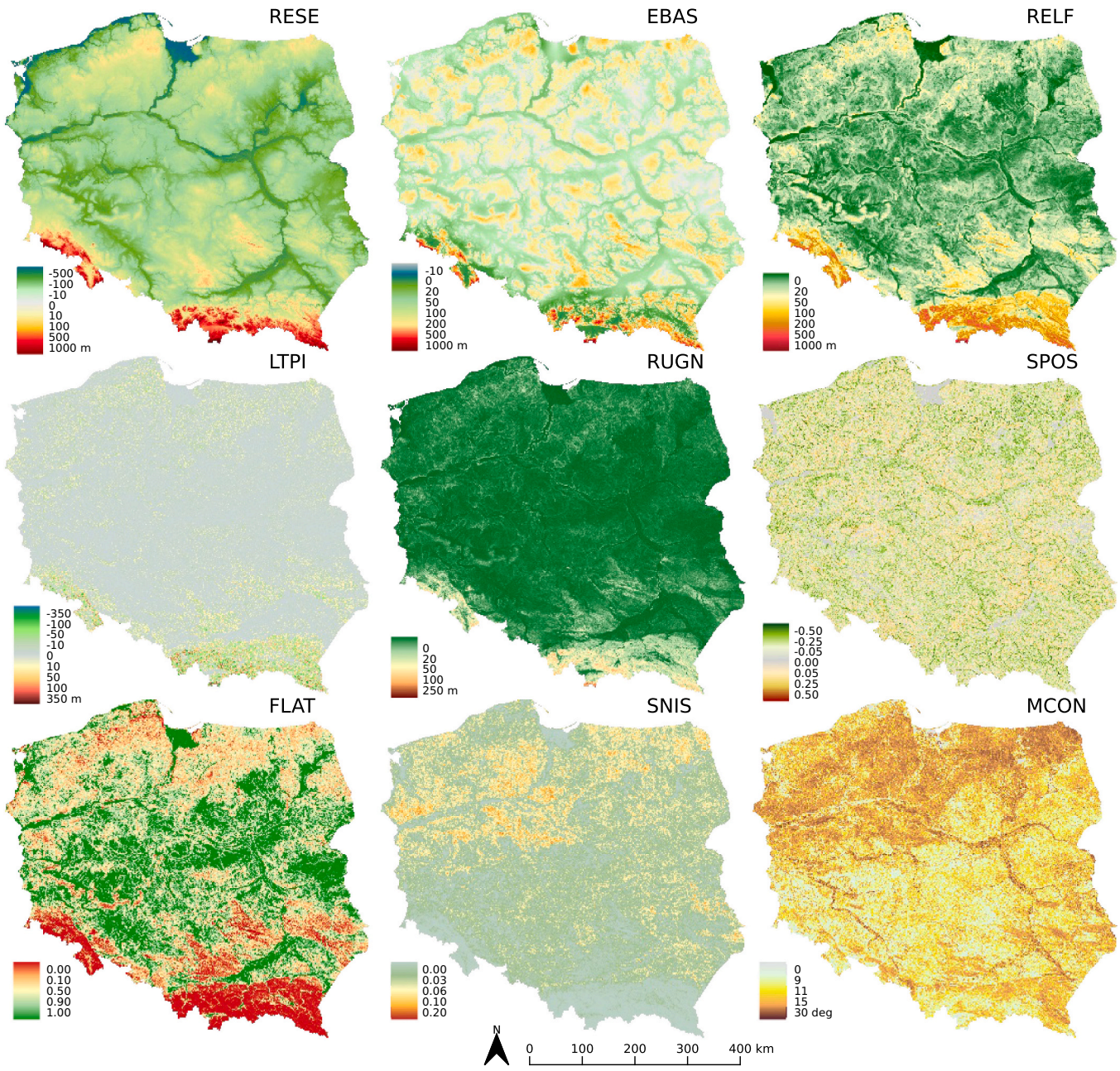

Fig. 2. Nine morphometric variables (ELEV is presented in Fig. 1), used in the analysis.

substitutes the slope inclination in low-relief areas. EBAS shows the relative height of each pixel above the hypothetical hydrologic surface, also known as isobase. It plays a similar role as HAND – Height Above Nearest Drainage (Rennó et al., 2008) but is not affected by watershed boundaries. That surface was interpolated from the 3rd and 2nd Strahler's order streams (Grohmann, 2005) for the stream network generated with *r.stream* package (Jasiewicz and Metz, 2011). The minimal basin size required to start the first-order stream was set to 1 km^2. ELEV is rarely used in classification systems, mainly because its role in recognizing elementary forms is marginal. Only Brown et al. (1998) and Burrough et al. (2000) used this variable in supervised classifications. We include ELEV because it closely relates to the variation in the elevation and morphogenetic zones (Gilewska, 1991; Solon et al., 2018).

The horizontal variability is a compound parameter of an elevation and distance (Pike, 1988) and is expressed in many forms, usually as a frequency of forms (Iwahashi and Pike, 2007). Due to the low frequency of forms in the plains, we use three variables to describe the texture features: Flatness (FLAT), Mean Convergence (MCON), and Surface Noise (SNIS). FLAT represents the percentage of near-flat cells in the local neighborhood and is closely related to the portion of near-level previously proposed by Hammond (1954). We assume the cell as near-flat if its slope inclination is below 1°, which, for 30 m DEM, corresponds to 2° in actual surface (Schmidt and Hewitt, 2004). MCON and SNIS have not been used so far and represent minor variability of the terrain surface. The first is calculated as a mean convergence index (Böhner et al., 2008) in the local neighborhood, and the second is a density of isolated pits and peaks calculated at the same scale. Both variables are especially useful for distinguishing between areas of different morphogenesis, i.e., surfaces covered during the last glaciation and those extensively denuded during the last glacial period (Dylik, 1969; Rotnicki, 1974).

Two other variables illustrate relief distribution (Etzelmüller et al., 2007). The Ruggedness (RUGN) is a mean difference between the central cell and each cell in the neighborhood of the local neighborhood (Riley

et al., 1999). Slope Position (SPOS) is the variable calculated as (*ELEV*) − (*min*(*ELEV*) + (*max*(*ELEV*)) / 2 and shows the position of the cell relative to the midpoint of the range and was used to describe dissection (Nir, 1957; Pike and Wilson, 1971) of the terrain.

Finally, we discuss the departure from using popular variables based on the first (slope, aspect) and the second derivative of terrain, namely different types of curvatures (Evans, 1972). In lowland areas, the values of these variables are minimal. Schmidt and Hewitt (2004) and Vaze et al. (2010) discussed the impact of DEM scale on terrain derivatives and noticed that low resolution of DEM may lead to high uncertainty and incorrect estimation of these variables, especially in flat areas. To minimize the impact of DEM resolution, variables based on terrain derivatives were excluded in favor of the variables operating on height differences. We noticed that relief is correlated with slope, and slope position (SPOS) is correlated with profile curvature, both at least $R^2 > 0.96$. Since both pairs of variables carry the same information, we used those with less uncertainty (Fig. 3).

2.3. Clustering process

Cluster analysis aims to simplify the complexity of a dataset by breaking the set into smaller units. The clusters obtained as a result of the analysis should be characterized by maximum internal cohesiveness and distinctiveness from other groups. This goal is easy to achieve when each of the variables describing the dataset has a regular distribution around the centers of the clusters, and the variables are not correlated. When these assumptions are met, simple and quick methods of hard clusterings, such as the popular k-means, are the best choice. However, such assumptions are difficult to meet in geomorphometric analysis. The terrain surface is a continuous phenomenon. It is difficult to expect the data structure to contain distinguished natural clusters, especially in lowland areas, where single forms smoothly pass into each other. In such a situation, the data structure manifests itself only in changes in density, and hard clustering cannot find natural clusters. K-means, for example, divide such the dataset into more or less regular units, which often provides abstracted information (Patel and Kushwaha, 2020).

When the dataset is complex without clear clusters, methods based on the Gaussian process are recommended (Qiu, 2010). The GMM (Day, 1969; Dempster et al., 1977) is the most popular variant of Gaussian clustering and is robust against the limitations mentioned above. GMM assumes that data structure is a mixture of normal distributions. This assumption is not fundamental but common in such a methodology (Day, 1969). When the number of variables is more than one, such a mixture becomes multivariate normal. A model-based clustering assumes that each observation (case) belongs to one of the components with a given weight, where each component is the density function of normal distribution, expressed by different mean and standard deviation. This algorithm allows the discovery of complex patterns by unmixing them into cohesive components that represent real patterns within the dataset. Weights assigned to each observation categorize GMM as so-called "soft clustering". The minimum weight will occur when the observation belongs to each component with equal probability. The $1 - weight$ determines the uncertainty of this assignment; thus, the best situation occurs when observation belongs to one cluster with $weight = 1$.

The GMM algorithm starts with randomly allocated Gaussian distributions for each separate variable in multidimensional space. At each iteration, the data likelihood is maximized via the two steps, an expectation and maximization (Dempster et al., 1977). During the first step, the algorithm estimates the probability that observation belongs to a given component, and in the second step, each component is updated to fit best all the assigned observations. The soft clustering is more robust than the hard one, so when the data comes from the dataset where "clusters" are more changes in density than separate subspaces, the lack of clear boundaries is masked by "soft" assignment to individual groups. It means that the size of clusters adopts given variables, the method does not require prior data transformations, and clusters follow the local densities in the data.

Despite strengths such as flexibility and the ability to handle uncertainty in data, selecting the optimal number of clusters is still a heuristic, supported by more or less formal criteria. The GMM supports Bayesian Information Criterion (BIC) (Schwarz, 1978) as formal support for determining the optimal number of clusters. The BIC limits the number of clusters by applying the penalty to too complex models. When data structure creates separate groups, the BIC allows finding the optimal number at the local minimum. However, when the dataset forms a continuous cloud with changes in density, the BIC values decrease monotonically, and the optimal number is somewhere where the BIC value becomes constant. An analysis of the differences in the BIC values between 2 and 30 clusters (Fig. 4) indicates that the data structure forms a cloud with densities rather than a series of clusters, and the optimal value is somewhere between 8 and 24. Thus, the formal criterion plays an auxiliary role, and the main factor becomes the understandability of the obtained patterns. The latter means that selecting the optimal number of clusters requires manual analysis of a series of maps, and the researcher's knowledge influences the result.

2.4. Software and data processing

We used the R programming language (R Core Team, 2021) for the analysis, in particular the following packages: *sf* (Pebesma, 2018) to manage vector data, *stars* (Pebesma, 2021) to manage raster data, *mclust* (Scrucca et al., 2016) to train the Gaussian mixture models, *future* (Bengtsson, 2021) to parallelize models training, and *recipes* (Kuhn and Wickham, 2021) to prepare a pipeline for data pre-processing. Additionally, GRASS GIS (GRASS Development Team, 2020) was used to generate derivative products from the DEM.

To avoid possible problems with non-normal distributions, all input data were standardized using Yeo-Johnson power transformation. Since it was impossible to perform the entire 5 billion cells, a representative sample of 3 million cells from the entire study area was selected, and GMM was fitted for this sample (Fig. 3). The model was then predicted over the entire area resulting in cluster and uncertainty maps with the same resolution as the input data.

