Adam Mickiewicz University
Doctoral School of Natural Sciences
Faculty of Geographical and Geological Sciences

DISSERTATION

Krzysztof Dyba

Interpretation of geospatial data using explainable machine learning methods

Supervisor
Prof. UAM dr hab. Jarosław Jasiewicz

Assistant Supervisor
Prof. UAM dr hab. inż. Cezary Kaźmierowski

Poznań, 2024

Acknowledgments

I thank my parents Mirosław and Irena for their support and motivation in carrying out my scientific work.

I thank my supervisors Professor Jarosław Jasiewicz and Professor Cezary Kaźmierowski for their time, inspiration for scientific research, assistance and critical comments.

I thank my colleagues Prof. Jan Piekarczyk, Prof. Sławomir Królewicz, Prof. Jakub Nowosad and Dr. Jakub Ceglarek for cooperation, discussions and kindness. In addition, I thank Dr. Jakub Ceglarek for proofreading, which significantly improved the dissertation.

Table of contents

1. Introduction

In scientific research, the transparency of the data processing is crucial, both for the interpretation of the results and the comparability of the results achieved using different methods. Numerical methods used in geographic information systems are mainly based on geoprocessing algorithms. If the sequence of algorithms is well-documented – from the acquisition of the data to the obtaining of the result, the analytical process is transparent to the user; thus allowing for repetition, but also for critical analysis. This approach reflects the knowledge of the researcher, understanding of the input data and knowing the analytical procedures that lead to the result. However, models created solely on the basis of expert knowledge are usually characterized by a certain approximation of the modeled process, i.e. they are its simplification.

The rapid development of geoinformatics technologies is closely linked to the tremendous growth of spatial data, which is a mutual process. Data growth requires new efficient and effective analytical techniques, which contributes to the increased demand for data and investment in new sensors. These data are acquired from various sensors, including multispectral imagery and geographical coordinates from satellites, point clouds from laser scanning, as well as field surveys and digitization of archival data (Chiang et al., 2014; Gotway & Young, 2002). In the face of such large datasets, the expert-based methods used so far turn out to be insufficient, especially for complex and non-linear problems occurring in the natural sciences. For this reason, researchers are turning to the use of advanced machine learning methods to analyze spatial data (Bergen et al., 2019; Casali et al., 2022; Du et al., 2020; Karpatne et al., 2019; Lary et al., 2016; Nikparvar & Thill, 2021).

Machine learning is a subfield of artificial intelligence that involves creating regression or classification models based on self-learning algorithms and training data. The main goal of machine learning is the development of models that effectively recognize patterns and relationships on new datasets, resulting in easier and more automated work. The machine learning process is intended to produce useful tools whose effectiveness has been verified on an independent test set. According to the nature of the learning process, the following can be distinguished: 1) supervised learning (depending on the type of predicted variable, it involves classification or regression); 2) unsupervised learning. The main difference between supervised and unsupervised classification is the *a priori* presence of the dependent (reference) variable in supervised classification, while unsupervised classification uses only independent variables (features) without any assumptions about the expected outcome. It also requires an additional step in the *a posteriori* analytical process, in which the groups resulting from clustering are interpreted and then labeled.

Another division is related to the interpretability of the models and two types can be distinguished, i.e. black-box and white-box models. Black-box models such as neural networks (Anderson, 1995), random forests (Breiman, 2001), and support vector machines (Cortes & Vapnik, 1995) have high prediction accuracy, but their rationale is not overt. In contrast, white-box models such as linear regression (Seber & Lee, 2003), logistic regression (Menard, 2008), decision trees (Quinlan, 1986), generalized linear models (Nelder & Wedderburn, 1972) have lower accuracy due to the greater requirements of feature engineering, but the mechanisms behind their operation are understood through explicit interactions between the explanatory variables and the dependent variable (Figure 1). These interactions are defined by the values of regression coefficients in linear models or the threshold values of divisions in decision-regression trees. Currently, in research practice, black-box models are increasingly being used due to their high efficiency.

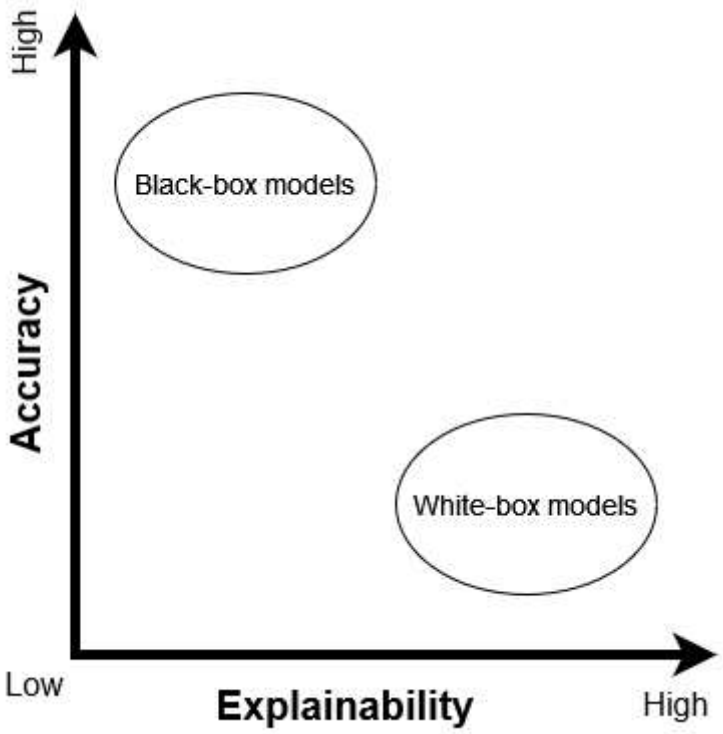

Figure 1. Relationship between accuracy and explainability of white-box and black-box models.

Naturally, in the case of black-box models, questions arise: Which factors led to the outcome and how did they affect it? Emerging work related to explainable machine learning attempts to answer these questions (Belle & Papantonis, 2021; Biecek & Burzykowski, 2021; Linardatos et al., 2020; Molnar, 2022; Roscher et al., 2020) by providing indirect methods that approximate the

operation of these models, making the process of creating decision rules more understandable to the user (Figure 2).

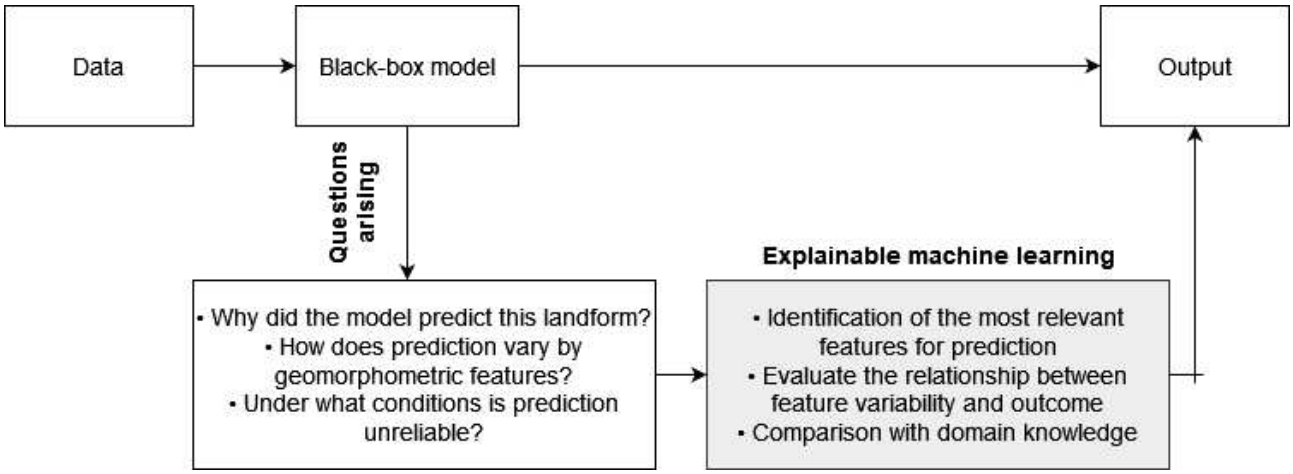

Figure 2. The role of explainable machine learning in black-box model analysis.

A current research gap is the lack of interpretation of the impact of explanatory variables (for example, spectral bands or geomorphometric variables) on the results of modeling spatial data using machine learning. Another major issue is the effectiveness of such solutions and, more specifically, the compatibility with domain knowledge – does the model consider factors in the same way as experts, or are decision rules constructed in a completely different way? The modeling result must be reproducible, useful and understandable to potential users (Fayyad et al., 1996). It can be hypothesized that interpretable machine learning can be a tool to support the process of interpreting spatial data.

Despite a significant number of publications using machine learning to solve problems in remote sensing (Chang & Bai, 2018; Hänsch et al., 2018; Maxwell et al., 2018), soil science (Heung et al., 2016; Heuvelink et al., 2021; Padarian et al., 2020; Wadoux et al., 2020), archaeology (Bickler, 2021; Cacciari & Pocobelli, 2022), or ecology (Crisci et al., 2012; Lorena et al., 2011; Rubbens et al., 2023; Shan et al., 2006; Stupariu et al., 2022), the interpretive approach in machine learning have not been widely used. Moreover, there is an overload of studies that only compare the effectiveness of complex models without understanding how they actually work (Linardatos et al., 2020).

Nevertheless, researchers are paying attention to the problem of interpretability of the modeling process in Earth and environmental sciences. The first work is considered to be the visualization of regression trees on a map for the problem of bird species richness (White & Sifneos, 2002). However, in subsequent years, greater emphasis was placed on research that included: the analysis of factors influencing settlement (Jasiewicz & Hildebrandt-Radke, 2009),

analysis of the relationship between human activities in the past and the natural environment (Jasiewicz & Sobkowiak-Tabaka, 2015), analysis of the usefulness of meteorological data for predicting high pollen concentrations (Nowosad, 2016), analysis of the influence of landscape on the attractiveness of a residence (Dąbrowski, 2018), analysis of the usefulness of meteorological and remote sensing data for monitoring phenological phases of plants (Czernecki et al., 2018), or analysis of the environmental variables to determine indicators of the trophic state of lakes (Jasiewicz et al., 2022; Zawiska et al., 2023). The current state of research requires intensified work on the interpretive approach in machine learning, to answer the question whether the discussed methods will actually bring benefits in the process of acquiring new knowledge.

1.1 Objective

The purpose of this dissertation is to reveal whether explainable machine learning methods can be a tool supporting the process of modeling geospatial data, where the measure of success is not training an effective model, but the ability to obtain additional knowledge about the factors leading to the result. Thus, the aim of the dissertation shifts the issue of exclusively assessing the accuracy of machine learning models towards understanding their operation and connection with other data processing methods.

The above objective was achieved through three research experiments covering various machine learning paradigms, including regression analysis, supervised classification and unsupervised classification, and selected explainable machine learning methods for global model interpretation (Figure 3):

1) **Dyba, K.**, Ermida, S., Ptak, M., Piekarczyk, J., & Sojka, M. (2022). Evaluation of methods for estimating lake surface water temperature using Landsat 8. *Remote Sensing*, *14*(15), 3839. https://doi.org/10.3390/rs14153839
2) **Dyba, K.**, & Jasiewicz, J. (2022). Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland). *Geomorphology*, *413*, 108373. https://doi.org/10.1016/j.geomorph.2022.108373
3) **Dyba, K.** (2024). Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland. *Scientific Reports*, *14*(1), 5447. https://doi.org/10.1038/s41598-024-56066-6

Article	Issue	Materials	Methods	Result
Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8	Regression analysis	• Satellite imagery (including thermal bands) • In-situ temperature measurements	• Multiple linear regression • Random forest • Variable importance plot	Comparison of model performance and identification of crucial explanatory variables
Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland	Supervised classification	• Digital elevation model • Geomophometric variables • Digital geomorphological map	• Random forests • Gradient boosting • Convolutional neural network • Variable importance plot • Accumulated local effects	Understanding how the geomorphometric variables used by the black-box model affect landform classification
Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland)	Unsupervised classification	• Digital elevation model • Geomophometric variables	• Gaussian mixture models • Bayesian information criterion • Principal component analysis • Interpretation of clusters	Explanation of *a posteriori* classification principles for the clustering results of geomorphometric variables

Figure 3. Summary of articles comprising the dissertation. The second and third studies were conducted as part of the "*The use of computer vision for geomorphological mapping*" project funded by Polish National Science Centre (2021/41/N/ST10/00347, Principal investigator: Krzysztof Dyba).

2. Materials and methods

Machine learning encompasses a diverse range of models, each designed to solve specific types of problems and data characteristics. There are many algorithms, each with its own strengths, weaknesses and applications (Sarker, 2021). Moreover, progress in research is constantly leading to the development of existing models as well as the creation of new ones, further expanding the capabilities of machine learning.

Multiple linear regression is an extension of simple linear regression and is used to analyze the relationship between many independent variables (features) and a single dependent (modeled) variable. This model assumes that the relationship between the dependent variable and each independent variable is linear. Other assumptions concern a constant variance of the error terms for independent variables (homoscedasticity), a distribution of residuals close to the normal distribution and the lack of multicollinearity of predictors. The model is defined by the following formula:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \quad (1)$$

where:

Y is the dependent variable.

$X_1, X_2, ..., X_n$ are the independent variables.

$\beta_1, \beta_2, ..., \beta_n$ are the coefficients (slopes) associated with each independent variable.

ϵ is the error, representing the difference between the actual and predicted values of Y (includes the effect of all unexplained factors on Y).

While linear regression is exclusively used for regression, random forests are also used for classification (Breiman, 2001). Its operation is based on creating many regression or classification trees. In the case of classification, the final prediction is the class that receives the most votes among all trees in a majority vote. For regression, the predictions of individual trees are averaged. Tree construction uses a technique called bagging (bootstrap aggregating), so that each tree is trained on a different subset of data using random sampling with replacement. In addition, a random subset of the features is drawn during tree construction, so that the trees are not correlated with each other. This approach helps reduce the variance of the model, making it less prone to overfitting.

XGBoost (Chen & Guestrin, 2016) and LightGBM (Ke et al., 2017) are currently state-of-the-art algorithms using gradient boosting. While random forests reduce the variance of a large number of complex trees, gradient boosting reduces the error of simple trees with low variance in a sequential manner (this is known as weak learner). Each tree is trained on the prediction errors from

the previous tree, focusing on its improvement in subsequent iterations. The main difference between these algorithms is how the trees are built – XGBoost uses a level-wise (horizontally) tree growth strategy, while LightGBM utilizes a leaf-wise (vertically) strategy. The leaf-wise approach allows tree growth to be stopped earlier when sufficient improvement has been achieved, thereby increasing the speed of computing. However, it can increase the risk of overfitting.

U-Net is a convolutional neural network designed for semantic segmentation (Ronneberger et al., 2015). That is, it is trained to classify each image pixel into a specific category. Basically, the U-Net architecture consists of two paths, i.e. a contracting path (encoder) and an expansive path (decoder). The former extracts features from the input image while reducing its spatial resolution, whereas the latter does the opposite, i.e. upsamples the feature maps back to the original input resolution while incorporating information from the encoding path to enhance the segmentation results.

There are two main divisions of explainability methods for machine learning models (Guidotti et al., 2019; Molnar, 2022), i.e. the first one according to the level (local or global) and the second one according to the inclusiveness (agnostic and specific models). Agnostic methods can be applied to any machine learning model because they are independent of the model structure. Otherwise, specific models are used that are tailored to particular types of models or architectures (e.g. only for decision trees). The global approach explains the general functioning of the model, the relevance of the explanatory variables and the interactions between them, while the local approach explains how the model makes a prediction for a specific example, i.e. how it works in given conditions.