3. Results

3.1. Number of clusters

In order to define the optimal number of clusters, we analyzed maps from 4 to 24 clusters. Each map was compared with the extent of morphogenetic zones (Fig. 1) and the patterns of geomorphometric variables (Fig. 2) in terms of knowledge retrieved by the given pattern. A visual comparison between individual cluster patterns and variables (Figs. 2 and 5) shows that the 4-cluster model follows the variability of FLAT and, to a lesser extent of RELF. The 8- and 12-cluster models are related to MCON, while 16 and 20 clusters (Fig. 6A) disclose the role of SPOS and RUGN. The simplest 4-cluster model does not yet show the division into morphogenetic zones; this differentiation becomes apparent with an increasing number of variables. The 8-cluster pattern uncovers differences between Lakelands and Denudated plains, while the 12-cluster model also reveals the distinctiveness of Coastal plains. The 16-clusters model is close to the final, 20-cluster version of the clustering and marks off the bottoms of broad river valleys. The differences between 16- and 20-cluster models concern mainly the presence of clusters that appear on edges between plateaus and valleys. Further increasing the number of clusters creates units whose distinctiveness is questionable and difficult to interpret. For this reason, further divisions have been omitted.

The results presented in this way indicate that the optimal number of clusters is between 16 and 20. The smaller number does not reveal the distinctiveness of an important type of plains, which are the bottoms of

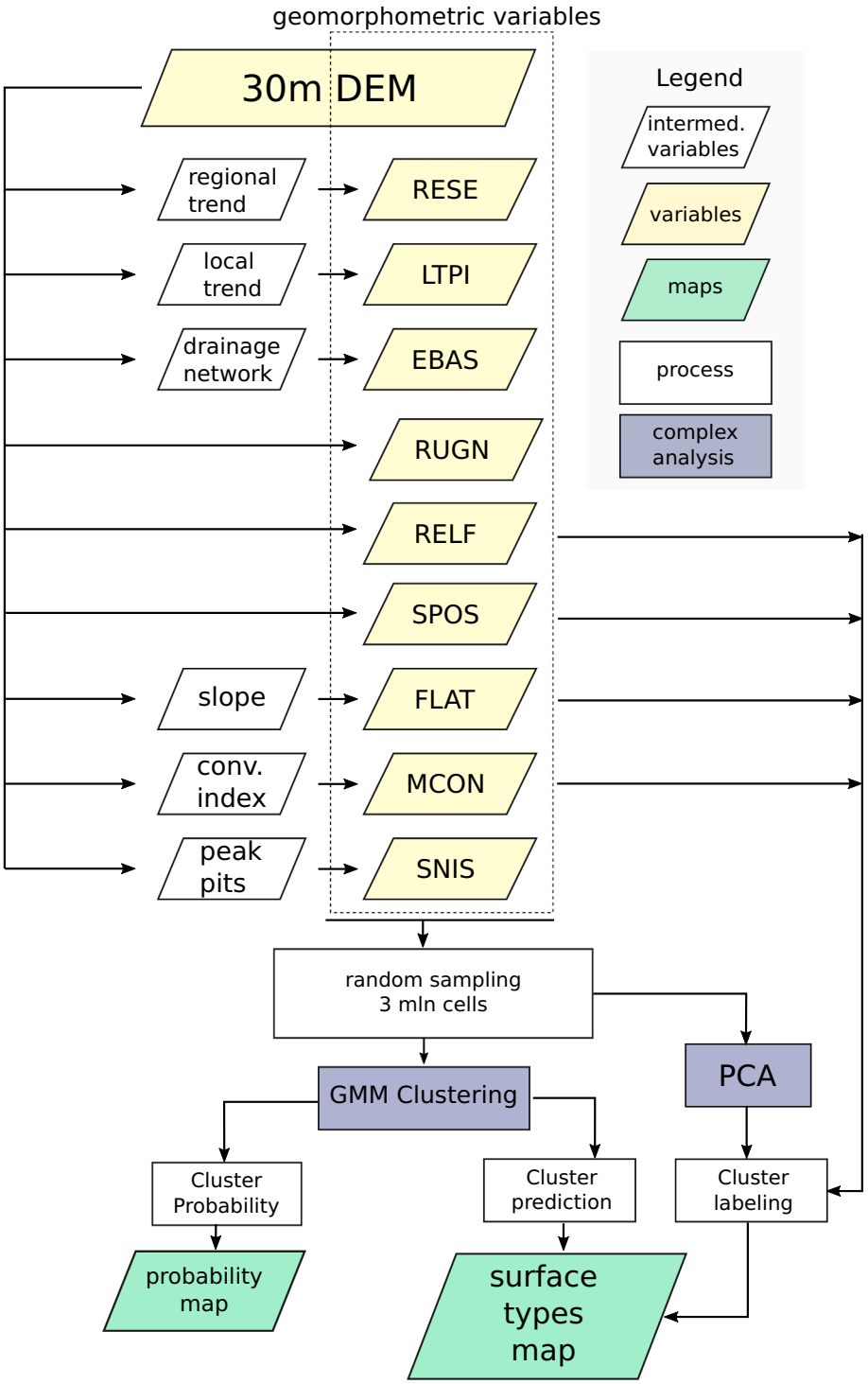

Fig. 3. Graphical summary of the data processing. For variable abbreviation, see Table 1.

river valleys, while the larger number creates units that are difficult to interpret. The 20-cluster version reveals the next type of surface - associated with the edges between valleys and plateaus and narrow valleys cutting through the latter. Dissections are important types of terrain forms, the removal of which reduces the value of the classification. Thus, we considered the number of 20 clusters optimal from the point of view of the problem of the paper. Fig. 6A presents the finally accepted result. It also should be emphasized that the increment of clusters is not hierarchical, i.e., clusters of the higher tier are not just sub-clusters of lower-tier divisions. This is primarily the result of the "soft" nature of the GMM clustering, namely the overlapping of Gaussian components.

3.2. *Cluster labeling*

The second step of the unsupervised analysis is labeling, a posterior heuristic process intended to give meaning to clusters. We already noticed that the spatial distribution of clusters is not even, and surface types of the same class are spatially related to the surfaces with a similar genesis. Such an observation tempts us to label units according to their

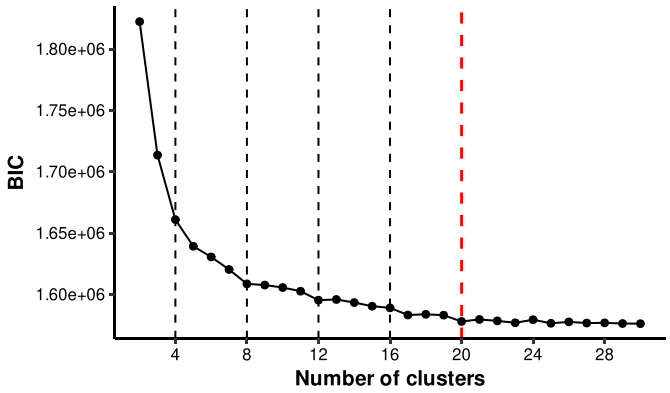

Fig. 4. The changes of Bayesian Information Criterion between 2 and 30 clusters. The red dashed line shows the final number of clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dominant genesis to refer to the earlier cartographic concepts; however, we want to avoid situations where the result is only an imperfect approximation of the already existing expert-driven maps (Iwahashi et al., 2021). Thus, the labeling aims to find an origin-agnostic classification scheme by discovering the divisive rules and then using these rules for final labeling. The remaining question is, what is represented by clusters? Clusters are intentionally delineated solely on the hard data, namely elevation and its transformations, and not taking into account the morphogenesis or chronology because there is either an interpretation or exists in the form of sparse data (Szuman et al., 2021). We observe that cells belonging to the same clusters create relatively homogeneous groups (Fig. 6), often containing an agglomeration of minor units, and are bounded by areas where the homogeneity is disrupted (Minar and Evans, 2008). Such entities definitely do not correspond to the term "landform" sensu Mark and Smith (2004) or Evans (2012), so we define the units on a more general level as "Land Surface Types". Following these assumptions, the labeling process (Fig. 7) includes the analysis of cluster variability in the principal components space and is supported by analyzing the relationship between individual morphometric variables.

3.2.1. *Surface variability inside the principal components space*

We applied Principal Component Analysis – PCA (Hotelling, 1933; Jolliffe, 2002) in the first step, reducing data complexity to a few gradients describing a significant part of variable space. Three first principal components (PC) applied to the scaled data explain almost 70 % of the data variability (Fig. 8A). The first PC explains that 39.7 % is negatively related to ELEV, RUGN, and EBAS and positively related to FLAT. It describes the gradient explained by local and regional vertical elevation variability, from most diverse (negative) to completely flat (positive). The second PC represents 15.9 % which is less than half of the first PC and is negatively related to all remaining variables representing both vertical and horizontal variability. The third PC represents 14.1 % of the variance (Fig. 8B), which is close to the second PC and is positively related to textural properties (MCON, SNIS) and negatively to local topographic position (RESE, LTPI, and SPOS). It means that areas with the highest positive third PC indicate high horizontal variation, while negative values represent low-lying, thus poorly textured areas.

The relationships between PC and variables show an interesting situation. While the first gradient divides the study area into uplands and plains, the second PC reveals only local variability; the third gradient relates to the textural features and shows that those features depend on the local relief (second PC). The first PC can divide the study area into uplands and lowlands and the second one into rough and smooth surfaces. The remaining 30 % of the variability requires deeper insight and heuristic use of selected variables. Based on the analysis of the mutual relationship between all applied morphometric features, we found four variables that best describe the differentiation of Land Surface Types. The first pair includes RELF and SPOS (Fig. 8C), both variables are parallel to the first PC and describe the vertical variability of the surface. The second pair (FLAT, MCON) is parallel to the second PC – and describes the details of the horizontal variance. It finally led to the 4-level classification scheme, where a surface type is a target unit, and the fourth level has only been applied to a few divisions (Fig. 6B). The labels are descriptive and hierarchical; the first capital letter refers to the group, the second to the subgroup, and the third to the land surface type. The fourth, optional lowercase letter, describes land surface subtypes.

Five clusters with negative first PC and local relief above 50 m were labeled as Uplands (U), and the remaining 15 clusters were tagged as Plains (P). Four clusters with the highest values of the first PC and FLAT close to 1 were labeled as Near-Flat Plains (PF). Although the latter category is distinguished solely inside the first PC gradient, for the semantic reason, we decided to narrow its role as a subgroup of Plains rather than a third distinct group separated from both Plains and Uplands.

3.2.2. *Uplands group*

Inside the Uplands group, three clusters with relief above 100 m form a subgroup of mountains (UM). The further division within Uplands solely based on the SPOS variable is relatively straightforward. Positive values of SPOS identify the highest parts of the mountains and highlands, mostly elevated parts or ridges (UME and UHE), while negative values indicate the inner part of the mountain (UMI) or, in particular, inter-mountain valleys (UMV) lower mountains and highlands (UMS).

3.2.3. *Plains group*

After separating the PF subgroup, and the remaining 11 clusters belonging to the Plains, we can distinguish two subgroups using the PCON variable: rolling (PR) and smooth (PS). We decided on the term "rolling" as an intermediate between "hilly" and "plain" (Pike and Wilson, 1971). Such division follows the second PC's gradient, but Fig. 8C reveals the existence of a transitional subgroup, which combines the features of both previous surfaces. Detailed analysis of spatial distribution shows that the given subgroup connects with slope edges and deeply incised valleys. Thus, we labeled this as Dissections (on) Plains (PD).