The most popular global methods include *partial dependence plot* (Greenwell et al., 2018), *accumulated local effect* (Apley & Zhu, 2020), and *permutation feature importance* (Breiman, 2001; Fisher et al., 2018). While to local methods *individual conditional expectation* (Goldstein et al., 2015), *local interpretable model-agnostic explanations* (Ribeiro et al., 2016), *anchors* (Ribeiro et al., 2018), *shapley values (Shapley, 1953;* Štrumbelj & Kononenko, 2014), and *counterfactual explanations* (Dandl et al., 2020; Wachter et al., 2017). This dissertation focuses on the use of global interpretations, since the local approach (i.e., explaining individual cases) has already been examined by Jasiewicz et al. (2022) and Zawiska et al. (2023). Moreover, for huge datasets, such as multi-temporal satellite imagery with billions of cells (observations), a local interpretation approach is not justified.

The first experiment employed thermal optical images from the Landsat 8 satellite (National Aeronautics and Space Administration, United States Geological Survey) to model the surface temperature of lakes. Archived in-situ measurements conducted by the Institute of Meteorology and

Water Management for 38 lakes from 2013 to 2020 were used as reference data. Both datasets are publicly available (the former was accessed from https://earthengine.google.com/ and the latter from https://danepubliczne.imgw.pl/). To compare the effectiveness of prediction and interpretability of the modeling results, two different regression models were developed. The first was a simple multiple regression model representing a white-box model, and the second was a regression tree model based on random forests representing a black-box model. The *t-statistic* and Pearson's linear correlation coefficient were used to determine the significance of the predictors for the linear model, while the *impurity* algorithm was used for random forests.

The supervised classification experiment employed available sheets of the Digital Geomorphological Map of Poland at a scale of 1:100,000 representing different morphogenetic zones (Rączkowska & Zwoliński, 2015), including the coastal area (Świnoujście), the young and old glacial areas (Toruń and Kutno), the upland areas (Katowice, Kraków Zachodni and Tomaszów Lubelski), and the areas of young and old mountains (Jelenia Góra and Nowy Targ), while post-processed Digital Terrain Elevation Data Level 2 (DTED L2) was used as a source of topographic data (Jasiewicz et al., 2014). In the process of exploratory data analysis, the suitability of potential 15 geomorphometric variables for landform classification was considered using model performance metrics, visual inspection of the results, and model information gain metrics. In addition, Pearson's linear correlation between explanatory variables was examined to minimize redundancy. Above all, an attempt was made to explain how the geomorphometric variables used affect the model's classification decisions. For this purpose, accumulated local effects (Apley & Zhu, 2020) were used to assess the relationship between model features and prediction. The impact of a feature on the prediction is analyzed, while keeping all other features constant. The result comes from aggregating values over small feature intervals, showing how changing a specific feature affects the model's prediction on average.

In the last experiment, unsupervised classification (clustering) was used to analyze the diversity of land surface types and classification uncertainty based on the same digital terrain model employed in the previous experiment. Expert knowledge was used to appropriately select geomorphometric variables focusing on reflecting the vertical and horizontal variability of lowland areas, and visual inspection was used for verification. Some of the geomorphometric variables were highly correlated, conveying the same information, so as in the previous experiment, preselection of variables was applied using Pearson's linear correlation coefficient. For this purpose, the Bayesian information criterion was used to reduce overly complex models with a large number of clusters. The final stage of the analysis was related to the interpretation of the results, in which principal components analysis was used to identify and name the clusters.

3. Summary of articles

3.1. Regression analysis

The experiment presented in the article *„Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8”* compared four models for estimating water surface temperature based on satellite thermal data, considering two atmospheric models and two empirical models (multiple linear regression and random forests). The first empirical model represented the white-box type, while the second represented the black-box type. The best results were achieved for the random forests model, while systematic errors were found for atmospheric models in the coastal zone.

In the context of analyzing the influence of explanatory variables, 12 variables were evaluated considering 9 spectral bands, 2 spectral indices (Normalized Difference Vegetation Index and Normalized Difference Water Index) and the month of the year. Thermal band B10 (10.6 - 11.19 μm) was found to result in a smaller prediction error of surface water temperature than band B11 (11.5 - 12.51 μm). This may be related to the calibration of the thermal infrared sensor, which is susceptible to scattered light, which may cause measurement instability. Another important variable was the month of the year due to the seasonality of water temperature (water temperature is significantly higher in summer than in spring and autumn months). There was also a slight effect of the green band B3 (0.53 - 0.59 μm), which may be related to the depth of penetration in this range of electromagnetic radiation depending on water clarity. The remaining spectral variables were not found to be significant. In addition, the relationship between the magnitude of the prediction error and morphometric parameters of the lakes (area, volume and average depth) was analyzed, but no significant relationship was found.

Although the random forests model provided greater efficiency, its interpretation is limited to identifying the role of variables only, without indicating the relationship how they influenced the modeled value. This gap is filled by the linear model, as it precisely defines the relationship between the dependent variable and the explanatory variables:

$$WST = 2.9 * B10 - 2.07 * B11 + 48.48 \quad (2)$$

where:

WST is Water Surface Temperature in Kelvin.

B10, B11 are thermal bands in Kelvin.

The biggest advantage of the linear model is its mathematical simplicity, since it consists of a simple equation in which the regression coefficients determine how the prediction was made. In addition, from a technical point of view, greater simplicity is associated with lower memory requirements and greater computational performance compared to the random forests model.

The study fills a research gap in the topic of using satellite thermal data to monitor lake temperatures in Poland. This issue is poorly understood, and the presented article is the first such extensive study on modeling the surface temperature of lakes in Poland.

3.2. Supervised classification

The experiment presented in the article „*Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland*” included a comparison of three algorithms based on machine learning (random forests and gradient boosting) and deep learning (convolutional neural networks) for automatic geomorphological mapping. The best result was achieved for the XGBoost model relying on gradient boosting, while the lowest result was recorded for convolutional neural networks represented by the U-Net model. The main reason for the difference in the performance of these models was the relatively small input dataset and the unbalanced representativeness of landforms. Deep learning models have hundreds of thousands of parameters to tune compared to machine learning models, which typically have just a few and therefore require much more input data.

The interpretation of the decisions made by the most effective XGBoost model consisted of two steps: 1) identification of the most relevant geomorphometric variables for classification using the information gain metric; 2) detailed examination of the influence of explanatory variables on the classification result.

Regarding the first stage, the most useful were entropy, aggregated median elevation in the 1000 and 500 m cell, absolute elevation, standard deviation of elevation, surface convexity, topographic openness, slope and multi-scale topographic position index. The convergence index, surface texture, tangential and profile curvatures and multidirectional hillshade turned out to be less significant. The exclusion of less significant explanatory variables allowed the model to be simplified in terms of interpreting the results, as well as to be computationally faster.

In the second stage of the analysis, the non-linear effect of individual geomorphometric variables on the probability of classifying geomorphological units was examined by applying accumulated local effects. To the best of the author's knowledge, the referred method has not previously been used to interpret the results of automatic mapping in geomorphology. The analysis was conducted for all landforms for the morphogenetic zones considered. Furthermore, it was investigated how the area of geomorphological units affects the probability of correct classification of these units. The existence of such a positive relationship was confirmed, and it was found that the threshold value for the dataset used is approximately 45 km^2. This means that above this threshold, the magnitude of the influence of geomorphometric variables became significant. Otherwise, the impact was negligible.

Another notable observation is the varying impact of geomorphometric variables on the classification depending on the morphogenetic zone. The probable reason is the morphogenetic

diversity of the studied areas and different levels of geomorphological features, thus geomorphometric variables are useful to varying extents for characterizing the landforms occurring there, depending on the research area. This means that while the slope surface can be easily classified in a young glacial area using the slope, this may not be possible in an area of young mountains due to the completely different structure and characteristics of this morphogenetic zone. This also indicates that the same landforms in various morphogenetic areas can be characterized differently by the same geomorphometric variables.

Ultimately, accumulated local effects allowed for a thorough examination of how the explanatory variables affect the model's classification decisions, which helped interpret the functioning of the black-box model.

3.3. Unsupervised classification

While supervised classification in the modeling process requires reference data, in contrast, unsupervised classification has no such requirement. The clustering process uses the input data to discover patterns or clusters within it. The experiment described in the previous subsection focused on the best possible representation of reference data derived from the Digital Geomorphological Map of Poland, while the experiment conducted in the article „*Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland)*" was based primarily on the procedure for interpreting the obtained clusters. Iwahashi et al. (2021) noted that research on cluster analysis in geomorphological studies is mainly based on comparing clustering results with other existing reference maps, which limits unsupervised methods to only a subset of supervised classification. This experiment proposes a different approach based on discovering new knowledge in the data by finding useful patterns (Fayyad et al., 1996). In other words, the verification of the result of the analysis was not mainly the consistency with other geomorphological maps, but mainly the possibility of interpreting the results using a semi-formal procedure. The matching of the spatial patterns of the obtained clusters with the physical-geographic division of the country confirms the usefulness of this method.

One of the crucial issues in unsupervised classification is choosing the right number of clusters to produce meaningful and interpretable results. However, determining the optimal number of clusters is a difficult challenge, as it requires a balance between model complexity and the ability to reflect useful patterns. The choice depends primarily on the structure of the input data - selecting too few clusters may lead to the merger of separate heterogeneous groups, while too many clusters make them impossible to interpret. The number of clusters considered ranged from 4 to 24, and they were compared with the extent of morphogenetic zones and the spatial pattern of geomorphometric variables. The Bayesian information criterion indicated 20 clusters as the optimal choice. A larger number of clusters was questionable in terms of homogeneity and caused significant problems in interpretation, while a smaller number did not reveal important landforms such as bottoms of river valley.

The most important stage in unsupervised classification is the interpretation of clusters, i.e. finding the answer to the question of what they actually represent. This stage involves discovering the classification rules that cause the creation of individual land surface divisions. The aforementioned procedure was based mainly on principal component analysis supported by analysis of spatial patterns of individual geomorphometric variables. The first three principal components explained almost 70% of the data variability, including the first principal component (39.7%) and

was negatively related to absolute elevation, roughness and elevation above erosional base, while it was positively related to the terrain flatness. The second principal component explained 15.9% of the data variability and was negatively related to variables representing textural properties including mean convergence, surface noise, local topographic position and slope position. The third principal component represented 14.1% of the variance and was positively related to textural properties (mean convergence, surface noise) and negatively to the local topographic position. Additionally, four geomorphometric variables were used for interpretation, i.e. mean convergence, flatness, slope position, and relief, which had the greatest potential to distinguish landforms in lowland areas. Finally, 15 lowland units in 4 subgroups and 5 upland units were described. The obtained results showed strong convergence with the physical-geographic division on the scale of mesoregions. The developed map of land surface types on a national scale is a valuable product and can be used as a supplement to existing geomorphological studies.

4. Discussion

Explainable machine learning is a new but rapidly developing field. For this reason, it is not yet well established in science. There are currently ongoing discussions on the scope (Arrieta et al., 2020) and the distinction between the definitions of model explainability and interpretability, which were previously used as synonyms. Broniatowski (2021) proposes to distinguish these terms based on experimental psychology, where interpretation refers to the ability to contextualize the model results by linking them to the designed functional purpose of the system and user preferences, while explanation refers to the ability to precisely describe the mechanism or implementation of the model that led to the result. A simpler division is proposed by Rudin (2019), in which interpretation refers to white-box models, while explanation refers to black-box models as an *a posteriori* process, which requires the creation of a second model to explain the modeling method.

Therefore, the question arises about the validity of using black-box models in the context of explaining their operation. White-box models are inherently interpretable, which allows to accurately reproduce their operation and understand how the result was obtained. Some researchers recommend avoiding black-box models in favor of white-box models due to their transparency, and point out problematic issues with *a posteriori* explanations considering: 1) the explanation is only an approximation of what the model calculates; 2) the explanation may be an oversimplification of a complex model decision; 3) the explanation is susceptible to errors in the input data (Rudin, 2019). Moreover, the researchers point out that the performance of white-box models can be similar to black-box models, provided the explanatory variables are properly selected and prepared (Affenzeller et al., 2020; Loyola-Gonzalez, 2019). Further issues related to the concept of explainable machine learning itself are raised by Watson (2022), who highlights three current problems: 1) ambiguity (the impact of pragmatic assumptions such as context, level of abstraction, or purpose of analysis); 2) lack of rigorous testing; and 3) emphasis on the product instead of the iterative process.

Nevertheless, in light of the results obtained, we can point to the relevance of using explanatory machine learning methods to interpret spatial data, resulting in increased knowledge about geospatial processes. The first study identified the most important explanatory variables for the estimation of water surface temperature using remote sensing data, with the main limitation being the lack of consideration of how the examined variables affect the temperature prediction, i.e. we do not know neither the direction (positive or negative influence) nor the course of change (for which range of feature values it is increasing, and for which it is decreasing). Information on the significance of the explanatory variables alone is not sufficient, as it does not make it clear which

ranges of variable values are significant. However, the subsequent study on automated geomorphological mapping provided a detailed understanding of how geomorphometric variables affect the probability of landform classification by applying the accumulated local effects method. The method used revealed interactions between the model's prediction and the explanatory variables, showing the ranges of values to which the model responded and in what way. A separate, and very important, issue is the interpretation of unsupervised classification results often overlooked in research work. The limitations are related to the subjective selection of the optimal number of clusters without statistical justification, and ending the interpretation of clustering results on labeling the clusters with the letters of the alphabet and presenting their characteristics with descriptive statistics (minimum, maximum, mean, standard deviation). In the case of huge spatial datasets consisting of numerous objects, features and clusters, this approach makes interpretation almost impossible, thus preventing valuable conclusions from the analysis. The third study addresses the aforementioned problems by using the Bayesian information criterion to select the optimal number of clusters in a formalized way (nevertheless, this is one of many metrics that can be used for this purpose) and principal component analysis to interpret and finally label 20 clusters based on 10 complex geomorphometric variables. This approach facilitated understanding the clustering results of a large dataset covering the entire country, which would be more challenging in the case of descriptive statistics.

5. Conclusion

Based on the obtained results, we can conclude that the explainable methods employed to analyze spatial data provide added value in the context of interpreting black-box models used to solve classification and regression problems. They enable a better understanding of the decisions made by the models and examination of the impact of the features used on the prediction, which thus contributes to increasing knowledge about the existing patterns and relationships in spatial processes. The involvement of interpretation techniques also increases the transparency and accessibility of the modeling process, making audiences more likely to trust the results. Moreover, the way the predictions are explained can be confronted with domain knowledge and expert expectations, which allows to detect bias or incorrect model assumptions. In view of this, it is possible to improve the model, which may ultimately improve the accuracy of predictions.

Nevertheless, even despite the prominent benefits of explaining the behavior of machine learning models, it is necessary to take into account the limitations that are primarily related to the input data used to train the model. Data errors, outliers, unreliable representativeness or inappropriate selection of explanatory variables may lead to incorrect interpretations and conclusions. Another source of problems may be the model itself, i.e. using an inappropriate model for the task, resulting in low prediction accuracy and low sensitivity to features. We can infer that insufficient model performance will result in incorrect interpretations. Therefore, the quality of the input data and the accuracy of the model should be carefully considered before drawing conclusions.

The field of explainable artificial intelligence is new and has not yet found wide application in Earth and environmental sciences. Geospatial data, due to their specificity (spatial context), pose an additional challenge in interpretation resulting from spatial autocorrelation or spatial heterogeneity, which significantly complicates the modeling and interpretation process. As a final conclusion, the author recommends wider use of explanatory methods in research using machine learning and further development of methods for explaining model decisions, taking into account the spatial context.

6. Bibliography

1) Affenzeller, M., Burlacu, B., Dorfer, V., Dorl, S., Halmerbauer, G., Königswieser, T., Kommenda, M., Vetter, J., & Winkler, S. (2020). White Box vs. Black Box Modeling: On the Performance of Deep Learning, Random Forests, and Symbolic Regression in Solving Regression Problems. In R. Moreno-Díaz, F. Pichler, & A. Quesada-Arencibia (Eds.), *Computer Aided Systems Theory – EUROCAST 2019* (Vol. 12013, pp. 288–295). Springer International Publishing. https://doi.org/10.1007/978-3-030-45093-9_35
2) Anderson, J. A. (1995). *An introduction to neural networks*. MIT press.
3) Apley, D. W., & Zhu, J. (2020). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377
4) Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
5) Belle, V., & Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data*, *4*, 688969. https://doi.org/10.3389/fdata.2021.688969
6) Bergen, K. J., Johnson, P. A., De Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433), eaau0323. https://doi.org/10.1126/science.aau0323
7) Bickler, S. H. (2021). Machine Learning Arrives in Archaeology. *Advances in Archaeological Practice*, *9*(2), 186–191. https://doi.org/10.1017/aap.2021.6
8) Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: Explore, explain, and examine predictive models* (1st ed.). CRC Press.
9) Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
10) Broniatowski, D. A. (2021). *Psychological foundations of explainability and interpretability in artificial intelligence* (NIST IR 8367; p. NIST IR 8367). National Institute of Standards and Technology (U.S.). https://doi.org/10.6028/NIST.IR.8367
11) Cacciari, I., & Pocobelli, G. F. (2022). Machine Learning: A Novel Tool for Archaeology. In S. D'Amico & V. Venuti (Eds.), *Handbook of Cultural Heritage Analysis* (pp. 961–1002). Springer International Publishing. https://doi.org/10.1007/978-3-030-60016-7_33

12) Casali, Y., Aydin, N. Y., & Comes, T. (2022). Machine learning for spatial analyses in urban areas: A scoping review. *Sustainable Cities and Society*, *85*, 104050. https://doi.org/10.1016/j.scs.2022.104050

13) Chang, N.-B., & Bai, K. (2018). *Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing* (1st ed.). CRC Press. https://doi.org/10.1201/9781315154602

14) Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

15) Chiang, Y.-Y., Leyk, S., & Knoblock, C. A. (2014). A Survey of Digital Map Processing Techniques. *ACM Computing Surveys*, *47*(1), 1–44. https://doi.org/10.1145/2557423

16) Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

17) Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, *240*, 113–122. https://doi.org/10.1016/j.ecolmodel.2012.03.001

18) Czernecki, B., Nowosad, J., & Jabłońska, K. (2018). Machine learning modeling of plant phenology based on coupling satellite and gridded meteorological dataset. *International Journal of Biometeorology*, *62*(7), 1297–1309. https://doi.org/10.1007/s00484-018-1534-2

19) Dąbrowski, A. (2018). *The impact of landscape structure on the attractiveness of the residential area in Poznań* [Doctoral thesis, Adam Mickiewicz University]. http://hdl.handle.net/10593/24495

20) Dandl, S., Molnar, C., Binder, M., & Bischl, B. (2020). Multi-Objective Counterfactual Explanations. In T. Bäck, M. Preuss, A. Deutz, H. Wang, C. Doerr, M. Emmerich, & H. Trautmann (Eds.), *Parallel Problem Solving from Nature – PPSN XVI* (Vol. 12269, pp. 448–469). Springer International Publishing. https://doi.org/10.1007/978-3-030-58112-1_31

21) Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., Li, E., Su, H., & Liu, W. (2020). Advances of Four Machine Learning Methods for Spatial Data Handling: A Review. *Journal of Geovisualization and Spatial Analysis*, *4*(1), 13. https://doi.org/10.1007/s41651-020-00048-5

22) Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37. https://doi.org/10.1609/aimag.v17i3.1230

23) Fisher, A., Rudin, C., & Dominici, F. (2018). *All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*. https://doi.org/10.48550/ARXIV.1801.01489

24) Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

25) Gotway, C. A., & Young, L. J. (2002). Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, *97*(458), 632–648. https://doi.org/10.1198/016214502760047140

26) Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). *A Simple and Effective Model-Based Variable Importance Measure* (arXiv:1805.04755). arXiv. http://arxiv.org/abs/1805.04755

27) Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

28) Hänsch, R., Schulz, K., & Sörgel, U. (2018). Machine learning methods for remote sensing applications: An overview. In U. Michel & K. Schulz (Eds.), *Earth Resources and Environmental Remote Sensing/GIS Applications IX* (p. 1). SPIE. https://doi.org/10.1117/12.2503653

29) Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, *265*, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014

30) Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., Van Den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, *72*(4), 1607–1623. https://doi.org/10.1111/ejss.12998

31) 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. *Progress in Earth and Planetary Science*, *8*(1), 3. https://doi.org/10.1186/s40645-020-00398-0

32) Jasiewicz, J., & Hildebrandt-Radke, I. (2009). Using multivariate statistics and fuzzy logic system to analyse settlement preferences in lowland areas of the temperate zone: An

example from the Polish Lowlands. *Journal of Archaeological Science*, *36*(10), 2096–2107. https://doi.org/10.1016/j.jas.2009.06.004

33) 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

34) Jasiewicz, J., & Sobkowiak-Tabaka, I. (2015). Geo-spatial modelling with unbalanced data: Modelling the spatial pattern of human activity during the Stone Age. *Open Geosciences*, *7*(1), 20150031. https://doi.org/10.1515/geo-2015-0031

35) Jasiewicz, J., Zawiska, I., Rzodkiewicz, M., & Woszczyk, M. (2022). Interpretative Machine Learning as a Key in Recognizing the Variability of Lakes Trophy Patterns. *Quaestiones Geographicae*, *41*(1), 127–146. https://doi.org/10.2478/quageo-2022-0009

36) Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554. https://doi.org/10.1109/TKDE.2018.2861006

37) Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 3149–3157.

38) Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

39) Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, *23*(1), 18. https://doi.org/10.3390/e23010018

40) Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., Giovanni, R. D., Lohmann, L. G., De Carvalho, A. C. P. L. F., & Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, *38*(5), 5268–5275. https://doi.org/10.1016/j.eswa.2010.10.031

41) Loyola-Gonzalez, O. (2019). Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access*, *7*, 154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286

42) Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

43) Menard, S. W. (2008). *Applied logistic regression analysis* (2. ed., [Nachdr.]). Sage Publ.

44) Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (Second edition). Christoph Molnar.

45) Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, *135*(3), 370. https://doi.org/10.2307/2344614

46) Nikparvar, B., & Thill, J.-C. (2021). Machine Learning of Spatial Data. *ISPRS International Journal of Geo-Information*, *10*(9), 600. https://doi.org/10.3390/ijgi10090600

47) Nowosad, J. (2016). Spatiotemporal models for predicting high pollen concentration level of Corylus, Alnus, and Betula. *International Journal of Biometeorology*, *60*(6), 843–855. https://doi.org/10.1007/s00484-015-1077-8

48) Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *SOIL*, *6*(1), 35–52. https://doi.org/10.5194/soil-6-35-2020

49) Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

50) Rączkowska, Z., & Zwoliński, Z. (2015). Digital geomorphological map of Poland. *Geographia Polonica*, *88*(2), 205–210. https://doi.org/10.7163/GPol.0025

51) Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

52) Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11491

53) Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

54) Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

55) Rubbens, P., Brodie, S., Cordier, T., Destro Barcellos, D., Devos, P., Fernandes-Salvador, J. A., Fincham, J. I., Gomes, A., Handegard, N. O., Howell, K., Jamet, C., Kartveit, K. H.,

Moustahfid, H., Parcerisas, C., Politikos, D., Sauzède, R., Sokolova, M., Uusitalo, L., Van den Bulcke, L., … Irisson, J.-O. (2023). Machine learning in marine ecology: An overview of techniques and applications. *ICES Journal of Marine Science*, *80*(7), 1829–1853. https://doi.org/10.1093/icesjms/fsad100

56) Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

57) Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

58) Seber, G. A. F., & Lee, A. J. (2003). *Linear Regression Analysis* (1st ed.). Wiley. https://doi.org/10.1002/9780471722199

59) Shan, Y., Paull, D., & McKay, R. I. (2006). Machine learning of poorly predictable ecological data. *Ecological Modelling*, *195*(1–2), 129–138. https://doi.org/10.1016/j.ecolmodel.2005.11.015

60) Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

61) Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

62) Stupariu, M.-S., Cushman, S. A., Pleşoianu, A.-I., Pătru-Stupariu, I., & Fürst, C. (2022). Machine learning in landscape ecological analysis: A review of recent approaches. *Landscape Ecology*, *37*(5), 1227–1250. https://doi.org/10.1007/s10980-021-01366-9

63) Wachter, S., Mittelstadt, B., & Russell, C. (2017). *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*. https://doi.org/10.48550/ARXIV.1711.00399

64) Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

65) Watson, D. S. (2022). Conceptual challenges for interpretable machine learning. *Synthese*, *200*(2), 65. https://doi.org/10.1007/s11229-022-03485-5

66) White, D., & Sifneos, J. C. (2002). Regression Tree Cartography. *Journal of Computational and Graphical Statistics*, *11*(3), 600–614. https://doi.org/10.1198/106186002484

67) Zawiska, I., Jasiewicz, J., Rzodkiewicz, M., & Woszczyk, M. (2023). Relative impact of environmental variables on the lake trophic state highlights the complexity of eutrophication controls. *Journal of Environmental Management*, *345*, 118679. https://doi.org/10.1016/j.jenvman.2023.118679

7. Attachments

7.1. Academic achievements

Other publications:

- Dyba, K., Wąsala, R., Piekarczyk, J., Gabała, E., Gawlak, M., Jasiewicz, J., & Ratajkiewicz, H. (2022). Reflectance spectroscopy and machine learning as a tool for the categorization of twin species based on the example of the Diachrysia genus. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 273, 121058. https://doi.org/10.1016/j.saa.2022.121058
- Dyba, K., & Nowosad, J. (2021). rgugik: Search and Retrieve Spatial Data from the Polish Head Office of Geodesy and Cartography in R. *Journal of Open Source Software*, 6(59), 2948. https://doi.org/10.21105/joss.02948

Conferences:

- Ogólnopolska konferencja naukowa "Geoinformacja: Nauka – Praktyka – Edukacja": "*Klasyfikacja form rzeźby terenu z wykorzystaniem uczenia maszynowego na przykładzie Polski*", 1 December 2022, Adam Mickiewicz University in Poznań (National scientific conference "Geoinformation: Science – Practice – Education": "*Classification of landforms using machine learning on the example of Poland*") [Presentation]
- 3rd Forum Conference: "Classification of landforms using machine learning on the example of Poland" 19-21 September 2022, Doctoral School of Natural Sciences, Adam Mickiewicz University in Poznań [Poster]
- XXIV Ogólnopolska Konferencja Fotointerpretacji i Teledetekcji: "*Analiza temperatury jezior na podstawie termalnych zdjęć satelitarnych*", 27-28 September 2021 (24th National Conference on Photointerpretation and Remote Sensing: "*Analysis of lake temperatures from thermal satellite images*") [Presentation]
- 2nd Forum Conference: "*Analiza temperatury jezior na podstawie termalnych zdjęć satelitarnych*" 20-21 September 2021, Doctoral School of Natural Sciences, Adam Mickiewicz University in Poznań ("*Analysis of lake temperatures from thermal satellite images*") [Poster]
- IV Ogólnopolska Konferencja Hydrologiczna z okazji Światowego Dnia Wody: "*Analiza temperatury jezior na podstawie termalnych zdjęć satelitarnych*", 12 December 2020, Adam Mickiewicz University in Poznań (4th National Hydrological Conference on the occasion of

World Water Day: "*Analysis of lake temperatures from thermal satellite images*") [Presentation]

- 1st Forum Conference: "Use of machine learning and spectrometry to classify butterfly species" 21 September 2020, Doctoral School of Natural Sciences, Adam Mickiewicz University in Poznań [Presentation]
- I Ogólnopolska Konferencja dla studentów i doktorantów „Środowisko przyrodnicze jako obszar badań": "Przestrzenna predykcja parametrów glebowych z wykorzystaniem metod uczenia maszynowego", 30 May 2019, Adam Mickiewicz University in Poznań (1st National Conference for students and doctoral students "The natural environment as a research area": "Spatial prediction of soil parameters using machine learning methods") [Presentation]

Grants:

- [Principal investigator] "*The use of computer vision for geomorphological mapping*" funded by Polish National Science Centre (2021/41/N/ST10/00347) in 2022 – 2024
- [Employee] "*Optimization of agricultural crop productivity management using satellite imagery, alternatively with personalized GIS services generated in an autonomous way using artificial intelligence shared in a virtual way*" funded by Polish National Centre for Research and Development (POIR.01.01.01-00-2257/20-00)
- [Employee] "*Sustainable management of agricultural productivity using satellite imagery, based on personalized GIS services made available in a dedicated portal*" funded by Polish National Centre for Research and Development (POIR.04.01.02-00-0110/17-00)

Teaching:

I taught several classes on remote sensing, spatial analysis, and workshops on geographic information system and programming in R language. I was also an invited lecturer at the OpenGeoHub Summer School:

- 2022, Siegburg (Germany): "*Benchmarking R and Python for spatial data processing*" (https://opengeohub.org/summer-school/siegburg-2022/)
- 2023, Poznań (Poland): "*Unsupervised classification (clustering) of satellite images*" (https://opengeohub.org/summer-school/opengeohub-summer-school-poznan-2023/)

Other achievements:

- The open source rgugik package for searching and downloading spatial data was awarded by the Polish Head Office of Geodesy and Cartography in the competition for the best application using spatial data provided by this office in 2020
- My work on parallel raster processing in R programming language was discussed by Dr. Eric Nantz on the *"R Weekly Highlights"* podcast (https://podverse.fm/episode/ljbh5q9us, starting at 31:14)

7.2. Copies of articles

remote sensing

MDPI

Article

Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8

Krzysztof Dyba [1,*], Sofia Ermida [2], Mariusz Ptak [3], Jan Piekarczyk [4] and Mariusz Sojka [5]

[1] Laboratory of Applied Geoinformatics, Adam Mickiewicz University, Krygowskiego 10, 61-680 Poznan, Poland
[2] Instituto Português do Mar e da Atmosfera (IPMA), 1749-077 Lisbon, Portugal
[3] Department of Hydrology and Water Management, Adam Mickiewicz University, Krygowskiego 10, 61-680 Poznan, Poland
[4] Department of Environmental Remote Sensing and Soil Science, Adam Mickiewicz University, Krygowskiego 10, 61-680 Poznan, Poland
[5] Department of Land Improvement, Environmental Development and Spatial Management, Poznan University of Life Sciences, Piątkowska 94E, 60-649 Poznan, Poland
* Correspondence: krzysztof.dyba@amu.edu.pl

Abstract: Changes in lake water temperature, observed with the greatest intensity during the last two decades, may significantly affect the functioning of these unique ecosystems. Currently, in situ studies in Poland are conducted only for 38 lakes using the single-point method. The aim of this study was to develop a method for remote sensing monitoring of lake water temperature in a spatio-temporal context based on Landsat 8 imagery. For this purpose, using data obtained for 28 lakes from the period 2013–2020, linear regression (LM) and random forest (RF) models were developed to estimate surface water temperature. In addition, analysis of Landsat Level-2 Surface Temperature Science Product (LST-L2) data provided by United States Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA) was performed. The remaining 10 lakes not previously used in the model development stage were used to validate model performance. The results showed that the most accurate estimation is possible using the RF method for which RMSE = 1.83 °C and R^2 = 0.89, while RMSE = 3.68 °C and R^2 = 0.8 for the LST-L2 method. We found that LST-L2 contains a systematic error in the coastal zone, which can be corrected and eventually improve the quality of estimation. The satellite-based method makes it possible to determine water temperature for all lakes in Poland at different times and to understand the influence of climatic factors affecting temperature at the regional scale. On the other hand, spatial presentation of thermics within individual lakes enables understanding the influence of local factors and morphometric conditions.