3.2.3.1. Rolling plains. The PR subgroup can be divided by the density of secondary landforms forming a surface noise into three Land Surface Types: high (density) (PRH), moderate (PRM), and low (PRL). This attribute is expressed by the pair FLAT-MCON (Fig. 8C), where the mean value of the convergence index increases within the LR group with the decreasing percentage of flat areas. Finally, the SPOS variable allows dividing the PMR and PRL type into upper (PRMu, PRLu) and lower (PMRl, PRLl) subtypes.

3.2.3.2. Dissections on plains. The PD is a subgroup representing mostly elongated landform elements connected with valleys. The subgroup divides further into three types that correspond to deeply cut valleys (PDV) or represent upper edges (PDEu), more inclined and partially eroded, and lower parts (PDEl), of narrow slopes, usually almost flat.

3.2.3.3. Smooth plains. Clusters belonging to PS arrange along FLAT-PCON gradients similar to PR and PD groups, but values of the MCON are significantly lower. All three clusters in this subgroup have very similar relief but vary with FLAT and SPOS. It allows for labeling them as inclined (PSI) and gentle (PSG). The third remaining cluster is very close to near-flat surfaces and was marked as PSF.

3.2.3.4. Near-flat plains. The PF subgroup shows the variation primarily by MCON variable and includes the following Land Surface Types:

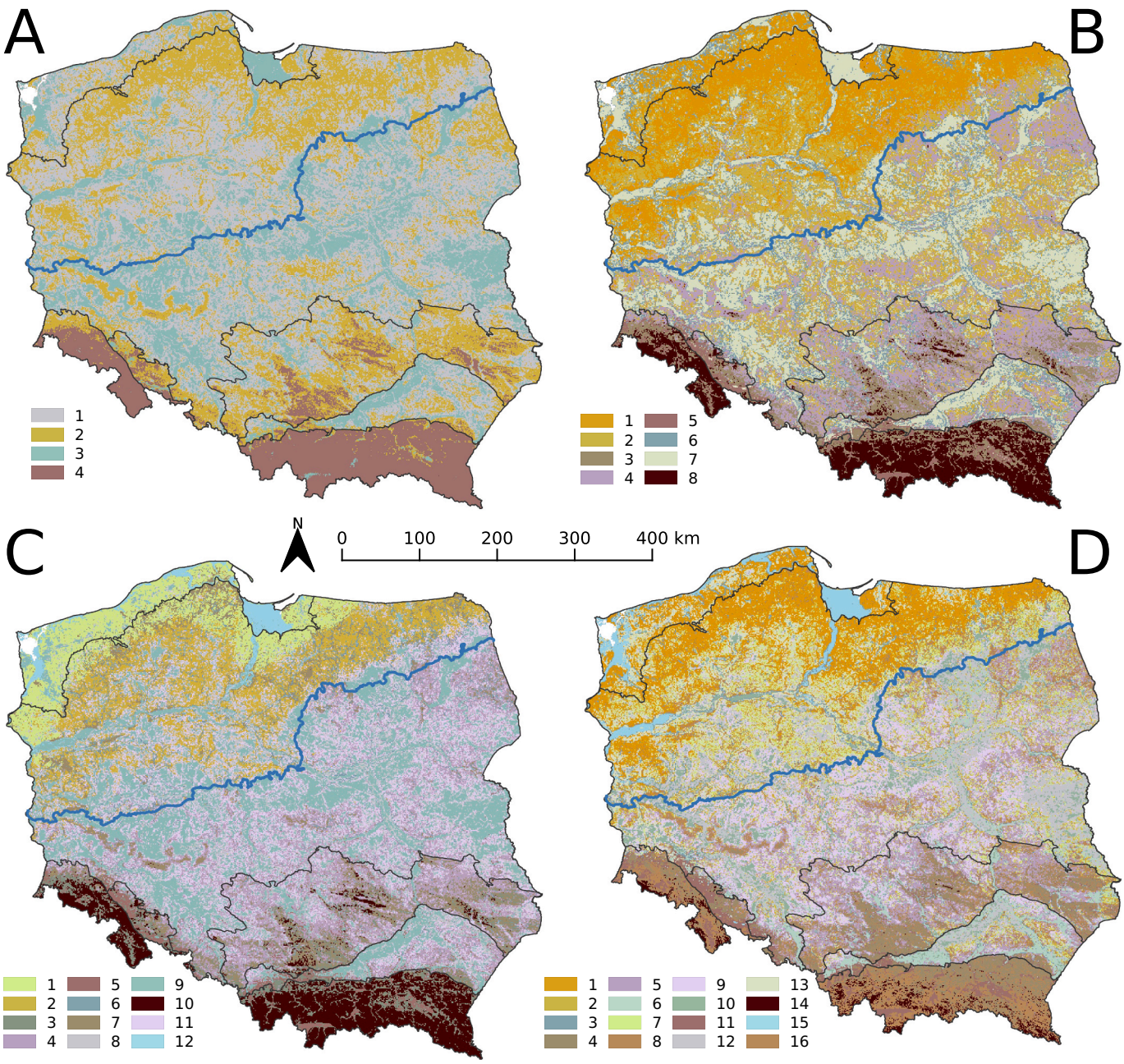

Fig. 5. Result of clustering with: A) four; B) eight; C) twelve; D) sixteen number of classes. Colors are meaningless, however partially relates to the symbology presented in Fig. 6. Blue line marks extend of Last Glacial Maximum (LGM). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

rough (PFR) with little secondary relief, dissected (PFD), and smooth (PFS) or completely flat. The latter divides further into upper (PFSu) and lower (PFSl) subgroups by the SPOS variable.

3.3. Uncertainty of the clustering

The GMM assesses the probability that the observation comes from the given cluster and thus assesses the uncertainty to what extent a given type of surface describes the actual terrain. Fig. 9 presents the spatial distribution of uncertainty and its relation to the Land Surface Types; namely, the pattern of uncertainty follows the spatial distribution of the classes. We observed that Land Surface Types have different uncertainties, and these differences affect the spatial distribution of uncertainty. A detailed analysis of these differences is beyond the scope of this paper; however, we notice that upland classes (UME, UMI, UMS, UMV) have uncertainty significantly lower than plains, except for the complete flat PFSl. Other classes have skewed distributions (Fig. 9C), with a minimum close to 0, or even uniform (PRMu and PRLl), which means that belonging to these classes is vague. This situation is not surprising and results from a low variation of geomorphometric variables in lowland areas.

For the same reason, the spatial distribution of uncertainties (Fig. 9B) at the surface type-level shows that the uncertainty increases at the boundaries between patches of Land Surface Types. The low uncertainty in the elevated areas results primarily from the high height variability and thus minimizes transition areas between land surface classes. In lowland areas, on the other hand, slight differences in altitude make the transition zones extensive and the boundaries between the forms fuzzy. This is particularly evident in rolling plains, where landforms inherent in the normal morphogenetic cycle have not developed, and thus

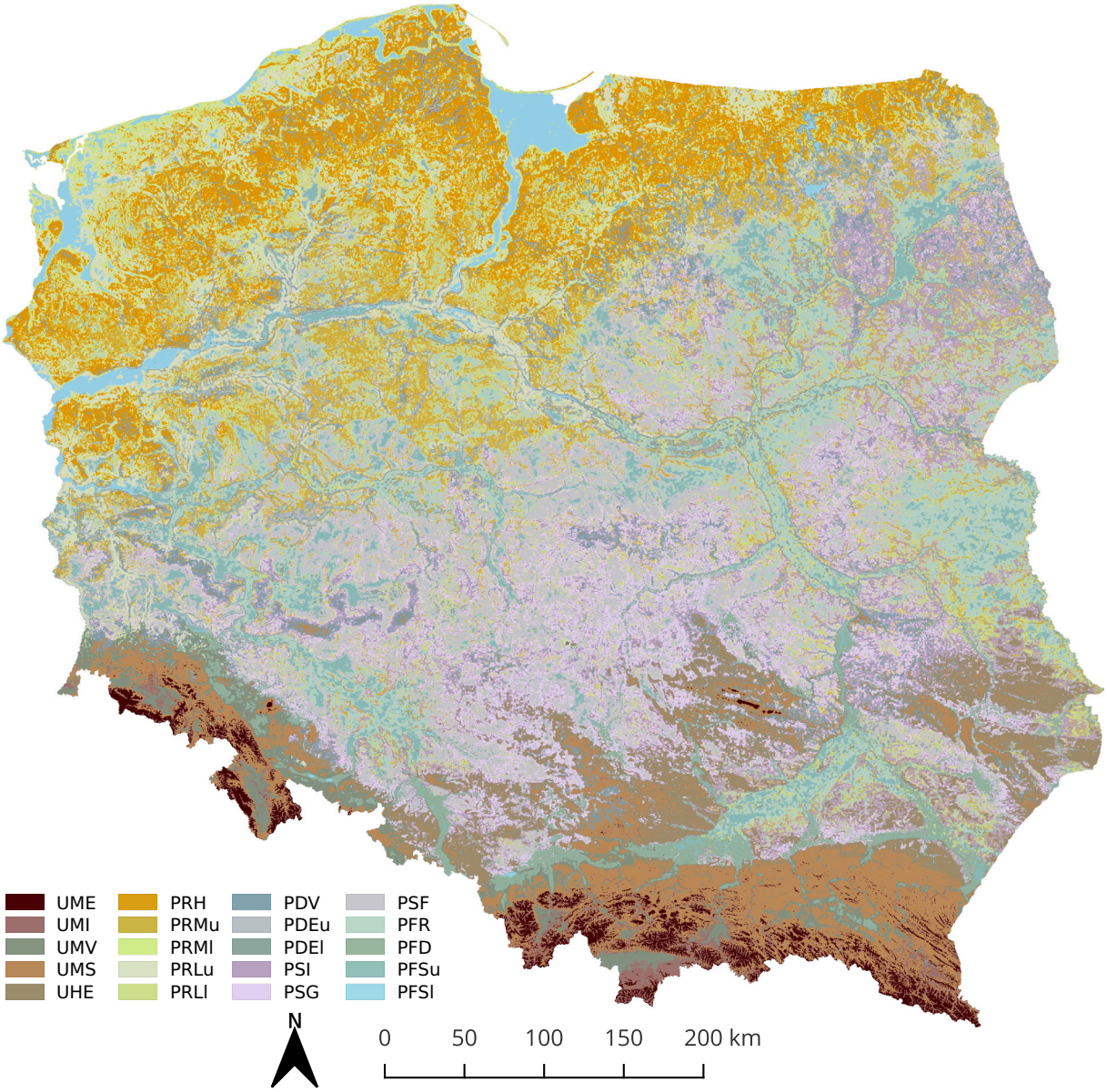

Fig. 6. Result of clustering with 20 clusters. See the Fig. 7 for detail classification and Table 2 for full land surface types names.

"transition zones" are the dominant element in the relief classification.

4. Discussion

The quality assessment of the unsupervised classification is a combination of understandability, usefulness, and novelty (Fayyad and Uthurusamy, 1996) that belong to the domain of subjective evaluation. The successful labeling (Fig. 6) fulfills the understandability test, and the next two will be evaluated by discussing the spatial distribution of surface types, especially concerning morphogenetic zones, and comparing the results with other mosaics.