Keywords: lakes; water temperature; Landsat; thermal images; Poland

check for updates

Citation: Dyba, K.; Ermida, S.; Ptak, M.; Piekarczyk, J.; Sojka, M. Evaluation of Methods for Estimating Lake Surface Water Temperature Using Landsat 8. *Remote Sens.* **2022**, *14*, 3839. https://doi.org/10.3390/rs14153839

Academic Editors: Andrzej Stateczny, Marta Wlodarczyk-Sielicka and Mariusz Specht

Received: 1 July 2022
Accepted: 4 August 2022
Published: 8 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Water ecosystems are characterized by considerable dynamics of energy and matter exchange, and therefore reflect modern transformations of the natural environment. Among the key properties of rivers and lakes is water temperature. Its course and distribution strongly determine a number of other processes [1]. It is therefore not surprising that the first water temperature measurements in lakes date back to the 18th century, and the frequency and accuracy of such measurements have been successively improving since. Traditional methods of temperature records based on in situ measurements are of point character, and are often not representative for complex and dynamic ecosystems such as lakes. The methods also have logistic limitations, and are costly and time consuming [2]. Moreover, the measurements are performed at various time intervals and at different depths, leading to high heterogeneity of the available data [3]. A complex understanding of the functioning of lake ecosystems requires collecting credible data on the state of water

at the global scale, over long time periods, with high spatial and temporal resolution [2,4]. Data regarding the surface temperature of lakes have been collected using remote sensing for many years [5]. Satellite image data have been recorded at the global scale since 1972, when the first civil Earth observation mission commenced under the name Landsat (formerly known as the Earth Resources Technological Satellite). For the first time in the Landsat program, a sensor with a thermal band was installed on the Landsat 3 satellite (launched in 1978), but it quickly failed. Each of the following satellites of this program was equipped with sensors that recorded thermal radiation emitted from the Earth's surface. An increase in water temperature in lakes shows a rate approximate to that of the growth of regional air temperature [6,7], although in the region of the Great Lakes and in northern Europe, a faster increase in lake water temperature was recorded than that of the surrounding air [7,8]. Satellite sensors used for the observation of the temperature of the surface water layer record emitted radiation at a range of 7–13 μm. The temperature of a very thin water layer is measured that can be referred to the temperature of the epilimnion in the conditions of lack of wind, although the external water layer is usually cooler than the 50 cm layer of sub-surface water [9]. Thermal scanners mounted on planes measure the temperature of the water layer on the interface of air and water with a thickness of 10–20 μm, but it can be several tenths of a degree warmer than temperature measured by a conventional sensor [10]. The difference between the temperature of the surface water layer and temperature in water depths is determined by factors such as the time of day, cloudiness, and wind speed [11,12]. To estimate the surface water temperature, data from one ("mono-window method") or two ("split-window method") thermal bands of the Landsat satellite can be used. The latter method is based on the difference in energy intensity in two thermal bands, which has the potential to improve temperature estimation [13]. In research determining the correlation between water temperature measured in situ and estimated from satellite data, thermal bands with low spatial resolution (i.e., Advanced Very High Resolution Radiometer) obtained very high values of determination coefficients ($R^2 > 0.9$) [14]. Similar results were obtained using data from Landsat satellites with higher spatial resolution, although with lower temporal resolution. Tavares et al. [15] recommends the use of thermal data from the Landsat 7 ETM+ sensor for the observation of small lakes because at a relatively high spatial resolution the temperature estimation error is very close to the error obtained from the MODIS sensor (RMSE 1.07 and 1.05 °C, respectively). The usefulness of other medium-resolution thermal satellite sensors for water surface temperature assessment, such as Landsat TM, ETM+, TIRS [16–20] and Terra ASTER [21–23], was also investigated. Lake temperature is primarily determined by meteorological factors (insolation, cloudiness, air temperature, and wind speed), and to a lower degree by geomorphometric factors (surface area and depth) [24]. The temperature of the surface layer of the lake increases with its insolation, although an increase in temperature also causes cloudiness, because the amount of long-wave radiation reaching the surface increases [25]. A negative correlation occurs between water temperature and mean lake depth, i.e., shallow lakes warm up faster and to higher temperatures [26,27].

In the case of the territory of Poland, a lot of research has been recently conducted on lake water temperature, in reference to both measurement techniques [28,29] and improvement of methods of its modelling [30]. Polish lakes (approximately 7000 natural lakes in total) with multiannual in situ datasets are suitable for searching solutions in both of the aforementioned aspects of research. Water temperature measurements in Polish lakes have a history of several decades, and systematic observations commenced at the end of the 1950s. Over the years, the population of lakes covered by field measurements has changed, and today it concerns several tens of objects. No comprehensive research with the application of satellite images in reference to lake water temperature in Poland has been undertaken to date, although studies in the scope dynamically develop in different regions of the globe [31–34]. Polish research so far covers three lakes, and the obtained results revealed high cohesion with in situ measurements [35], encouraging a broader

approach to the issue. Polish lakes already currently show a considerable increase in water temperature [29,36] that will continue to progress in the future [37,38].

The objective of this article is the comparison of four methods of estimation of water surface temperature in lakes in Poland based on thermal images from Landsat 8. Other detailed objectives include identification of limitations of the applied methods, and determination of the spatial variability of water temperature within the selected lakes and in a broader regional approach.

2. Materials and Methods

To organize the presentation of the results obtained in the subsequent stages, the following subsections were distinguished in this paper. Section 2.1 describes the study lakes, while Section 2.2 presents methods for conducting daily water temperature measurements. Sections 2.3 and 2.4 describe the acquisition of Landsat 8 Collection 2 Level 1 (L8L1) data and Landsat Level-2 Surface Temperature Science Product data (LST-L2). Sections 2.5–2.7 describe the method of calculating lake water temperature using linear regression model (LM), random forest regression (RF), and Ermida et al. (2020) [39] method (LST), respectively. Section 2.8 describes the validation process. Finally, Section 2.9 shows the possibility of using results to present the spatial variability of water temperature at the scale of a single lake and at the regional scale for multiple lakes. The flowchart of this study is shown in Figure 1.

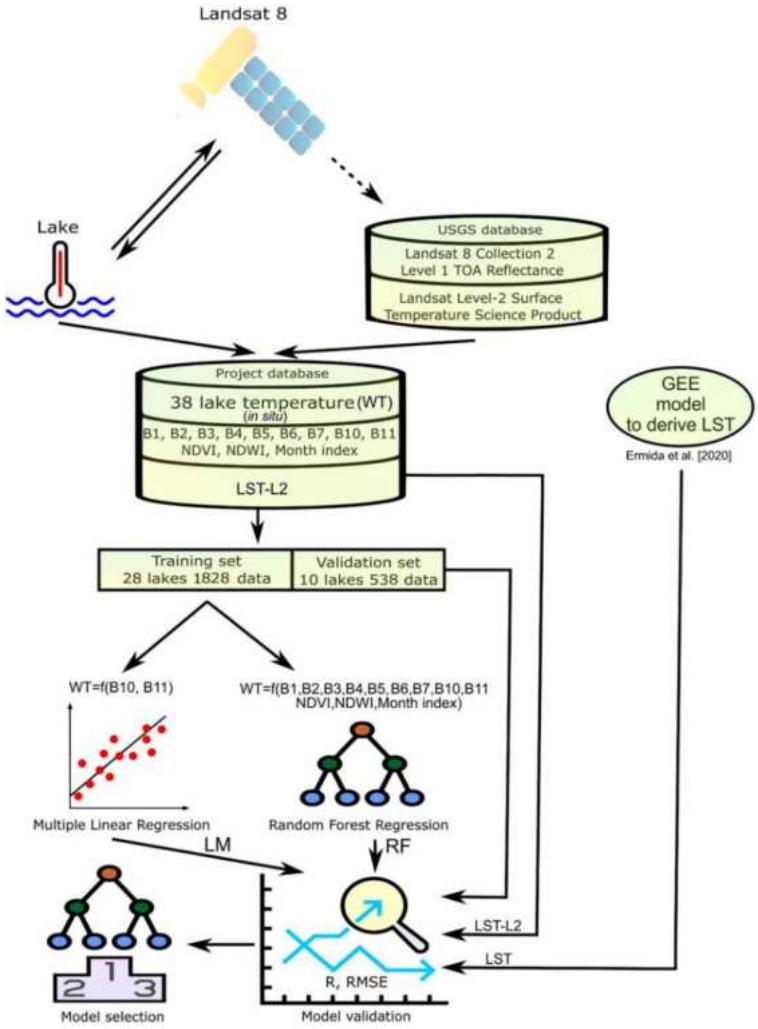

Figure 1. Flowchart of lake water temperature prediction. One method was developed by Ermida et al. (2020) [39].

2.1. Study Sites Description

The analysis of the quality of water temperature determination based on satellite images from Landsat 8 covered 38 lakes in north Poland (Figure 2). All lakes selected for the analysis are natural water bodies. Their genesis is primarily related to the course of geomorphological processes during the Weichselian glaciation. The maximum extent of the ice during the Weichselian in north Poland is marked with a dotted line in Figure 2. It also constitutes the boundary of three main lakelands. The northern area of Poland is the richest in lakes, covering more than 500 lakes with a surface area larger than 1 km^2. Lakes selected for the study have a variable surface area from 1.6 to 111.9 km^2, and mean depths from 1.3 to 38.7 m. In genetic terms, postglacial lakes are largely dominant (channel, moraine, and kettle) related both to the erosional and accumulation activity of the Scandinavian ice sheet. Lakes developed in the Holocene also occur, namely Jamno (No. 28), Gardno (No. 29) and Łebsko (No. 30) which are also the northernmost study objects. Lake Sławskie (No. 1) is the southernmost one. Lake Morzycko (No. 2) is the westernmost lake, and Lake Studzieniczne (No. 27) has the easternmost location. Details of the morphometric parameters of lakes are presented in Supplementary Table S1. Mean annual temperatures of the analyzed lakes are at a level from 8.6 to 11.4 °C. In the regional approach, evident differences are observed in the period of winter–spring, when ice cover is still present on the lakes in the eastern part of Poland, but already absent in the lakes of west Poland. It is due to the features of transitional climate of the analyzed region, when a longer effect of cold continental air is recorded in the east.

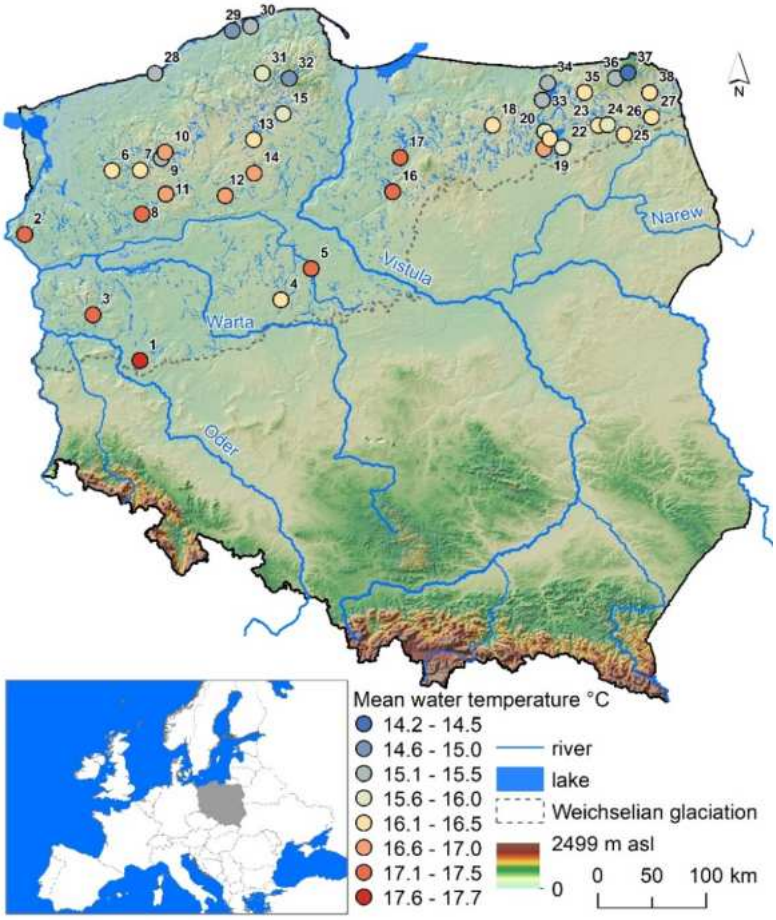

Figure 2. Location of the studied lakes with average water temperature from April to October. Numbering of lakes (1–38) according to Supplementary Table S1.

2.2. In Situ Data

In this study, the results of daily water temperature measurements conducted between 2013 and 2020 by the Institute of Meteorology and Water Management—National Research Institute (IMWM) were used. The period from April to October, when there is generally no ice cover on the analyzed lakes, was selected for the analysis. Water temperature measurements by the IMWM are performed with a frequency of once a day at 6 UTC. The measurement points are located in the coastal zone and the sensor is placed approximately 0.4 m under the water level. Temperature measurements are performed both automatically and manually (especially for previous years), and their accuracy is 0.1 °C. Based on earlier hourly measurements of lake water temperature in Poland [40], it was evidenced that the difference between water temperature measurements at 6 UTC (in accordance with the IMWM standard) and ~9 UTC (overpass time of Landsat) in the summer half-year averages 0.1 °C, reflecting high thermal stability of water. Moreover, it should be emphasized that water temperature variability in the near-surface water layer (from the surface to a depth of 1 m) is inconsiderable, and as shown by earlier studies in the case of lakes in Poland [41], it averages 0.2 °C.

2.3. Landsat 8 TOA Data

In this paper, Landsat 8 data provided by the United States Geological Survey (USGS) was used. The data from April 2013 to October 2020 over north Poland were analyzed. The Landsat 8 Collection 2 Level 1 Tier 1 (L8L1) data were obtained using the Google Earth Engine (GEE; accessed on 24 February 2022) [42]. The L8L1 dataset contains calibrated top-of-atmosphere (TOA) reflectance derived from the data produced by the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). These L8L1 products contain 5 visible and near-infrared (VNIR) bands, 2 short-wave infrared (SWIR) bands with spatial resolution 30 m and 2 thermal infrared (TIR) bands resampled from 100 m to 30 m. Only scenes covering positions in the reference in situ data were considered (Figure 3). Due to the fact that in situ lake water temperature measurement sites are located within the shoreline range, it is not possible to obtain L8L1 data for these locations specifically. That results in the pixels (resolution 30 m) that extend over land or shoreline vegetation as well. Therefore, L8L1 data were collected from a point located 60 m inside the lake in relation to the in situ lake water temperature measurement sites. The water difference in this neighborhood is negligible, and we have the confidence to exclude pixels that are land. For the analysis all scenes with cloud cover less than 60% were taken. Moreover, we excluded data marked as clouds and cloud shadows based on the CFMask algorithm [43]. CFMask derives from the Function of Mask (FMask) developed by Zhu and Woodcock [44] and Zhu et al. [45]. Additionally, we removed data with reflectance values below 0, temperature values below 0 °C in thermal band B10, and outliers based on the ultra-blue band (B1 values $\geq$ 0.14). Additionally, some observations were made on two scenes on the same day—we treated them as duplicates and removed them.

Due to cloud cover patterns in Poland, for each of the analyzed lakes a slightly different imagery dataset was obtained. The least, 38 L8L1 images were acquired for Białe Augustowskie Lake (No. 26), while 83 images were acquired for Niesłysz Lake (No. 3). The dates for Niesłysz Lake (No. 3) are shown in Figure 4. The red dots indicate imageries that were rejected from the analysis and did not meet previously defined criteria, while the green dots indicate imageries that were used for estimating water temperature in the lake. Only 43% of all available scenes were suitable for analysis. Supplementary Figure S1 shows the exact availability of satellite scenes by lake, month, year and water temperature.

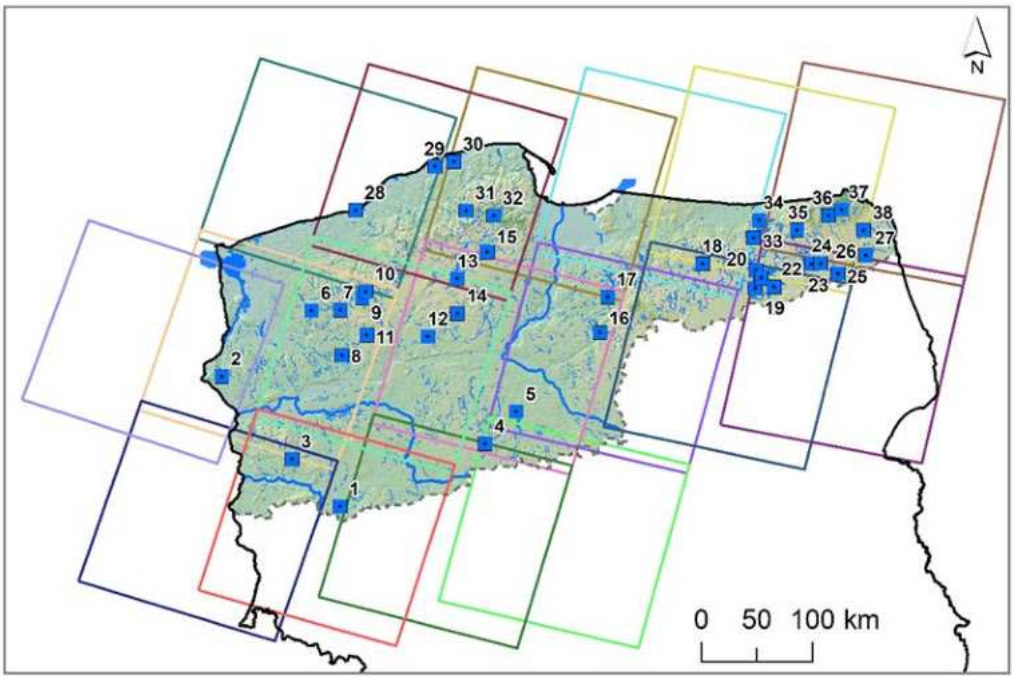

Figure 3. Landsat 8 Collection 2 Level 1 processed scenes. Numbering of lakes (1–38) according to Supplementary Table S1.

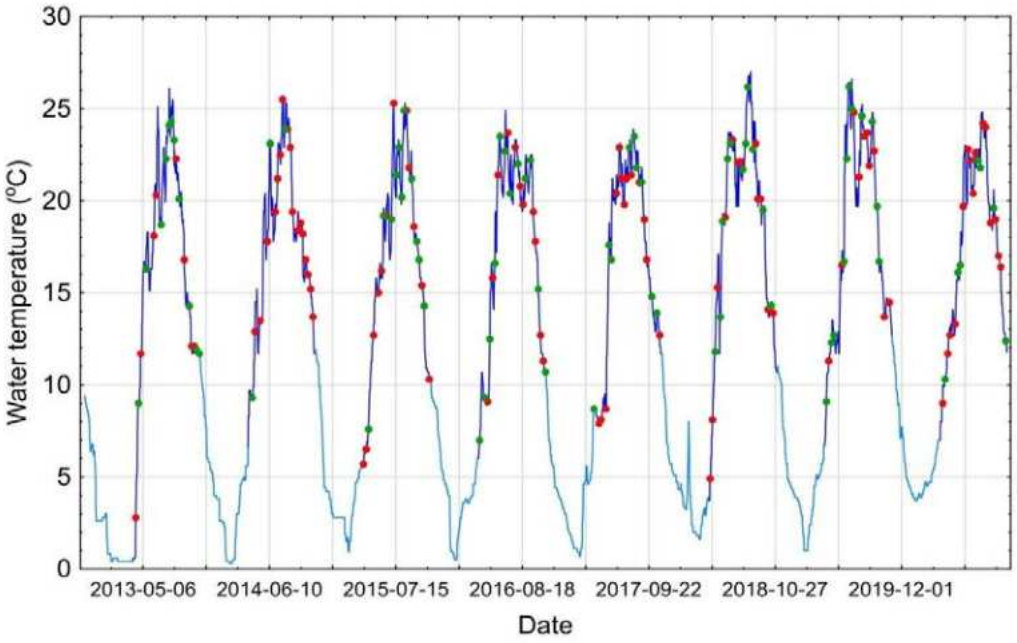

Figure 4. Data availability of Landsat 8 Collection 2 Level 1 in the background of the daily water temperatures of the Niesłysz Lake—the dark blue line presents the water temperatures in the period from April to October, the light blue line presents the water temperatures in the period from November to March, green and red dots present the dates of the imageries applied and excluded from the analysis, respectively.

2.4. Landsat Level-2 Surface Temperature Science Product

In this paper, the results of Landsat Level-2 Surface Temperature Science Product provided by the USGS were used [46]. The LST-L2 product contains the temperature of the Earth's surface in Kelvin (K). This product is generated mainly from the top-of-atmosphere thermal infrared data using a single-channel algorithm. Nevertheless, the algorithm is very extensive and uses auxiliary data, including data from the ASTER satellite and atmospheric

data from reanalysis. The product is available in a resolution of 30 m as georeferenced rasters (*.tif* format) in the Universal Transverse Mercator coordinate system. Information on the applied techniques related to atmospheric compensation, calibration methodology, and validation are widely presented and discussed by Cook [47], Cook et al. [48], Malakar et al. [18], Schaeffer et al. [49]. In the case of LST-L2, the data availability differed for all analyzed lakes due to cloud cover pattern in Poland. From among all analyzed lakes, the least number of LST-L2 images, i.e., 38, was acquired for Biale Augustowskie Lake (No. 26), while the most number of 88 was acquired for Jamno Lake (No. 28).

2.5. Simple Linear Model

We used multiple linear regression as the first choice method because of its simplicity. The purpose of this model is to determine the relationship between multiple explanatory variables and the modelled (response) variable. In other words, we want to predict the unknown values of one variable based on the known values of other variables using the following formula:

$$f(x) = b_1x_1 + b_2x_2 + \ldots + b_nx_n + a \quad (1)$$

where

$f(x)$—is the response variable;
b_1, b_2, b_n—are regression coefficients calculated for individual explanatory variables;
x_1, x_2, x_n—are the values of the explanatory variables;
a—is the intercept.

Therefore, our modelled variable is water temperature and our explanatory variables are the two thermal bands B10 and B11. The main assumption of linear regression is the linear relationship between the modelled variable and the predictors, which is obviously not clear in the case of complex natural processes. However, this is not the only assumption—further limitations are related to the lack of multicollinearity, independence and normal distribution of errors, and a limited number of explanatory variables. For this reason, a more robust machine learning model may be a better solution.

2.6. Random Forest Model

Random Forest (RF) is a method that enables the estimation of searched values based on a large set of uncorrelated and random trees [50]. In our work, we used it for regression analysis. RF is based on bagging and random subspace methods. Bagging is related to the creation of regression trees, which are then combined into ensembles to obtain an overall prediction. To design regression trees, a number of independent bootstrap samples are created from the original training dataset. Each bootstrap sample (D_b) is created by randomly sampling n subsamples from the original training data D, containing N subsamples. The bootstrap sample (D_b) usually consists of 2/3 of D and contains no duplicate subsamples. Then, K independent regression trees are created for bootstrap samples with input vector x. In the regression analyses, the average prediction of the K regression trees, $h_k(x)$, is calculated to obtain an overall prediction [51,52] as follow,

$$y = \frac{1}{K}\sum_{k=1}^{K} h_k(x) \quad (2)$$

Since the regression trees have high variance, the bagging is designed to reduce the variance and to prevent overfitting the complex RF model. Therefore, the learning trees cannot be correlated. Samples derived from the dataset that were not selected to train the k-th regression tree during the bagging process are compiled into an out-of-bag (*OOB*) subdataset. The *OOB* subdataset contains the remainder of D dataset ($^1/_3$ samples). Based on *OOB* subdataset, the performance of the k-th regression tree is calculated by the mean squared error (MSE_{OOB}) as,

$$MSE_{OOB} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y_i})^2 \quad (3)$$

where y_i and $\overline{y_i}$ are the prediction and the mean of the i-th prediction from all regression trees. The determination coefficient R^2_{OOB} of the OOB subdataset, can be calculated based on MSE_{OOB} and the total variance Var_y, of the output parameter, using the following formula,

$$R^2_{OOB} = \frac{MSE_{OOB}}{Var_y} \quad (4)$$

The input data during the RF analyses were TOA reflectance values (B1, B2, B3, B4, B5, B6, B7, B10, and B11). Moreover, the spectral indices the Normalized Difference Vegetation Index (NDVI) [53] and the Normalized Difference Water Index (NDWI) [54], the month index as auxiliary variables to maximize prediction performance were used. All calculations related to the RF were accomplished using the *ranger* [55] package within the statistical software R 4.1.2 [56].

2.7. Land Surface Temperature Model

Ermida et al. (2020) [39] provides a code on GEE to derive LST from Landsat Collection 1 Level-1 thermal infrared bands. The method, referred to hereafter as LST, is based on the Statistical Mono-Window (SMW) algorithm developed by the Climate Monitoring Satellite Application Facility (CM-SAF) for deriving LST climate data records from Meteosat First and Second Generation [57]. The algorithm uses a single thermal infrared band (band 10 in the case of the Landsat 8) for consistency across all Landsat series. This method also uses the ASTER Global Emissivity Dataset (GED) database and applies vegetation cover correction based on Landsat NDVI. Since not all lakes considered in this study were correctly identified by the ASTER GED dataset, we have updated the code to use the Landsat water mask (available through the quality flag). For water bodies, the emissivity was set to 0.99.

2.8. Model Validation Procedure

To assess the accuracy of the LM, RF, LST and LST-L2 model predictions, we randomly selected 10 of the 38 lakes studied. In total, we used 538 in situ lake water temperature measurements to validate the temperature prediction results, which is approximately 23% of the input dataset. The root mean square error (RMSE), mean bias error (MBE), coefficient of Pearson correlation (R), between measured and predicted values were used for validation. Additionally, we calculated the standard deviation (SD) of the predicted values.

2.9. Model Application

Results obtained from the best model were used for the presentation of the spatial distribution of water temperature in lakes in the studied area at the local and regional scale. The objective of the presentation of the spatial variability of water temperature in single lakes was to show the role of local factors, i.e., the morphometry of the lake basin, range of impact of river waters, groundwater supply, and effect of wind on the distribution of water temperatures in the lake. The presentation of the thermal regime of lake waters at the regional scale aimed at illustrating the effect of climatic factors on the spatial distribution of water temperature in the lakes of north Poland in different time periods from April to October.

We tested the applicability of the developed RF model for the entire example scene of 28 April 2021 (ID: LC08_191023_20210428) and made the temperature prediction for 1568 lakes located in the north-western part of Poland with an area larger than 3 ha (approximately 30 pixels). On a local scale, we chose 4 lakes on different dates—Gopło, Łebsko, Drawsko, Ełckie (IDs: LC08_190023_20210608, LC08_191022_20150818, LC08_192023_20180529, and LC08_187022_20211025). We used the *terra* [58] package to process spatial data (both raster and vector) and perform spatial predictions.

3. Results

3.1. Comparison of Methods

At the first stage of work, based on a training sample covering measurement data from 28 lakes (1828 measurements) and the corresponding data from Landsat 8 Collection 2 Level 1 (L8L1) meeting the criteria described in the methodology, a multiple linear regression model (LM) was developed, as well as a random forest model (RF). The following equation parameters were obtained for the linear model:

$$\text{Water temperature} = 2.9 \times \text{B10} - 2.07 \times \text{B11} + 48.48 \quad (5)$$

where bands B10 and B11 represent Top of Atmosphere Reflectance in Kelvins.

For model LM at the stage of development, the following statistics were obtained: R = 0.91, MBE = 0 °C, and RMSE = 2.28 °C. Considerably better statistics were obtained for the RF model, where the values were R = 0.95, MBE = 0.01 °C, and RMSE = 1.66 °C. Notice, however, that the RF algorithm strongly fits the training data. The analysis of the effectiveness of the RF model at the stage of its development within particular lakes showed very high results of correlation coefficients at a range from 0.98 to 1.0. MBE values were at a level from −0.52 to 0.58 °C, and RMSE values from 0.44 to 1.24 °C. The best results of temperature prediction at the stage of development of the RF model were obtained for Lake Ińsko (No. 6), and the worst for Lake Gopło (No. 5). Considering model LM, results obtained for particular lakes were as follows: R from 0.84 to 0.97, MBE from −1.48 to 1.86 °C, RMSE from 1.22 to 3.54 °C. The best results of temperature prediction at the stage of development of model LM were obtained for Lake Roś (No. 22), and the worst for Lake Gopło (No. 5). See Supplementary Table S2 for detailed results.

Considerably more credible from the point of view of further application of the model are results of the validation of statistical models on independent data not used at the stage of model development. The results are obviously worse than those obtained at the stage of model development. In this study, for LM, the obtained values R and RMSE were lower than in the case of RF model (Table 1). In the case of the LST model, RMSE values were almost twice higher than for results obtained from the RF model. The worst results of the estimation of lake water temperature were obtained for Landsat Level-2 Surface Temperature Science Product (LST-L2). In this case, MBE and RMSE values were 2.55 and 3.68 °C, respectively.

Table 1. Statistics comparing model performance on a test set. The last row with band 10 Top of Atmosphere Reflectance (uncalibrated) is given as a reference.