4.1. Spatial distribution of land surface types

Fig. 10 shows that there is a strong correspondence between morphogenetic zones and land surface types, but it does not allow for the conclusion that Land Surface Types are assigned to particular zones. The correspondence is a matter of quantitative differences rather than a simple relationship between Land Surface Types and morphogenetic zones. The detailed regional analysis goes beyond the aims of this work; therefore, in the discussion, we will analyze to what extent the spatial distribution of surface types is related to the geomorphological processes and landforms taking place in a given zone.

Table 2 contains a list of the distinguished Land Surface Types compiled with the landforms and morphogenetic zones. Only the Mountains have a distinct set of Land Surface Types (U) which are only partially present in the highest parts of the Highlands. The variability of the Plains to the north and south of the Last Glacial Maximum corresponds well to the concept of the morphogenesis of the Central European plains (Dylik, 1969; Mojski, 1995; Hughes and Woodward, 2009; Murton, 2021) with the division into two zones, young post-glacial and older post-glacial, denudated under periglacial conditions. Plains belonging to the rolling subgroup, especially those with the highest density of irregularities, represent very young surfaces related to the activity of the last ice sheet with a shallow impact of denudation. On the other hand, smooth (PS) and near-flat (PF) subgroups cannot be considered explicitly as areas denudated in periglacial conditions (Dylik, 1969) because they appear both north and south of the Last Glacial Maximum. The Land Surface Types belonging to the PS subgroup north of the Last Glacial Maximum indicate various processes, including subglacial

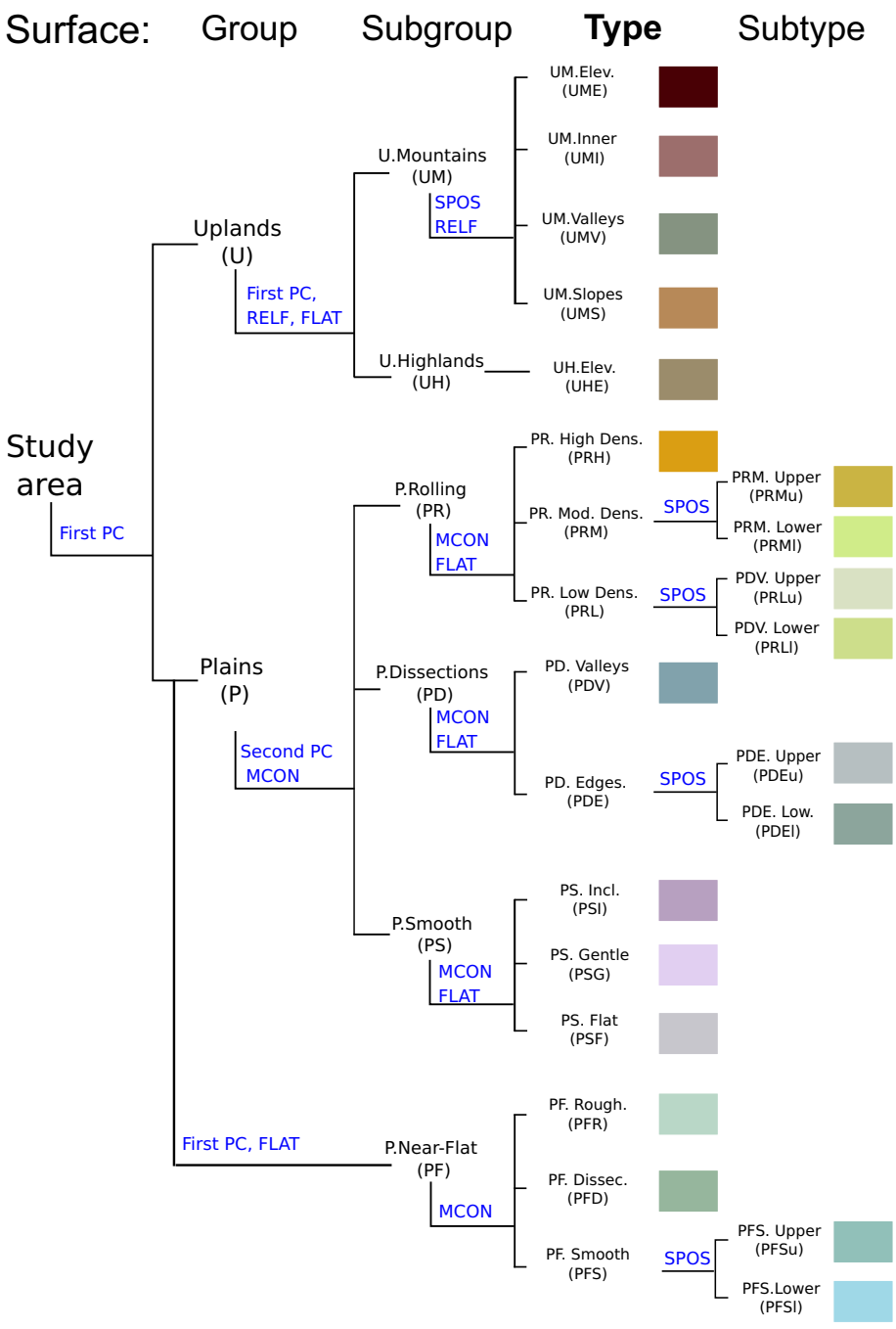

Fig. 7. Classification process (labeling) into land surface types Detail description of the process contains Subsection 3.2. See the *SEC:acronyms* section for full names.

exaggeration and ice sheet sideslips (Szuman et al., 2021). On the contrary PR subgroup south of Last Glacial Maximum appears mainly on highly carved valley edges and indicates intense erosional processes.

All classes belonging to the dissection plains (PD) subgroup appear both north and south of the line of PGM. Surfaces belonging to this subgroup form mostly relatively linear forms. Near-flat (PF) surfaces include terrains where areas inclined more than 1° appear across all plains. The most unusual appears in flat areas named rolling (PFR). Its extent coincides with boggy surfaces in river valleys and higher plains. The PFD (dissected) class encompasses flat but deep valley bottoms, and in practice, the class is limited only to the Forelands. Near-flat areas occur throughout Poland and are mainly associated with the bottoms of river valleys or boggy plains dominated by organic deposits developed during the Holocene. Surfaces PDSl and PRLl dominate in broad valleys in the Lakelands area, while smaller watercourses exist as deeply indented forms (PLDF). The PFS and PFR surface types dominate in wide river valleys cutting periglacial plains. The rough near-flat plains result from its frequent renewal during Holocene floods and the presence of numerous oxbow lakes.

Regardless of compliance with the extent of morphogenetic zones, we also observe a strong relation to the physico-geographical units (Solon et al., 2018) at the mezo-region level (Fig. 10). The physico-geographical division of Poland is an implementation of the European Landscape Convention. Mesoregions are the basic division unit belonging to higher-order units: macroregions, provinces, and mega-regions. The division is an update of the original division by Kondracki and Rychling (1994), adjusted to the new geological and geomorphological data. The course of the borders was made by hand on a scale of

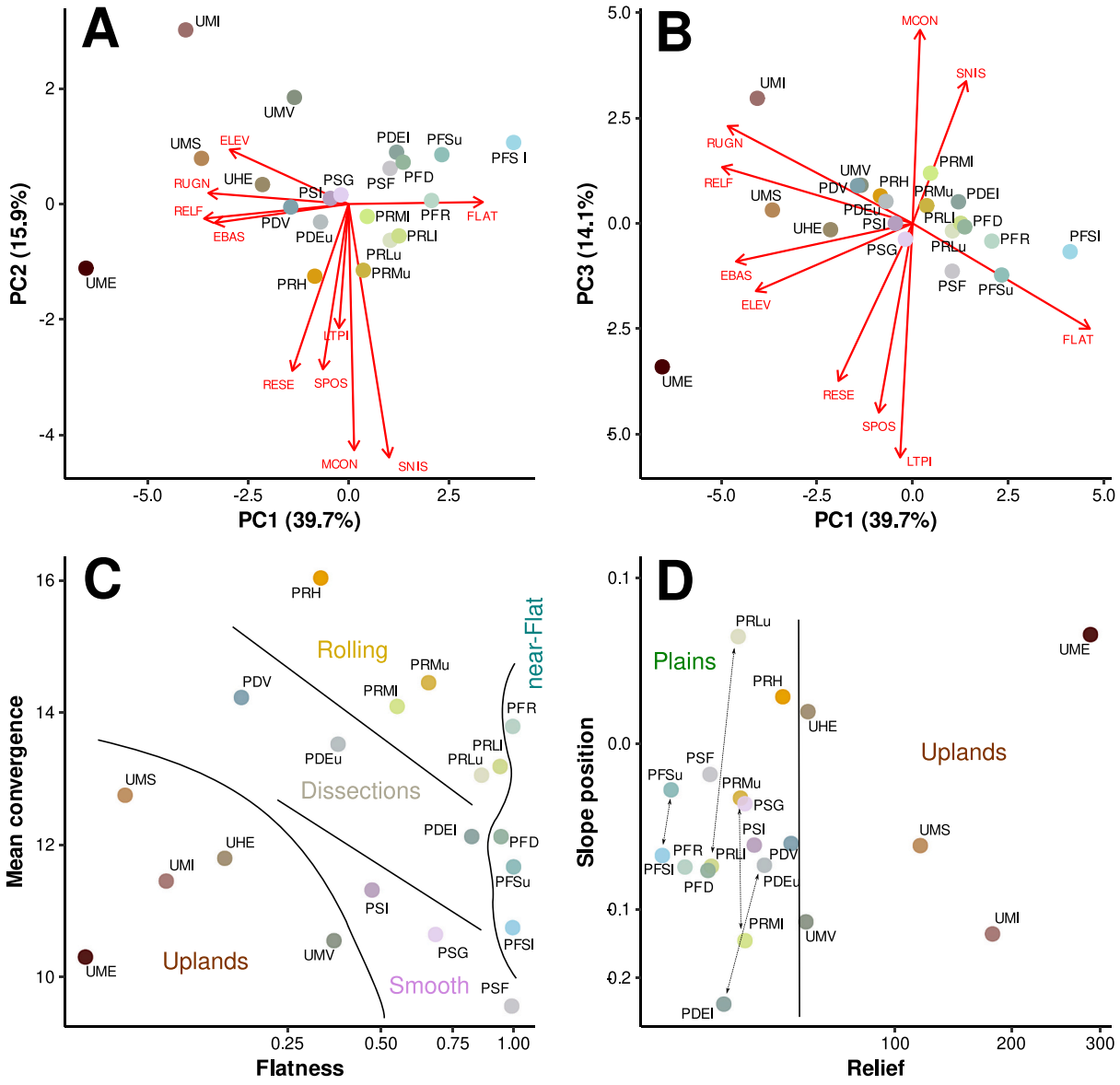

Fig. 8. Cluster distribution: A–B) against principal components (red arrows indicate variable's loadings); C–D) relief variables representing vertical variability (RELF and SPOS) and textural variables representing horizontal fluctuations (FLAT and MCON). Arrows on chart D show the position relation between lower (l) and upper (u) subtypes. Dots represent centers of clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1:50,000 and took Poland's topographic geological, geomorphological, and hydrological diversity. Thus, it becomes the best source for comparing the convergence of the classification results with the expert understanding of this phenomenon. Such a comparison confirms that the concept of classification, including the selection of variables and the method of grouping and labeling the variables, creates a pattern recognizable to those familiar with the geographic division of the country.