Method	MBE [°C]	RMSE [°C]	COR	SD [°C]
LM	−0.01	2.29	0.91	4.98
RF	−0.06	1.83	0.94	4.92
LST	2.04	3.35	0.88	5.52
LST-L2	2.55	3.68	0.9	5.94
B10 TOA	−2.11	2.70	0.88	4.83

Considering the fact that next to values of global statistics obtained for the entire dataset, the selection of the appropriate model for the estimation of water temperature in lakes requires broader analysis, at least in the temporal approach—of particular months. In the period from April to October, a substantial change in water temperature occurs in particular lakes (from 0.3 °C to 28.3 °C), and a thermal gradient occurs from the east to the west of Poland modified between lakes by the impact of the Baltic Sea. The analysis of results obtained from the validation of LM and RF models evidently shows that RMSE results are lower than those obtained from finished products LST-L2 and LST (Table 2). An advantage of LST-L2 and LST models is their global cover and no requirement for measurement data for their calibration. The comparison of RMSE values between LM and RF models using the Wilcoxon signed-rank test showed a significant difference between

them (p-value < 0.05). It suggested that RMSE values are lower for RF model than for LM. Differences between the estimation of water temperature are particularly evident in April (more than 1 °C), and the smallest differences occur in August (0.13 °C). Considering all models, the lowest RMSE values occur in October, and the highest in May. The analysis of the values of correlation coefficients for all methods showed that they are at a similar level, and the occurring differences are not statistically significant.

Table 2. Model performance per month on a test set. Note that the month index was used as a predictor in the random forest model.

Month	RMSE				Correlation			
	LM [°C]	RF [°C]	LST [°C]	LST-L2 [°C]	LM	RF	LST	LST-L2
April	2.91	1.84	4.39	4.28	0.71	0.75	0.7	0.69
May	2.97	2.43	4.42	4.87	0.69	0.66	0.68	0.67
June	1.81	1.48	2.84	3.39	0.76	0.82	0.74	0.76
July	1.8	1.46	2.68	3.59	0.68	0.71	0.66	0.71
August	2	1.87	3.13	3.42	0.72	0.67	0.63	0.69
September	1.93	1.45	2.2	2.4	0.81	0.8	0.75	0.77
October	1.75	1.58	1.8	1.62	0.88	0.79	0.8	0.84

Results of the validation of the LM, RF, LST, and LST-L2 models are presented in Figure 5. The greatest dispersion of points presenting the comparison of measured and estimated values evidently occurs for April and May. According to Skowron [59], in that period water temperature shows high dynamics due to water mixing related to the development of stratification. Moreover, water temperatures estimated by means of the LST and LST-L2 models are evidently higher than the measured ones, suggesting the occurrence of a systematic error. For models RF and LM, the dispersion of points focuses within a straight line of $y = x$, suggesting the occurrence of only incidental/random errors.

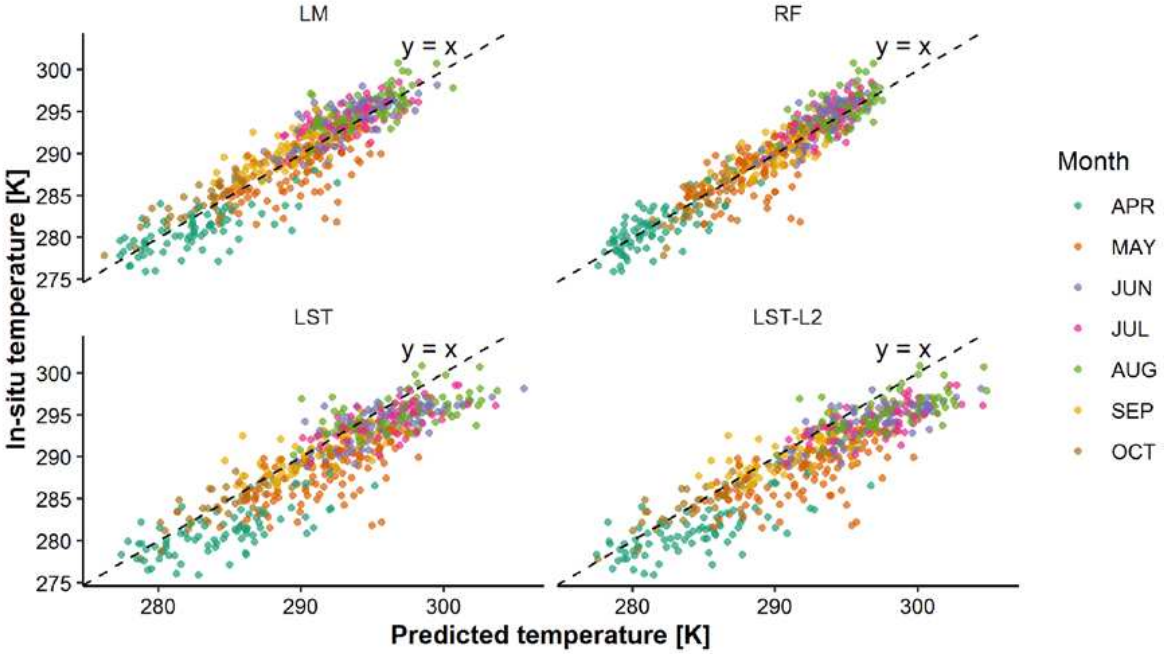

Figure 5. Scatterplots for the models tested. The black dashed line indicates the ideal prediction. A systematic error can be noticed for the LST and LST-L2 models.

Differences between measured water temperatures and those estimated by means of particular models are presented in Figure 6. Distributions of deviations obtained for the LST and LST-L2 models are evidently shifted right from the value of 0, pointing to

the occurrence of systematic errors of 2.0 and 2.6 °C, respectively (for a testing sample of 10 lakes, 538 measurements). Due to this, improvement of the estimation of water temperatures can be relatively easily obtained by means of LST and LST-L2 by introducing a correction factor. Results of such analyses are presented in the next sub-chapter. In the case of the remaining LM and RF models, the average deviation between the measured and forecasted value is 0.0 and 0.1 °C, respectively. This points to the lack of occurrence of systematic errors. Histograms between actual and predicted values by all models indicate that they are normally distributed, although this is not apparent from the Shapiro–Wilk tests (p-value < 0.05). This is mainly due to single outliers.

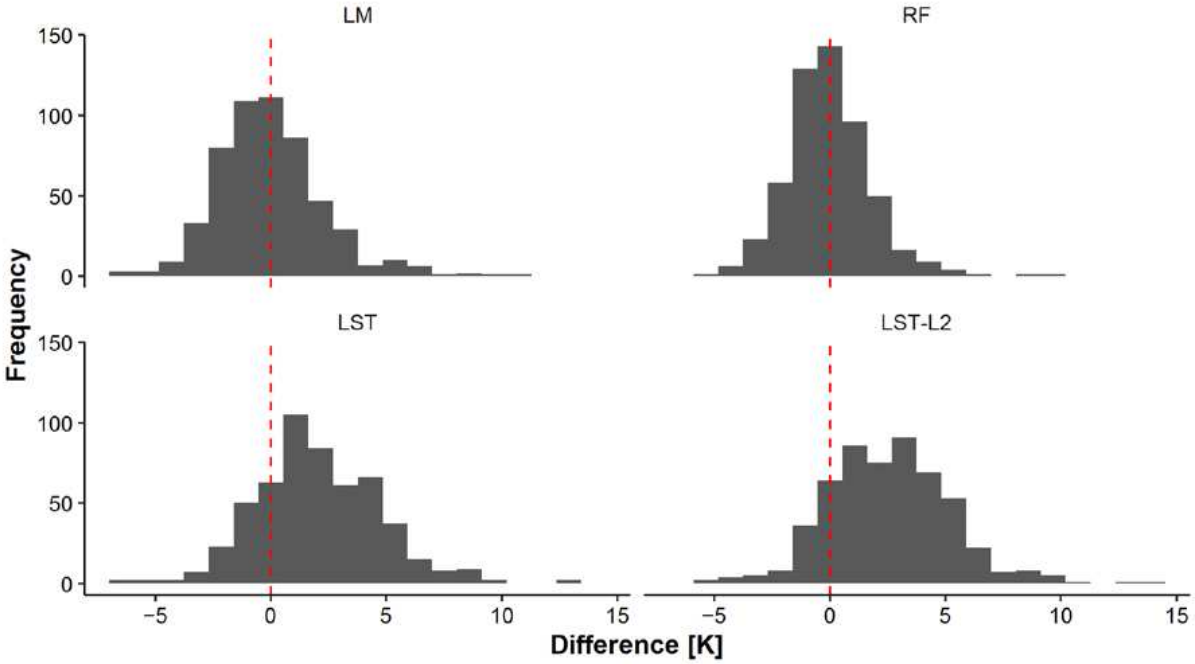

Figure 6. Differences between measured water temperatures and water temperatures estimated by models. The red dashed line marks the error equal to 0 K.

The results obtained both globally and for individual months suggest that the best estimate of lake water temperature is possible using the RF model. This probably results, among others, from the fact that the collection of explanatory (independent) variables covers as many as 12 parameters (B1, B2, B3, B4, B5, B6, B7, B10, B11, the NDVI, the NDWI and the month index), and in LM only 2 parameters are used (B10 and B11). Therefore, an analysis of the suitability of the explanatory variables for estimating water temperature using the impurity algorithm from the random forest model was conducted (Figure 7). We found that B10 was more significant than B11, which is consistent with other publications. This is also confirmed by the result obtained for the linear model—a slightly higher Pearson correlation coefficient was noted for band B10 (R = 0.88) than for band B11 (R = 0.84). The results of the study by Barsi et al. [60] showed that due to the occurrence of out-of-field stray light, the calibration system of the TIR sensor is unstable, and the recorded thermal data may carry a substantial error. The effect of dispersed light on data recorded in band B10 is relatively small and therefore should be preferred. Considerably greater errors can occur in the case of data in band B11, they should be subject to additional external calibration. Another important variable is the month index—this is related to the seasonality of water temperature. Specifying the month significantly reduces errors resulting from incorrect satellite measurements. For example, in the warm months, we expect statistically higher temperatures than in the cold months. Thus, excluding weather anomalies, it can be expected that the water temperature range in June will be approximately from 18 °C (lower quartile) to 21.6 °C (upper quartile). The other variables are of little importance. Of the

remaining spectral bands, the green band (B3) is the most significant. Green radiation that carries a lot of energy easily penetrates water. The depth of such penetration determines the magnitude of changes in water temperature, and depends on, among others, water transparency. Therefore, the amount of radiation reflected from water recorded in the green band can be related to water temperature [61–63].

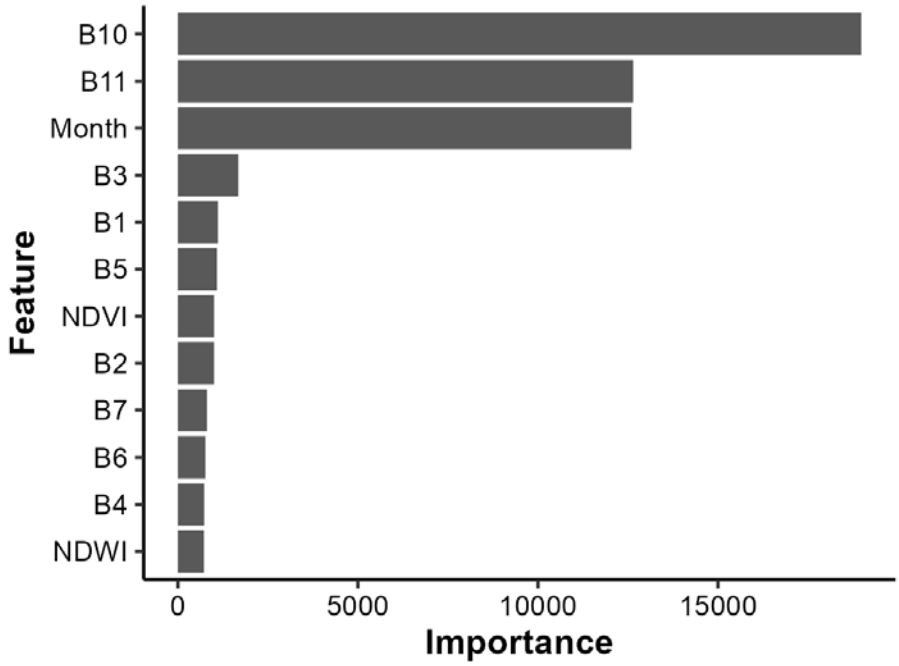

Figure 7. Feature importance using impurity algorithm.

We also evaluated the usefulness of spectral indices, the NDVI (used to calculate emissivity) and the NDWI (used to distinguish water from land), but they did not significantly affect the results.

Although the LM provides somewhat worse prediction results than the RF model, it has its evident advantages. Predictions of RF model employ 12 independent variables, which in the case of the analysis of the entire Landsat scene requires a total of approximately 5.4 GB RAM. In the case of LM, it will be only approximately 900 MB (2 variables). Next to the number of variables, the complexity of the model is also important—the linear model is much simpler, allowing for considerably faster prediction in terms of calculations. The last difference is related to the application of the month index—in the case of RF, it substantially improves prediction results, although it results in limiting the applicability of the model only to the geographic area of Poland (approximate latitudes and longitudes related to climate). In LM, no such assumptions were adopted; therefore, the potential of its applications is greater. Nonetheless, part of predictors can also be excluded post hoc in RF with no loss of effectiveness.

The spatial analysis of values of RMSE characteristics for the best RF model shows better adjustment of the model to the training sample of lakes (orange color) than for the test sample (points marked with red color) (Figure 8). The spatial analysis of RMSE values within the training sample showed no occurrence of spatial autocorrelation, i.e., grouping of lakes with evidently lower or higher RMSE values. Among the lakes used for testing the model, the lowest RMSE values were obtained for Lake Ełckie (No. 23): 1.40 °C. The highest RMSE values were obtained for Lake Hańcza (No. 37): 2.83 °C. Both Lakes Hańcza and Ełckie are located in east Poland. The correlation analysis between RMSE values and lakes' areas, depths and volumes showed no significant relationship for both the training set and the test set.

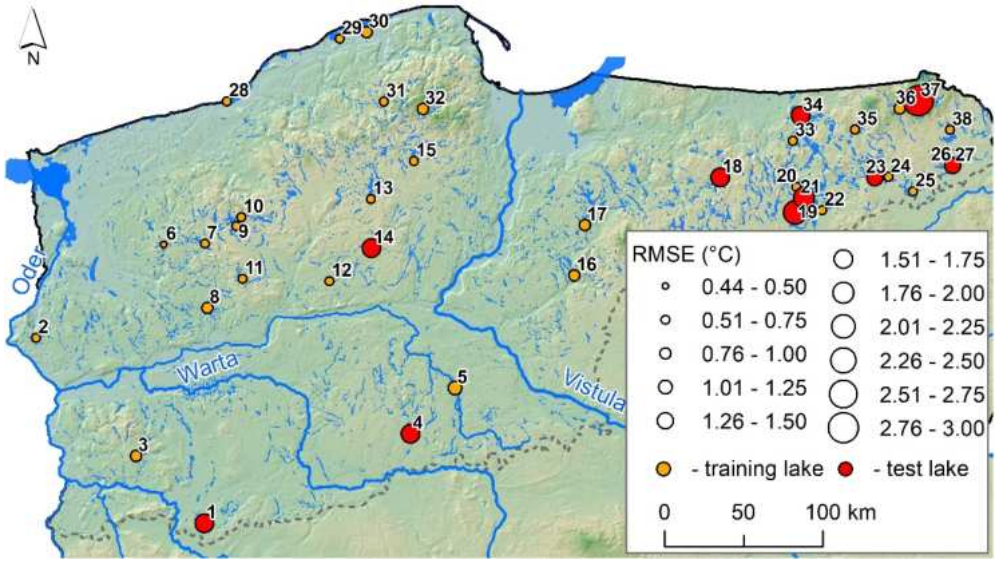

Figure 8. Spatial variability of RMSE values for the training sample and test sample in the random forest model. Orange color signifies training lakes, and red color marks test lakes. Numbering of lakes (1–38) according to Supplementary Table S1.