Within the Lakelands zone, there is a visible difference between the northern and southern parts. The PRH surface type dominates north, and to the south, it passes into rolling plains but with minor texture (PRM, PRL) mixed with smooth (PS) and near-flat (PF) plains. The diversity of Land Surface Types is mainly an effect of the differences in the dynamics of the Weichselian ice sheet, associated with the activity of the Baltic Ice Stream Complex (Punkari, 1997) with branches operating over a soft substratum of relatively gentle topography. The northern part was shaped during the Pomeranian phase, a regular part of the Baltic ice sheet (Punkari, 1997), and was subject to glacial thrusting and accumulation processes. The southern part is a zone of the influence of numerous but short-term and reduced advances of ice lobes that left a thin cover of glacial deposits (Szuman et al., 2021).

A variety of patterns also characterizes the denudated plains, and the lowest part of the Highlands zone. There is a difference between the western and central parts, which are dominated by smoother surfaces (PSF, PSG), and areas that dominate the eastern part with more textural features (PSI, PDV). The quantitative differences between the surface types in the western and eastern parts have a complex genesis, resulting from regional climatic differences affecting the intensity of denudation during the Plenivistulian and differences in the older, post-glacial relief (Rotnicki, 1974). The locally occurring inland dune fields are not associated with any distinctive Land Surface Type but are a mixture of different types of PD and PR, mostly PDV and PRH.

4.2. Comparison with other mosaics

The validity of spatial patterns resulting from clustering is usually

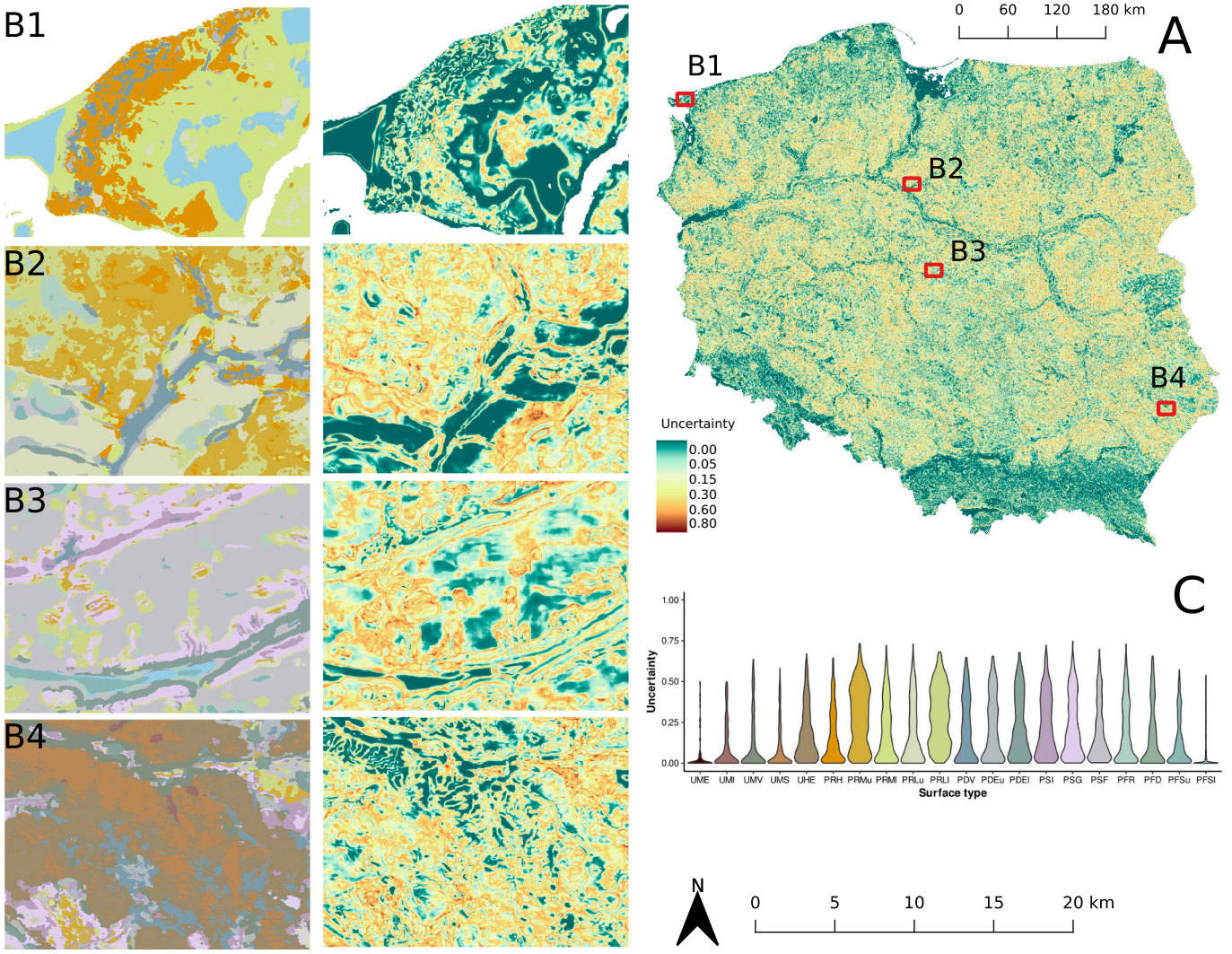

Fig. 9. The distribution of uncertainty: A) over entire study area; B) detailed, inside selected areas (location are marked on plate A); C) inside land surface types (violin plot).

determined by comparison with previous concepts, namely existing geomorphological or geological maps (Iwahashi et al., 2018, 2021). Although this approach does not assess the quality of clusters, such comparison gives an indirect answer to which of the existing divisions can be approximated by the morphometric parameters used in the clustering process. We compare the results of clustering with other mosaics (Fig. 11), including the classification made by the Hammond (1964), Terrain Position Index (Weiss, 2001), Iwahashi and Pike (2007) and geomorphons (Jasiewicz and Stepinski, 2013). Each of the automatic classifications was made in one step for the entire country based on the default set of input parameters, using the same elevation model that was used in the project. The classification results for the Hammond system and Terrain Position Index are not compared with the test areas, as their inventory of forms other than plains is minimal.

Additionally, as an expert-driven map, we use four test areas from a pilot project of a digital geomorphological map of Poland (Rączkowska and Zwoliński, 2015). Those testing areas include 1) Coastal plain (Wolin Island), 2) Lakelands (Chełmno-Dobrzyń Lakeland), 3) Periglacial plains (Kutno plain), and 4) Highlands (Roztocze) (Fig. 12, Appendix A).

A detailed comparison of mosaics shows that all surface patterns, regardless of the different number of classes, are similar and relate to expert-driven maps, but the extent of forms and exposed details differ in each case. The differences result primarily from a different approach to defining the boundaries in manual and automatic approaches. Manual delineation usually starts with setting main boundaries and iterative division into successive units. In an automatic case, each cell is labeled separately, and the spatial coherence of units results from the continuous nature of the Earth's surface. Moreover, the different extent of exposed forms is a consequence of calculating variables, such as textural parameters or relief by focal analysis that leads to the averaging of values in the neighborhood defined by the size of the window is one of the factors influencing the granulation of the distinguished classes (Drăguţ and Eisank, 2011).

Each of the presented in Fig. 12 has strengths and weaknesses. Expert-driven maps use complex chronological and morphogenetic units to differentiate surfaces containing similar landforms. For that reason, our method preserves different surface types, for example, various types of near-flat surfaces on Wolin Island (row 1). On the other hand, the proposed approach does not allow easy identification of ridge-valley systems, especially in upland areas (row 4), which is very easy for geomorphons, designed initially for such landscape types. Immature surfaces of postglacial areas (Lakelands, row 2) are the most significant challenge for automatic methods, especially those that use only fluvial landforms. Our approach distinguishes the surface of the valley bottom (green on the geomorphological map) from the Lakeland plateau (purple) and partially the variability of the plateau; the remaining methods recognized only river channels and valley slopes. In the periglacial

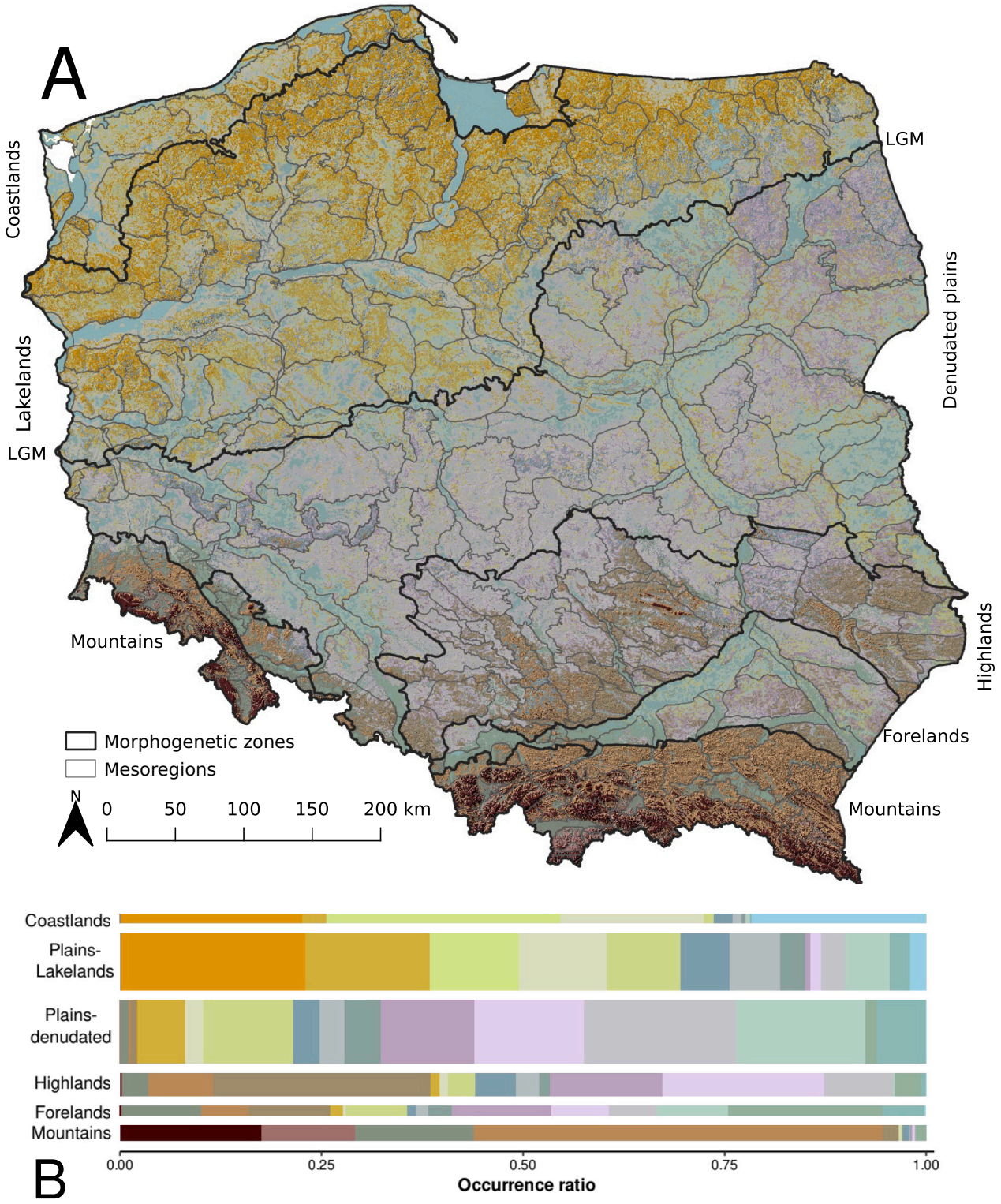

Fig. 10. Results of classification, compared with the extent of morphogenetic zones (see Fig. 1). A) Cluster distribution on the background of morphogenetic zones and physico-geographical mesoregions of Poland (Solon et al., 2018); B) Correspondence between morphogenetic zones and designated land surface types. LGM - Last Glacial Maximum.

plains (row 3), all three automatic methods recognize simple landforms, but only our method can distinguish between three types of planar surfaces, Except the size of slopes, which are more extensive compared to reality.