3.2. LST-L2 Calibration

Because LST-L2 data are a finished product that can be practically applied by a large group of scientists, including hydrologists, limnologists, and water ecologists with no knowledge of specialized tools for obtaining and processing satellite data or development of statistical models (such as LM or RF), a decision was made to develop a calibration system for Poland for the estimation of water temperatures in lakes based on LST-L2 data provided by USGS. The analysis employed an identical method of determination of calibration parameters (based on 1828 measurements) and validation (538 measurements) as described in Section 2. The comparison of LST-L2 values with values of temperature measured in situ showed the occurrence of a systematic error of approximately 2.42 °C (Figure 9).

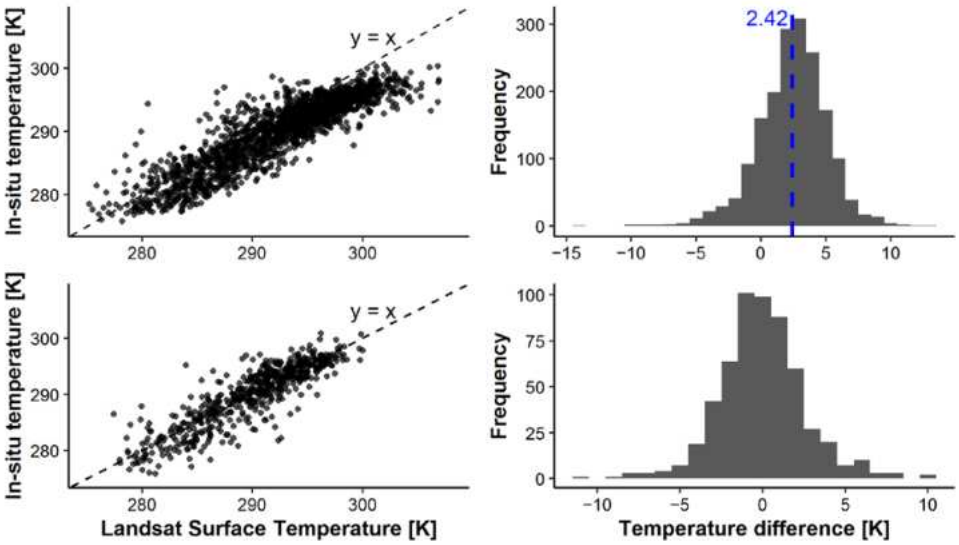

Figure 9. The difference between in situ temperature and LST-L2 in the scatter plot (first column) and histogram (second column). The first row represents uncalibrated training data, while the second row represents calibrated testing data. The blue dashed line indicates the mean systematic error of approximately 2.42 K. After the calibration, the systematic error was corrected.

We tested three ways of correcting it using a linear equation (with and without an intercept) and adding the mean value of this error. We obtained successively RMSE values on the test set: 2.59, 2.86, and 2.88 °C (the initial RMSE value of LST-L2 was 3.68 °C). We strongly recommend calibrating the LST-L2 data for estimating the water surface temperature by using the formula below:

$$\text{LST-L2 corrected} = 0.806 \times \text{LST-L2} + 54.37 \quad (6)$$

4. Discussion

Satellite images are an important tool in the case of research on water ecosystems, and find application in reference to many issues concerning both biotic [64–66] and abiotic processes [67–69]. A number of papers refer to the elementary feature, namely water temperature, and the applied methods analyze the issue at various temporal and spatial scales [15,70–74]. As mentioned in the introduction to this paper, research on lake water temperature in Poland with the application of satellite images has found no broader application to date, as confirmed by scarce papers in the scope. Ptak et al. [35] compared in situ water measurements with Landsat images (4, 5 and 7) for three coastal lakes, obtaining high coefficients of determination R^2 (0.87–0.95). Based on in situ measurements and Landsat 8 data regarding surface water temperature in Lake Raduńskie Górne, Szkwarek and Wochna [75] obtained credible results allowing for detailed presentation of its spatial distribution. Monitoring programs are often based on a limited spatial and temporal range. Remote sensing offers promising tools for large-scale observation, improving our possibilities of comprehensive research of the indices of lake properties [32]. Therefore, in the context of dynamically progressing remote sensing technology, its limited scope of applicability in the analysis of lake water temperature in Poland constitutes a research gap that is attempted to be filled based on Landsat 8.

The application of Landsat images in this paper corresponds with the intensively addressed global research trend regarding lake water temperature. The applied methodical approach showed an RMSE error at a level of 1.8 °C, a result approximate to that in other studies using Landsat images. Simon et al. (2014), analyzing two freshwater lakes in France, determined the relationship between Landsat thermal data and water temperature by measuring the temperature in situ at various depths, from 0 to 0.5 m [20]. The authors concluded that the obtained results (R^2 ranging from 0.90 to 0.94 and RMSE ranging from 1.753 to 2.397 °C) were satisfactory and coherent with literature data, where the values of mean square error are within a range from 1 to 2 °C. Giardino et al. [17] presented the Landsat TM-derived surface temperature of the sub-alpine lake with the RMSE of 0.328 °C; however, the procedure they used was not based on empirical relationships between satellite data and in situ measurements. For several hundred lakes analyzed based on data from the period 1999–2016 (Landsat 5 and 7), the total RMSE of temperature measurements obtained from the satellite was approximately 1.2 °C [19].

Mean absolute error for the analyzed lakes averaging from 1.38 to 2.39 °C (depending on the applied method) was approximate to the results of research conducted by Schaeffer et al. (2018) for lake waters in the United States, and reached 1.34 and 4.89 °C [49]. Results obtained in the study were also very approximate to those obtained by Tavares et al. (2019), who compared different methods of temperature estimation based on data from Landsat 7 and MODIS satellite [15]. In the estimation of water temperature (LSWT), the authors applied the single-channel algorithm method and two databases, SAFREE and TIGR3, obtaining similar R^2 values (0.936 and 0.938, respectively) as in this study. A weaker correlation with in situ water temperature was also shown by MOD28 data from the MODIS satellite (R^2 = 0.906), and a stronger one by MOD11 data (R^2 = 0.962). The most useful for the estimation of temperature (R^2 = 0.964) proved to be data from Landsat 7 satellite with the application of radiative transfer equation applied with atmospheric correction parameters from AtmCorr.

Considering the lack of possibility of conducting regular field observations, satellite remote sensing was recognized as a cost-effective way of monitoring water surface temperature at large spatial and temporal scales [76]. This statement is confirmed by results obtained in this paper, as exemplified by Figure 10, presenting a set of thermal data regarding three hundred lakes—including only several monitored through field observations. It should be emphasized that it is the first elaboration of the type for a fragment of the Polish territory, providing a promising basis for conducting monitoring of water temperature in lakes for a population of objects not achievable so far. Because water temperature is an elementary property of lake ecosystems [77,78] directly determining a number of processes and phenomena (biotic and abiotic), detailed information in the scope is necessary in reference to each lake. The obtained spatial image presented in Figure 10 points to the thermal variability of the analyzed lakes. The distribution in the orientation from the north (middle fragment) to the south is generally relatively clear, with occurrence of, respectively, cooler and warmer lakes. The situation illustrates the effect of a combination of environmental factors responsible for the thermal conditions of lakes. They include the surface area of the lake, land use structure around the lake, length, location (coordinates), and elevation above sea level [36]. According to the conducted research, multiple regression results showed that the recorded spatial distribution was significantly determined by, among others, the latitude, surface area, distance from the sea, presence of forests, and elevation above sea level. Similar research conducted for lakes on the Arctic coastal plain showed that, among others, parameters such as geographic location and morphometry of the lakes control their temperature at the regional scale [79].

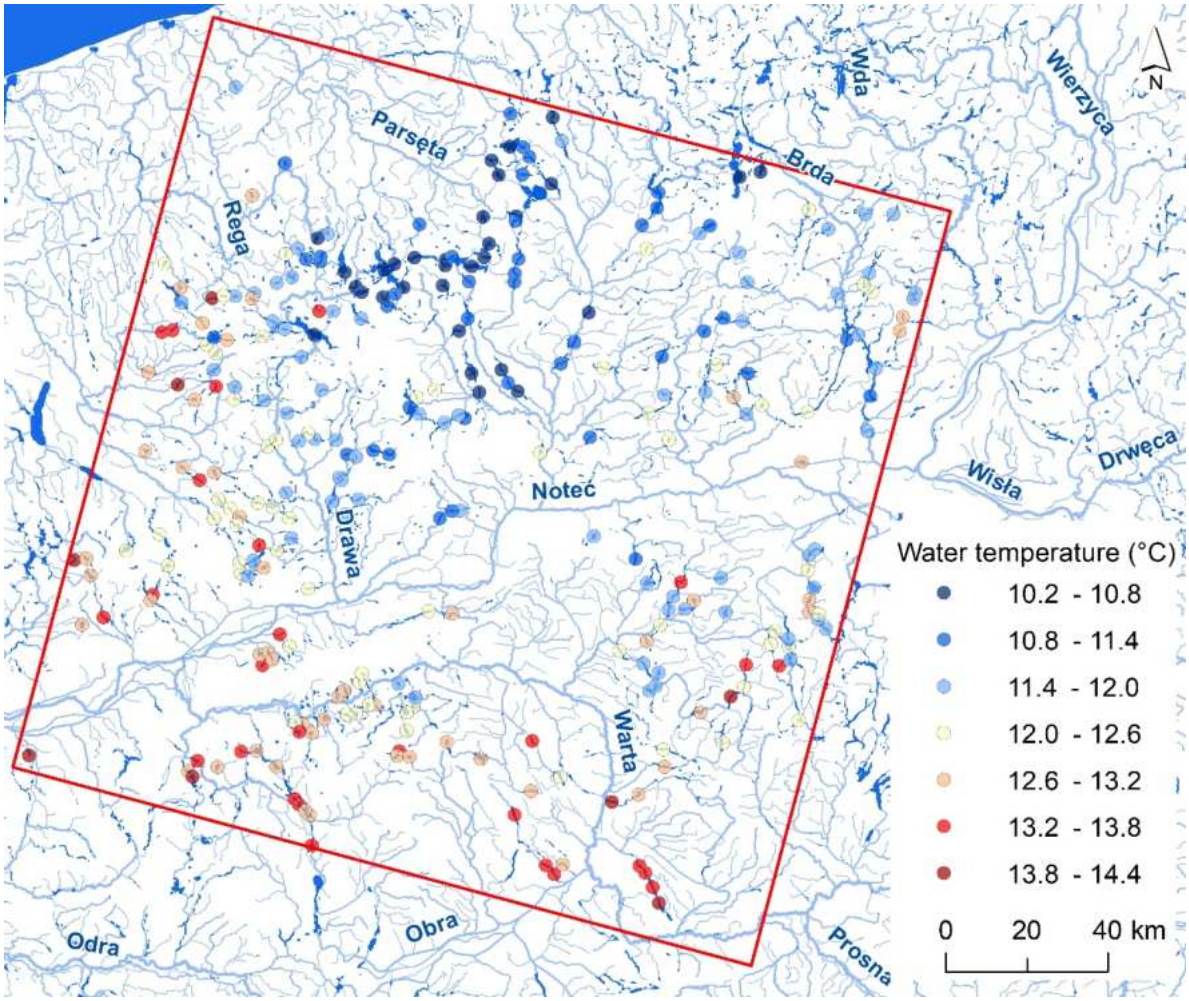

Figure 10. Median surface temperatures of non-monitored lakes as of 28 April 2021.

The analyzed lakes with cooler water temperature occurred with greater density in the catchments of the Gwda and Drawa Rivers, on postglacial outwash plains built of highly permeable formations, i.e., gravels and sands. The situation is evident in the case of runoff, where the groundwater component is dominant in both of the aforementioned catchments [80], causing greater supply of cooler waters to the lakes.

Detailed distribution of water temperature is presented for two lakes with variable morphometric parameters (Figure 11). In both cases, certain patterns were recorded also observed in other studies of the type, i.e., distribution of water temperature referring to the depth of particular sectors of the lake and changes in water temperature in estuarial zones of rivers. According to Cao et al. [81], temperature in the center of the East Lake (Wuhan, China) is higher than in the surrounding area. Maps of surface temperature obtained by means of Landsat 8 satellite for three lakes (Maggiore, Lugano, and Como) showed variability reaching two or three degrees between different areas, i.e., between centers of lakes and coastal zones where inlets of rivers are located [82]. Based on data from Landsat 8 satellite, Sener and Sener [83] determined surface water temperature for Lake Beysehir (Turkey), recording a range of water temperature from 20.1 to 26.8 °C. They also determined that it was the highest in shallower sectors of the analyzed lake. Aïtelghazi et al. [84] emphasise that important parameters to be considered in the context of thermal changes include depth and suspended particles transported by the river. In the case of dimictic Lake Drawsko (Figure 11a), the distribution of water temperature in reference to depth is evident—the warmest in the coastal zone, and coldest in the deepest areas. A different situation is recorded in the case of polymictic Lake Łebsko (Figure 11b), where water temperature is uniform throughout the water column, and its variability on the surface is determined by inflows of surface waters—the coldest zone occurs in the south-eastern part (mouth of the Łeba River with average water flow), and in the north-eastern part (inflow of marine water), as well as in the north-western part, where a cooler water mass was pushed to an isolated bay with limited circulation.

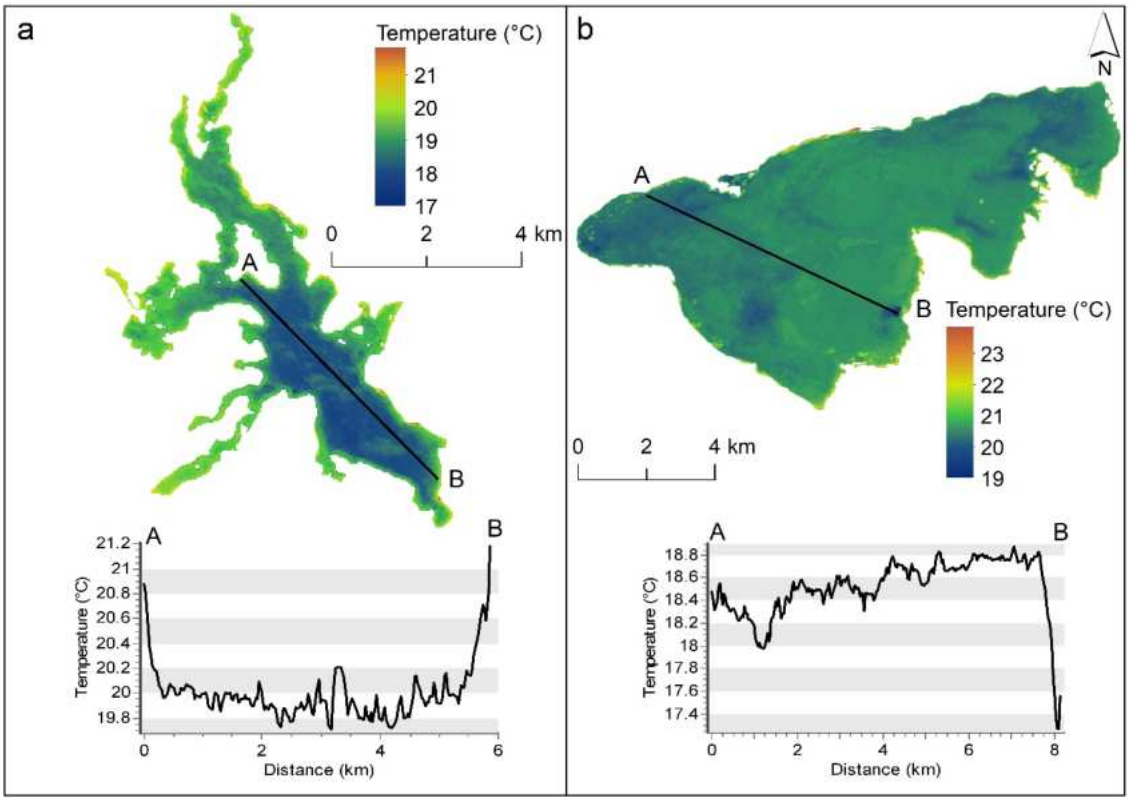

Figure 11. The surface temperature of (**a**) Drawsko Lake on 28 May 2018; (**b**) Łebsko Lake on 18 August 2015. The black lines labeled AB show a temperature profile from north-west (A) to south-east (B).