5. Conclusions and future works

The paper presents an unsupervised classification of surface types in low-relief areas, using Poland as a key test area. The procedure aims to define rules valid for the geomorphometric analysis of plains by various morphogenetic processes. As part of the procedure, we analyzed

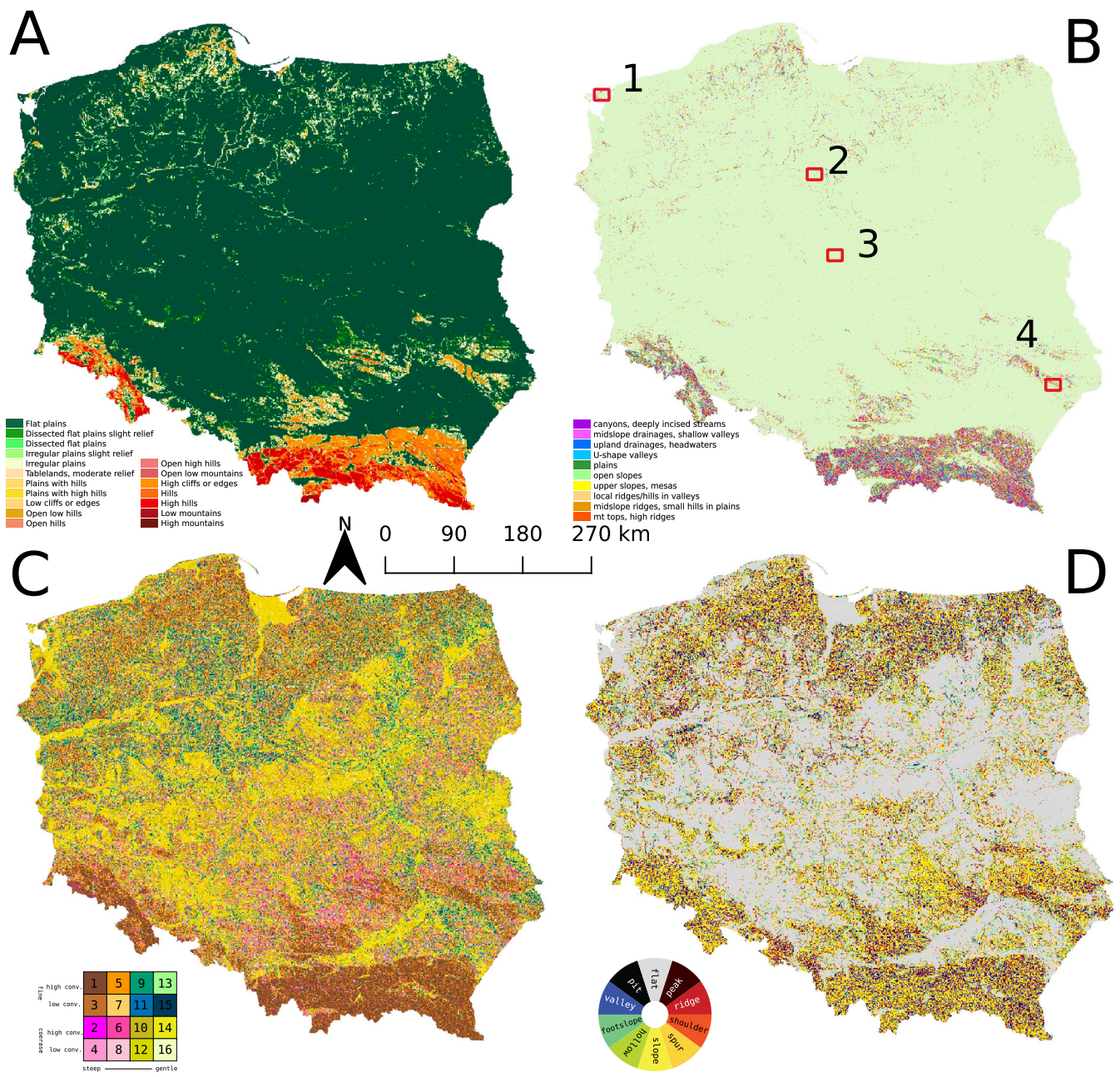

Fig. 11. Morphometric mosaics obtained with different methods. A) Hammond (1964) system; B) Topographic Position Index (Weiss, 2001); C) Iwahashi and Pike (2007); D) Geomorphons (Jasiewicz and Stepinski, 2013). Rectangles on Plate B marks location of testing areas presented on Fig. 12.

geomorphometric variables with great potential to differentiate between different types of plains. During the analytical process, we did not introduce any morphogenetic constraints leaving the determination of the extent of the plains and the surface types to the algorithm. We showed that the basis of an adequate classification is the balance between the variables describing the vertical and horizontal variability of the terrain. New variables describing secondary surface features (MCON and SNIS) have been proposed and shown to play an important role in distinguishing between different plains of glacial, periglacial, or fluvial origin. The Gaussian Mixture Model was used as a clustering algorithm. The Gaussian Mixture Model unmixes natural clusters and provides a soft clustering coupled with an assessment of the uncertainty of assignment to a given cluster. The obtained clusters were successfully labeled, recognizing four types of plains: rolling, smooth, near-flat, and dissections. Despite the application of the unsupervised classification and the use of different variables, the result refers to the Hammond (1954) rule system but is adapted to the postglacial plains of the European Lowlands.

The map of Land Surface Types is an additional effect of the research and was made using the limited cell-oriented method, but it proved to be very effective. The designated groups of surface types correspond partially to the range of the main morphogenetic zones; however, surface types cannot be used as direct indicators of morphogenesis. The relation is rather quantitative than qualitative and limited by the equifinality principle (Haines-Young and Petch, 1983; Beven, 1996), especially when the number of descriptors is limited. Nevertheless, our analyses open the way to detailed geomorphometric analyses of the plains, primarily carried out on a large spatial scale. Future work will include the implementation of recognized principles in cartographic-oriented research using object-oriented analyses (Drăguţ and Blaschke, 2006) and computer vision approach (Jasiewicz and Stepinski, 2013).

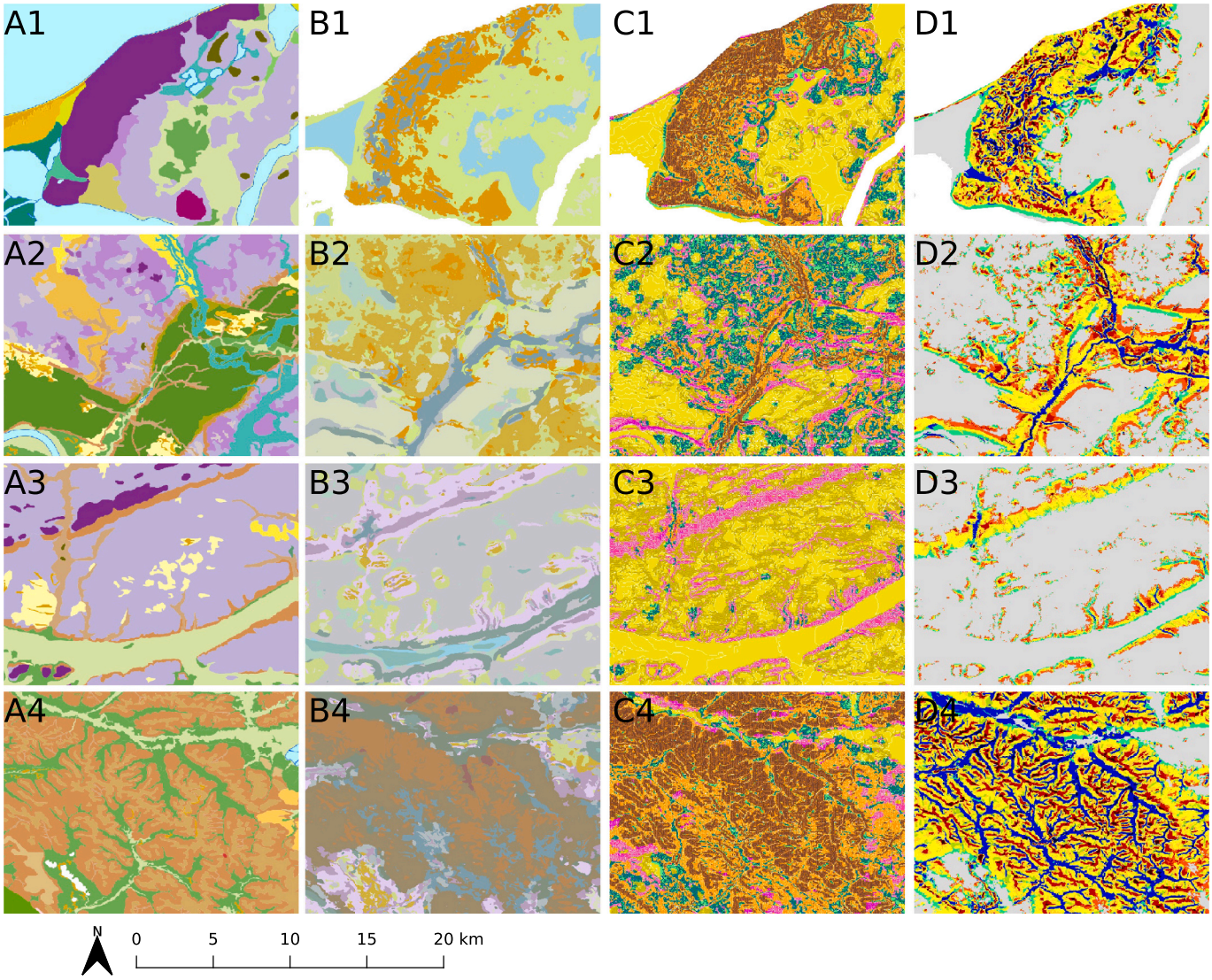

Fig. 12. Detailed comparison of expert-driven maps (Column A) with geomorphometric mosaics. Column B – our classification; Column C – Iwahashi and Pike (2007); column D – Geomorphons. Location of testing areas 1–4 is presented on Fig. 11, plate B. For expert-driven maps legend see Appendix A.

Acronyms

General

DEM	Digital Elevation Model
DTED L2	Digital Terrain Elevation Data Level 2
GMM	Gaussian Mixture Model
BIC	Bayesian Information Criterion
PCA	Principal Component Analysis
PC	Principal Component

Morphometric variables

ELEV	Absolute **Elev**ation
RESE	**Res**idual **E**levation
EBAS	**E**levation above erosional **bas**e
RELF	**Rel**ief
LTPI	**L**ocal **T**opographic **Pos**ition
RUGN	**Rug**ged**n**ess
SPOS	**S**lope **Pos**ition
FLAT	**Flat**ness
SNIS	**S**urface **Nois**e
MCON	**M**ean **Con**vergence

Land Surface Types

UME	**U**plands, medium **M**ountains, **E**levated
UMI	**U**plands, medium **M**ountains, **I**nner parts
UMV	**U**plands, **M**ountains, **V**alleys
UMS	**U**plands, low **M**ountains, **S**lopes/highlands
UHE	**U**plands, **H**ighlands, **E**levated
PRH	**P**lain, **R**olling, **H**igh density
PRMu	**P**lain, **R**olling, **M**edium density, **u**pper
PRMl	**P**lain, **R**olling, **M**edium density, **l**ower/dissected
PRLu	**P**lain, **R**olling, **L**ow density, **u**pper
PRLl	**P**lain, **R**olling, **L**ow density, **l**ower
PDV	**P**lain, **D**issection, **V**alleys
PDEu	**P**lain, **D**issection, **E**dges, **u**pper
PDEl	**P**lain, **D**issection, **E**dges, **l**ower
PSI	**P**lain, **S**mooth, **I**nclined
PSG	**P**lain, **S**mooth, **G**ently inclined
PSF	**P**lain, **S**mooth, near-**F**lat
PFR	**P**lain, near-**F**lat, **R**ough
PFD	**P**lain, near-**F**lat, **D**issections

Table 2
Compilation of Land Surface Types with the corresponding morphogenetic zones and landforms. For morphogenetic zones, see Fig. 1. Non-mountains zones represent all other morphogenetic zones except Mountains. Coverage shows the percentage of area covered by the given land surface type.

No.	Sym.	Coverage	Geomorphometric zones and typical landforms
11	UME	1.5 %	Mountains: ridges upper slopes
12	UMI	1 %	Mountains: inner parts of medium mountains, deeply incised valleys
13	UMV	2.6 %	Mountains: broad valleys and basins
14	UMS	5.8 %	Mountains: lower parts, slopes
15	UHE	4.5 %	Highlands: elevated areas
21	PRH	8.6 %	Coastal plains and Lakelands: Moraine plateaus of high diversity
22	PRMu	7.5 %	Lakelands: Moraine plateaus of medium diversity, upper parts. Denudated plains: edge zones of older moraine plateaus, upper parts of plateaus (limited extend)
23	PRMl	5.2 %	Coastal plains and Lakelands: Moraine plateaus of medium diversity, lower parts
24	PRLu	5.4 %	Coastal plains: lower part of plateaus, near coastal lowlands. Lakelands: upper terraces of broad river valleys
25	PRLl	7.9 %	Coastal plains: upper part of plateaus. Lakelands: upper terraces of broad river valleys
31	PDV	4 %	Lakelands: deeply incised narrow valleys. Denudated plains: dissected slopes of older moraine ramparts
32	PDEu	3.7 %	Non-mountains: valley slopes upper and medium parts
33	PDEl	2.9 %	Non-mountains: valley slopes lower parts and footslopes
41	PSI	6.8 %	Periglacial plains, Highlands, Forelands: upper parts of older moraine plateaus, slopes of narrow valleys
42	PSL	8.2 %	Periglacial plains, Highlands, Forelands: lower parts of older moraine plateaus
43	PSF	9.2 %	Non-mountains: elevated flat plateaus
51	PFR	8 %	Denudated plains, Highlands: flood terraces of broad river valleys, boggy plains. Denudated plains, outwash plains, pradolinas
52	PFD	2.1 %	Highlands and Forelands: deep valleys of rivers, gorges
53	PFSu	3.3 %	Denudated plains, uplands: upper parts of flood terraces of broad river valleys, boggy plains outwash plains, pradolinas
54	PFSl	1.9 %	Coastal plains and Lakelands: flat, low laying valley bottoms, near-coastline plains, deltas

PFSu **P**lain, near-**F**lat, **S**mooth, **u**pper
PFSl **P**lain, near-**F**lat, **S**mooth, **l**ower

CRediT authorship contribution statement

Krzysztof Dyba: Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Writing – original draft, Visualization, Funding acquisition. **Jarosław Jasiewicz:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

Authors declare no conflict of interests.

Acknowledgments

This research was funded in part by National Science Centre, Poland 2021/41/N/ST10/00347. For the purpose of Open Access, the author has applied a CC-BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission. The four sheets of the Digital Geomorphological Map of Poland used in this study were obtained from the Head Office of Geodesy and Cartography (license number: DIO.7211.342.2021_PL_N). Special thanks are owed to two anonymous reviewers for their very helpful and insightful comments.

Data and code availability

The code to reproduce the research results is available on GitHub under the MIT license: https://github.com/kadyb/geomorph_clustering. The final high-resolution maps with geomorphological units and uncertainty, Gaussian mixture model and data transformation model, and low-resolution rasters (500 m) with geomorphometric variables are available in the Zenodo repository under the CC BY 4.0 license: https://zenodo.org/record/6415362.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geomorph.2022.108373.

References

Adediran, A.O., Parcharidis, I., Poscolieri, M., Pavlopoulos, K., 2004. Computer-assisted discrimination of morphological units on north-Central Crete (Greece) by applying multivariate statistics to local relief gradients. Geomorphology 58, 357–370. https://doi.org/10.1016/j.geomorph.2003.07.024.
Badura, J., Jary, Z., Smalley, I., 2013. Sources of loess material for deposits in Poland and parts of central europe: the lost big river. Quat. Int. 296, 15–22.
Bengtsson, H., 2021. A unifying framework for parallel and distributed processing in R using futures. URL: R J. https://doi.org/10.32614/RJ-2021-048 https://journal.r-project.org/archive/2021/RJ-2021-048/index.html.
Beven, K., 1996. Equifinality and uncertainty in geomorphological modelling. The scientific nature of geomorphology. In: Proceeding of the 27th Binghamton Symposium in Geomorphology, pp. 289–314.
Böhner, J., Blaschke, T., Montanarella, L., 2008. SAGA – Seconds Out.. volume 19. Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie.
Brändli, M., 1996. Hierarchical models for the definition and extraction of terrain features. In: Geographic Objects With Indeterminate Boundaries. Taylor and Francis, London, pp. 257–270. https://doi.org/10.1201/9781003062660-22.
Brown, D., Lusch, D., Duda, K., 1998. Supervised classification of types of glaciated landscapes using digital elevation data. Geomorphology 21, 233–250. https://doi.org/10.1016/S0169-555X(97)00063-9.
Bue, B.D., Stepinski, T.F., 2006. Automated classification of landforms on Mars. Comput. Geosci. 32, 604–614. https://doi.org/10.1016/J.CAGEO.2005.09.004.
Burrough, P., van Gaans, P., MacMillan, R., 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets Syst. 113, 37–52. https://doi.org/10.1016/S0165-0114(99)00011-1.
Dan Capitan, R., Van De Wiel, M.J., 2012. Regional morphometric and geomorphologic mapping of Martian landforms. Comput. Geosci. 45, 190–198. https://doi.org/10.1016/J.CAGEO.2011.11.030.
Day, N.E., 1969. Estimating the components of a mixture of normal distributions. Biometrika 56, 463. https://doi.org/10.2307/2334652.
De Bruin, S., Stein, A., 1998. Soil-landscape modelling using fuzzy c-means clustering of attribute data derived from a Digital Elevation Model (DEM). Geoderma 83, 17–33. https://doi.org/10.1016/S0016-7061(97)00143-2.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. 39, 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
Dikau, R., 1989. The application of a digital relief model to landform analysis in geomorphology. In: Raper, J. (Ed.), Three Dimensional Applications in GIS, 1st ed. CRC Press, London, pp. 51–77. https://doi.org/10.1201/9781003069454-5.
Drăguţ, L., Blaschke, T., 2006. Automated classification of landform elements using object-based image analysis. Geomorphology 81, 330–344. https://doi.org/10.1016/j.geomorph.2006.04.013.
Drăguţ, L., Eisank, C., 2011. Object representations at multiple scales from digital elevation models. Geomorphology 129, 183–189. https://doi.org/10.1016/J.GEOMORPH.2011.03.003.
Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. ISPRS J. Photogramm. Remote Sens. 158, 63–75. https://doi.org/10.1016/j.isprsjprs.2019.09.018.
Dylik, J., 1969. Slope development under periglacial conditions. Biul. Peryglac. 18, 381–410.
Etzelmüller, B., Romstad, B., Fjellanger, J., 2007. Automatic regional classification of topography in Norway. Nor. Geol. Tidsskr. 87, 167–180.
Evans, I.S., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Spatial Analysis in Geomorphology, volume 6. Methuen, London, pp. 17–90. https://doi.org/10.4324/9780429273346-2.
Evans, I.S., 2012. Geomorphometry and landform mapping: what is a landform? Geomorphology 137, 94–106. https://doi.org/10.1016/j.geomorph.2010.09.029.
Fayyad, U., Uthurusamy, R., 1996. Data mining and knowledge discovery in databases. Commun. ACM 39, 24–26. https://doi.org/10.1145/240455.240463.
Fenneman, N.M., 1917. Physiographic subdivision of the United States. Proc. Natl. Acad. Sci. 3, 17–22. https://doi.org/10.1073/pnas.3.1.17.
Florinsky, I.V., 2017. An illustrated introduction to general geomorphometry. Prog. Phys. Geogr. 41, 723–752. https://doi.org/10.1177/0309133317733667.

Franklin, S.E., 2020. Interpretation and use of geomorphometry in remote sensing: a guide and review of integrated applications. Int. J. Remote Sens. 41, 7700–7733. https://doi.org/10.1080/01431161.2020.1792577.
Galon, R., 1972. G lówne etapy tworzenia sieę rzeźby Niżu Polskiego. In: Galon, R. (Ed.), Geomorfo- logia Polski. PWN, Warszawa, pp. 35–110.
Gilewska, S., 1991. Rzeźba. In: Starkel, L. (Ed.), Geografia Polski: Środowisko przyrodnicze. Państwowe Wydawnictwo Naukowe, Warszawa, pp. 248–296.
GRASS Development Team, 2020. Geographic Resources Analysis Support System (GRASS GIS) software, version 7.8. URL:. Open Source Geospatial Foundation http://grass.osgeo.org.
Grohmann, C., 2005. Trend-surface analysis of morphometric parameters: a case study in southeastern Brazil. Comput. Geosci. 31, 1007–1014. https://doi.org/10.1016/j.cageo.2005.02.011.
Haines-Young, R.H., Petch, J.R., 1983. Multiple working hypotheses: equifinality and the study of landforms. Trans. Inst. Br. Geogr. 8, 458–466. https://doi.org/10.2307/621962.
Hammond, E., 1954. Small-scale continental landform maps. Ann. Assoc. Am. 44, 33–42.
Hammond, E., 1964. Analysis of properties in land form geography: an application to broad-scale land form mapping. Ann. Assoc. Am. Geogr. 54, 11–19.
Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314.
Hengl, T., Rossiter, D.D.G., 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Sci. Soc. Am. J. 67, 1810–1822. https://doi.org/10.2136/sssaj2003.1810.
Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417.
Hughes, P., Woodward, J., 2009. Glacial and Periglacial environments. In: The Physical Geography of the Mediterranean. Oxford University Press, pp. 353–383. https://doi.org/10.1093/oso/9780199268030.003.0024.
Irvin, B., Ventura, S., Slater, B., 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77, 137–154. https://doi.org/10.1016/S0016-7061(97)00019-0.
Iwahashi, J., Pike, R.J., 2007. Automated classifications 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., Kamiya, I., Matsuoka, M., Yamazaki, D., 2018. Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. Prog Earth Planet Sci 1 (5), 1–31. https://doi.org/10.1186/S40645-017-0157-2, 2018 5.
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. https://doi.org/10.1186/s40645-020-00398-0.
Janowski, L., Tylmann, K., Trzcinska, K., Rudowski, S., Tegowski, J., 2022. Exploration of glacial landforms by object-based image analysis and spectral parameters of digital elevation model. IEEE Trans. Geosci. Remote Sens. 60, 4502817. https://doi.org/10.1109/TGRS.2021.3091771.
Jasiewicz, J., Metz, M., 2011. A new GRASS GIS toolkit for Hortonian analysis of drainage networks. Comput. Geosci. 37, 1162–1173. https://doi.org/10.1016/j.cageo.2011.03.003.
Jasiewicz, J., Stepinski, T.F., 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.
Jasiewicz, J., Netzel, P., Stepinski, T.F., 2014. Landscape similarity, retrieval, and machine mapping of physiographic units. Geomorphology 221, 104–112. https://doi.org/10.1016/j.geomorph.2014.06.011.
Jolliffe, I.T., 2002. Springer series in statistics. In: Principal Component Analysis, p. 29.
Klimaszewski, M., 1956. The principles of geomorphological survey of Poland. Prz. Geol. 28, 32–40.
Kondracki, J., 2002. Geografia regionalna Polski. PWN, Warszawa.
Kondracki, J., Rychling, A., 1994. Mapa 53.3. Regiony fizycznogeograficzne. In: Najgrakowski, M. (Ed.), Atlas Rzeczypospolitej Polskiej. Głòwny Geodeta Kraju, Warszawa, pp. 1993–1999.
Kozarski, S., 1986. Timescales and the rhythm of Vistulian geomorphic events in the polish lowland. Czasopismo Geograficzne 57, 247–270.
Kuhn, M., Wickham, H., 2021. recipes: preprocessing tools to create design matrices. URL: https://CRAN.R-project.org/package=recipes. r package version 0.1.16.
Lencewicz, S., 1937. Polska, volume 7. Nakl. Trzaski, Everta i Michalskiego.
Li, S., Xiong, L., Tang, G., Strobl, J., 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. Geomorphology 354, 107045. https://doi.org/10.1016/j.geomorph.2020.107045.
MacMillan, R., Pettapiece, W., Nolan, S., Goddard, T., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets Syst. 113, 81–109. https://doi.org/10.1016/S0165-0114(99)00014-7.
Mark, D.M., 1975. Geomorphometric parameters: a review and evaluation. Geogr. Ann. Ser. B 57, 165–177. https://doi.org/10.2307/520612.
Mark, D.M., Smith, B., 2004. A science of topography: from qualitative ontology to digital representations. In: Geographical Information Science and Mountain Geomorphology, pp. 75–100. July.
Marks, L., Ber, A., Gogo Lek, W., Piotrowska, K., 2006. Geological Map of Poland 1: 500000. Panǹstwowy Instytut Geologiczny, Warszawa.
Maxwell, A.E., Warner, T.A., 2015. Differentiating mine-reclaimed grasslands from spectrally similar land cover using terrain variables and object-based machine learning classification. Int. J. Remote Sens. 36, 4384–4410. https://doi.org/10.1080/01431161.2015.1083632.
Minar, J., Evans, I., 2008. Elementary forms for land surface segmentation: the theoretical basis of terrain analysis and geomorphological mapping. Geomorphology 95, 236–259. https://doi.org/10.1016/j.geomorph.2007.06.003.
Mojski, J.E., 1995. Pleistocene glacial events in Poland. In: Ehlers, J., Kozarski, S., Gibbard, P.L. (Eds.), Glacial Deposits in North-East Europe. A.A Balkema, Rotterdam, pp. 287–292.
Mojski, J.E., 2005. Ziemie polskie w czwartorzędzie: zarys morfogenezy. PiG, Warszawa.
Murton, J.B., 2021. What and where are periglacial landscapes? Permafr. Periglac. Process. 32, 186–212. https://doi.org/10.1002/ppp.2102.
Nir, D., 1957. The ratio of relative and absolute altitudes of Mt. Carmel: a contribution to the problem of relief analysis and relief classification. Geogr. Rev. 47, 564. https://doi.org/10.2307/211866.
Nowaczyk, B., 1995. The age of dunes in poland-selected problems. In: Quaestiones Geographicae. Zeszyt Specjalny, 4, pp. 233–239.
Patel, E., Kushwaha, D.S., 2020. Clustering cloud workloads: K-Means vs Gaussian mixture model. Procedia Comput. Sci. 171, 158–167. https://doi.org/10.1016/j.procs.2020.04.017.
Pebesma, E., 2018. Simple features for R: standardized support for Spatial Vector Data. R J. 10, 439–446. https://doi.org/10.32614/RJ-2018-009.
Pebesma, E., 2021. stars: spatiotemporal arrays, raster and vector data cubes. URL: https://CRAN.R-project.org/package=stars. r package version 0.5-3.
Peucker, T.K., Douglas, D.H., 1975. Detection of surface-specific points by local parallel processing of discrete terrain elevation data. Comput. Graphics Image Process. 4, 375–387. https://doi.org/10.1016/0146-664x(75)90005-2.
Pike, R.J., 1988. The geometric signature: quantifying landslide-terrain types from digital elevation models. Math. Geol. 20, 491–511. https://doi.org/10.1007/BF00890333.
Pike, R.J., Wilson, S.E., 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. Bull. Geol. Soc. Am. 82, 1079–1084. https://doi.org/10.1130/0016-7606(1971)82[1079:ERHIAG]2.0.CO;2.
Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T., 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology 78, 373–386. https://doi.org/10.1016/J.GEOMORPH.2006.02.005.
Punkari, M., 1997. Glacial and glaciouvial deposits in the interlobate areas of the Scandinavian Ice Sheet. Quat. Sci. Rev. 741–753. https://doi.org/10.1016/S0277-3791(97)00020-6.
Qiu, D., 2010. A comparative study of the K-means algorithm and the normal mixture model for clustering: bivariate homoscedastic case. J. Stat. Plann. Inference 140, 1701–1711. https://doi.org/10.1016/j.jspi.2009.12.025.
R Core Team, 2021. R: a language and environment for statistical computing. URL. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rączkowska, Z., Zwoliński, Z., 2015. Digital geomorphological map of Poland. Geogr. Pol. 88, 205–210. https://doi.org/10.7163/GPol.0025.
Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M.J., 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.
Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A Terrain Ruggedness Index That Qauntifies Topographic Heterogeneity.
Rotnicki, K., 1974. Slope development of Riss Glaciation end moraines during the Würm; its morphological and geological consequences. Questiones Geographicae 1, 109–139.
Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121, 243–256. https://doi.org/10.1016/j.geoderma.2003.10.008.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 461–464.
Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8, 289–317.
Shary, P.A., 1995. Land surface in gravity points classification by a complete system of curvatures. Math. Geol. 27, 373–390. https://doi.org/10.1007/BF02084608.
Shumack, S., Hesse, P., Farebrother, W., 2020. Deep learning for dune pattern mapping with the AW3D30 global surface model. Earth Surf. Process. Landf. 45, 2417–2431. https://doi.org/10.1002/esp.4888.
Solon, J., Borzyszkowski, J., Bid Lasik, M., Richling, A., Badora, K., Balon, J., Brzezińńska-Wójcik, T., Chabudziński, L., Dobrowolski, R., Grzegorczyk, I., Jodłowski, M., Kistowski, M., Kot, R., Krąż, P., Lechnio, J., Macias, A., Majchrowska, A., Malinowska, E., Migoń, P., Myga-Piątek, U., Nita, J., Papińska, E., Rodzik, J., Strzyż, M., Ziaja, W., Terpi Lowski, S., 2018. Physico-geographical mesoregions of poland: verification and adjustment of boundaries on the basis of contemporary spatial data. Geogr. Pol. 91, 143–170. https://doi.org/10.7163/GPol.0115.
Starkel, L., 1980. Przeglądowa mapa geomorfologiczna Polski w skali 1:500 000.
Starkel, L., 2001. Historia doliny Wisły: od ostatniego zlodowacenia do dziś. PAN, Warszawa.
Stepinski, T.F., Ghosh, S., Vilalta, R., 2006. In: Automatic Recognition of Landforms on Mars Using Terrain Segmentation and Classification. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4265 LNAI, pp. 255–266. https://doi.org/10.1007/11893318_26.
Stepinski, T., Vilalta, R., Ghosh, S., 2007. Machine learning tools for automatic mapping of Martian landforms. IEEE Intell. Syst. 22, 100–106. https://doi.org/10.1109/MIS.2007.114.
Szuman, I., Kalita, J.Z., Ewertowski, M.W., Clark, C.D., Livingstone, S.J., 2021. Dynamics of the last Scandinavian Ice Sheet's southernmost sector revealed by the pattern of ice streams. Boreas 50, 764–780. https://doi.org/10.1111/bor.12512.

Szymanowski, M., Jancewicz, K., Róycka, M., Migoń, P., 2019. Geomorphometry-based detection of enhanced erosional signal in polygenetic medium-altitude mountain relief and its tectonic interpretation, the Sudetes (Central Europe). Geomorphology 341, 115–129. https://doi.org/10.1016/j.geomorph.2019.05.022.

Szypuła, B., Wieczorek, M., 2020. Geomorphometric relief classification with the k-median method in the Silesian Upland, southern Poland. Front. Earth Sci. 14, 152–170. https://doi.org/10.1007/s11707-019-0765-9.

Torres, R.N., Fraternali, P., Milani, F., Frajberg, D., 2019. Mountain summit detection with Deep Learning: evaluation and comparison with heuristic methods. Appl. Geomatics. https://doi.org/10.1007/s12518-019-00295-2.

Vaze, J., Teng, J., Spencer, G., 2010. Impact of DEM accuracy and resolution on topographic indices. Environ. Model. Softw. 25, 1086–1098. https://doi.org/10.1016/j.envsoft.2010.03.014.

Wang, J., Cheng, W., Zhou, C., Zheng, X., 2017. Automatic mapping of lunar landforms using DEM-derived geomorphometric parameters. J. Geogr. Sci. 27, 1413–1427. https://doi.org/10.1007/s11442-017-1443-z.

Weiss, A.D., 2001. Topographic position and landforms analysis. In: Poster Presentation, ESRI User Conference, San Diego, CA, p. 200.

Wieczorek, M., Migoń, P., 2014. Automatic relief classification versus expert and field based landform classification for the medium-altitude mountain range, the Sudetes, SW Poland. Geomorphology 206, 133–146. https://doi.org/10.1016/J.GEOMORPH.2013.10.005.

Wood, J.J., 1996. The Geomorphological Characterisation of Digital Elevation Models. University of Leicester, UK. Ph.D. thesis.

Wood, W.F., Snell, J.B., 1960. A quantitative system for classifying landforms. In: Technical Report. Headquarters, Quartermaster Research and Engineering Command, US Army, Natick, MA.

Xie, Z., Haritashya, U.K., Asari, V.K., Young, B.W., Bishop, M.P., Kargel, J.S., 2020. GlacierNet: a deep-learning approach for debris-covered glacier mapping. IEEE Access 8, 83495–83510. https://doi.org/10.1109/ACCESS.2020.2991187.