The potential of Landsat images for monitoring small water bodies can be used for forecasting their evolution, and will be useful in the future [85]. It is important in the context of the progressing climate change and the related threat in reference to, among others, water quality or species composition of the fauna and flora of lake ecosystems.

5. Conclusions

This study concerned the comparison of the effectiveness of four methods of estimation of lake temperature based on satellite data. Two empirical models were developed (linear regression and random forest), and compared with two state of the art approaches employing atmospheric calibration. The comparative analysis shows the best performance of the random forest method, providing the highest correlation coefficient and the lowest RMSE.

Many papers recommend the application of atmospheric correction before using satellite data for the estimation of lake water temperature. It should be remembered, however, that the accuracy of these correction methods can be considerably reduced under certain conditions, e.g., with high content of water vapor in the atmosphere, and atmospheric models themselves show substantial limitations in spatial and temporal resolution. Therefore, continuous work on the improvement of correction algorithms should be accompanied by research on the possibility of application of new methods of data analysis (e.g., machine learning) and consideration of empirical data in the improvement of prediction results. This paper proposes an appropriate calibration correction for data from Landsat Level-2 Surface Temperature Science Product (LST-L2) that considerably improves results (root mean square error reduced by 30%) of estimation of lake temperature in the territory of Poland. Nevertheless, despite the correction, the random forest model provided the best results.

Future works should focus on determining the pixel based temperature uncertainty, analyzing the potential of using low-resolution data from the MODIS sensor for small surface lakes in Poland and determining the effectiveness of other water and vegetation indices. As of 31 October 2021, imagery is available from the next Landsat series satellite with the improved TIRS-2 sensor. Along with the increase in data availability, it should also be used for analysis to maintain continuous monitoring.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs14153839/s1, Table S1: Morphometric parameters of the analyzed lakes. Table S2: Linear and random forest model performance per lake on train and test sets. Test lakes are marked with "x". Figure S1: Availability of satellite scenes depending on (a) lake; (b) month; (c) year; (d) water temperature (as intervals). Reference [86] is cited in the Supplementary Table S1.

Author Contributions: K.D.: methodology, software, validation, formal analysis, data curation, writing—original draft preparation, review, editing, and visualization. S.E.: software and writing—review, editing. M.P.: writing—original draft preparation, review, and editing. J.P.: writing—original draft preparation, review, and editing. M.S.: formal analysis, writing—original draft preparation, review, editing, and visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: In this study, we used publicly available datasets. Hydrologic data, satellite measurements, random forest regression model, and statistical analysis results are available in this repository: https://github.com/kadyb/lakes_temp (accessed on 1 May 2022). This repository also includes fully functional and reproducible codes under the MIT license.

Acknowledgments: The source of the hydrological data is the Institute of Meteorology and Water Management—National Research Institute (https://www.imgw.pl/). Landsat-8 images courtesy of the U.S. Geological Survey (https://earthexplorer.usgs.gov/). Both websites accessed on 1 May 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ptak, M.; Sojka, M.; Graf, R.; Choiński, A.; Zhu, S.; Nowak, B. Warming Vistula River—The Effects of Climate and Local Conditions on Water Temperature in One of the Largest Rivers in Europe. *J. Hydrol. Hydromech.* **2022**, *70*, 1–11. [CrossRef]
2. Hestir, E.L.; Brando, V.E.; Bresciani, M.; Giardino, C.; Matta, E.; Villa, P.; Dekker, A.G. Measuring Freshwater Aquatic Ecosystems: The Need for a Hyperspectral Global Mapping Satellite Mission. *Remote Sens. Environ.* **2015**, *167*, 181–195. [CrossRef]
3. Lieberherr, G.; Wunderle, S. Lake Surface Water Temperature Derived from 35 Years of AVHRR Sensor Data for European Lakes. *Remote Sens.* **2018**, *10*, 990. [CrossRef]
4. van Puijenbroek, P.J.T.M.; Evers, C.H.M.; van Gaalen, F.W. Evaluation of Water Framework Directive Metrics to Analyse Trends in Water Quality in the Netherlands. *Sustain. Water Qual. Ecol.* **2015**, *6*, 40–47. [CrossRef]
5. Birk, S.; Ecke, F. The Potential of Remote Sensing in Ecological Status Assessment of Coloured Lakes Using Aquatic Plants. *Ecol. Indic.* **2014**, *46*, 398–406. [CrossRef]
6. Livingstone, D.M.; Lotter, A.F. The Relationship between Air and Water Temperatures in Lakes of the Swiss Plateau: A Case Study with Palaeolimnological Implications. *J. Paleolimnol.* **1998**, *19*, 181–198. [CrossRef]
7. Schneider, P.; Hook, S.J. Space Observations of Inland Water Bodies Show Rapid Surface Warming since 1985. *Geophys. Res. Lett.* **2010**, *37*, L22405. [CrossRef]
8. Austin, J.; Colman, S. A Century of Temperature Variability in Lake Superior. *Limnol. Oceanogr.* **2008**, *53*, 2724–2730. [CrossRef]
9. Crosman, E.T.; Horel, J.D. MODIS-Derived Surface Temperature of the Great Salt Lake. *Remote Sens. Environ.* **2009**, *113*, 73–81. [CrossRef]
10. George, G.D. Using Airborne Remote Sensing to Study the Physical Dynamics of Lakes and the Spatial Distribution of Phytoplankton. *Freshw. Rev.* **2012**, *5*, 121–140. [CrossRef]
11. Schluessel, P.; Emery, W.J.; Grassl, H.; Mammen, T. On the Bulk-Skin Temperature Difference and Its Impact on Satellite Remote Sensing of Sea Surface Temperature. *J. Geophys. Res.* **1990**, *95*, 13341. [CrossRef]
12. Wick, G.A.; Emery, W.J.; Kantha, L.H.; Schlüssel, P. The Behavior of the Bulk—Skin Sea Surface Temperature Difference under Varying Wind Speed and Heat Flux. *J. Phys. Oceanogr.* **1996**, *26*, 1969–1988. [CrossRef]
13. Rozenstein, O.; Qin, Z.; Derimian, Y.; Karnieli, A. Derivation of Land Surface Temperature for Landsat-8 TIRS Using a Split Window Algorithm. *Sensors* **2014**, *14*, 5768–5780. [CrossRef]
14. Politi, E.; Cutler, M.E.J.; Rowan, J.S. Using the NOAA Advanced Very High Resolution Radiometer to Characterise Temporal and Spatial Trends in Water Temperature of Large European Lakes. *Remote Sens. Environ.* **2012**, *126*, 1–11. [CrossRef]
15. Tavares, M.; Cunha, A.; Motta-Marques, D.; Ruhoff, A.; Cavalcanti, J.; Fragoso, C.; Martín Bravo, J.; Munar, A.; Fan, F.; Rodrigues, L. Comparison of Methods to Estimate Lake-Surface-Water Temperature Using Landsat 7 ETM+ and MODIS Imagery: Case Study of a Large Shallow Subtropical Lake in Southern Brazil. *Water* **2019**, *11*, 168. [CrossRef]
16. Cheval, S.; Popa, A.-M.; Șandric, I.; Iojă, I.-C. Exploratory Analysis of Cooling Effect of Urban Lakes on Land Surface Temperature in Bucharest (Romania) Using Landsat Imagery. *Urban Clim.* **2020**, *34*, 100696. [CrossRef]
17. Giardino, C.; Pepe, M.; Brivio, P.A.; Ghezzi, P.; Zilioli, E. Detecting Chlorophyll, Secchi Disk Depth and Surface Temperature in a Sub-Alpine Lake Using Landsat Imagery. *Sci. Total Environ.* **2001**, *268*, 19–29. [CrossRef]
18. Malakar, N.K.; Hulley, G.C.; Hook, S.J.; Laraby, K.; Cook, M.; Schott, J.R. An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5717–5735. [CrossRef]
19. Prats, J.; Reynaud, N.; Rebière, D.; Peroux, T.; Tormos, T.; Danis, P.-A. LakeSST: Lake Skin Surface Temperature in French Inland Water Bodies for 1999–2016 from Landsat Archives. *Earth Syst. Sci. Data* **2018**, *10*, 727–743. [CrossRef]
20. Simon, R.N.; Tormos, T.; Danis, P.-A. Retrieving Water Surface Temperature from Archive LANDSAT Thermal Infrared Data: Application of the Mono-Channel Atmospheric Correction Algorithm over Two Freshwater Reservoirs. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 247–250. [CrossRef]
21. Becker, M.; Daw, A. Influence of Lake Morphology and Clarity on Water Surface Temperature as Measured by EOS ASTER. *Remote Sens. Environ.* **2005**, *99*, 288–294. [CrossRef]
22. Kay, J.E.; Kampf, S.K.; Handcock, R.N.; Cherkauer, K.A.; Gillespie, A.R.; Burges, S.J. Accuracy of Lake and Stream Temperatures Estimated from Thermal Infrared Images. *J. Am. Water Resour. Assoc.* **2005**, *41*, 1161–1175. [CrossRef]
23. Sabol, D.E., Jr.; Gillespie, A.R.; Abbott, E.; Yamada, G. Field Validation of the ASTER Temperature–Emissivity Separation Algorithm. *Remote Sens. Environ.* **2009**, *113*, 2328–2344. [CrossRef]
24. Sharma, S.; Walker, S.C.; Jackson, D.A. Empirical Modelling of Lake Water-Temperature Relationships: A Comparison of Approaches. *Freshw. Biol.* **2008**, *53*, 897–911. [CrossRef]
25. Edinger, J.E.; Duttweiler, D.W.; Geyer, J.C. The Response of Water Temperatures to Meteorological Conditions. *Water Resour. Res.* **1968**, *4*, 1137–1143. [CrossRef]
26. Oswald, C.J.; Rouse, W.R. Thermal Characteristics and Energy Balance of Various-Size Canadian Shield Lakes in the Mackenzie River Basin. *J. Hydrometeorol.* **2004**, *5*, 129–144. [CrossRef]
27. Snucins, E.; John, G. Interannual Variation in the Thermal Structure of Clear and Colored Lakes. *Limnol. Oceanogr.* **2000**, *45*, 1639–1646. [CrossRef]
28. Choiński, A.; Ptak, M.; Strzelczak, A. Changeability of Accumulated Heat Content in Alpine-Type Lakes. *Pol. J. Environ. Stud.* **2015**, *24*, 2363–2369. [CrossRef]

29. Ptak, M.; Sojka, M.; Kozłowski, M. The Increasing of Maximum Lake Water Temperature in Lowland Lakes of Central Europe: Case Study of the Polish Lakeland. *Ann. Limnol. Int. J. Limnol.* **2019**, *55*, 6. [CrossRef]
30. Heddam, S.; Ptak, M.; Zhu, S. Modelling of Daily Lake Surface Water Temperature from Air Temperature: Extremely Randomized Trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. *J. Hydrol.* **2020**, *588*, 125130. [CrossRef]
31. De Santis, D.; Del Frate, F.; Schiavon, G. Analysis of Climate Change Effects on Surface Temperature in Central-Italy Lakes Using Satellite Data Time-Series. *Remote Sens.* **2021**, *14*, 117. [CrossRef]
32. Huovinen, P.; Ramírez, J.; Caputo, L.; Gómez, I. Mapping of Spatial and Temporal Variation of Water Characteristics through Satellite Remote Sensing in Lake Panguipulli, Chile. *Sci. Total Environ.* **2019**, *679*, 196–208. [CrossRef]
33. Irani Rahaghi, A.; Lemmin, U.; Barry, D.A. Surface Water Temperature Heterogeneity at Subpixel Satellite Scales and Its Effect on the Surface Cooling Estimates of a Large Lake: Airborne Remote Sensing Results From Lake Geneva. *J. Geophys. Res. Ocean.* **2019**, *124*, 635–651. [CrossRef]
34. Yu, Z.; Yang, K.; Luo, Y.; Wang, P.; Yang, Z. Research on the Lake Surface Water Temperature Downscaling Based on Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5550–5558. [CrossRef]
35. Ptak, M.; Choiński, A.; Piekarczyk, J.; Pryłowski, T. Applying Landsat Satellite Thermal Images in the Analysis of Polish Lake Temperatures. *Pol. J. Environ. Stud.* **2017**, *26*, 2159–2165. [CrossRef]
36. Ptak, M.; Sojka, M.; Choiński, A.; Nowak, B. Effect of Environmental Conditions and Morphometric Parameters on Surface Water Temperature in Polish Lakes. *Water* **2018**, *10*, 580. [CrossRef]
37. Czernecki, B.; Ptak, M. The Impact of Global Warming on Lake Surface Water Temperature in Poland—The Application of Empirical-Statistical Downscaling, 1971–2100. *J. Limnol.* **2018**, *77*, 330–348. [CrossRef]
38. Piccolroaz, S.; Zhu, S.; Ptak, M.; Sojka, M.; Du, X. Warming of Lowland Polish Lakes under Future Climate Change Scenarios and Consequences for Ice Cover and Mixing Dynamics. *J. Hydrol. Reg. Stud.* **2021**, *34*, 100780. [CrossRef]
39. Ermida, S.L.; Soares, P.; Mantas, V.; Göttsche, F.-M.; Trigo, I.F. Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series. *Remote Sens.* **2020**, *12*, 1471. [CrossRef]
40. Ptak, M.; Sojka, M.; Nowak, B. Characteristics of Daily Water Temperature Fluctuations in Lake Kierskie (West Poland). *Quaest. Geogr.* **2019**, *38*, 41–49. [CrossRef]
41. Ptak, M.; Nowak, B. Variability of Oxygen-Thermal Conditions in Selected Lakes in Poland. *Ecol. Chem. Eng.* **2016**, *23*, 639–650. [CrossRef]
42. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
43. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]
44. Zhu, Z.; Woodcock, C.E. Object-Based Cloud and Cloud Shadow Detection in Landsat Imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]
45. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4–7, 8, and Sentinel 2 Images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [CrossRef]
46. *USGS Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide*; United States Geological Survey: Asheville, NC, USA, 2022; pp. 1–42.
47. Cook, M. *Atmospheric Compensation for a Landsat Land Surface Temperature Product*; Rochester Institute of Technology: Rochester, NY, USA, 2014.
48. Cook, M.; Schott, J.; Mandel, J.; Raqueno, N. Development of an Operational Calibration Methodology for the Landsat Thermal Data Archive and Initial Testing of the Atmospheric Compensation Component of a Land Surface Temperature (LST) Product from the Archive. *Remote Sens.* **2014**, *6*, 11244–11266. [CrossRef]
49. Schaeffer, B.A.; Iiames, J.; Dwyer, J.; Urquhart, E.; Salls, W.; Rover, J.; Seegers, B. An Initial Validation of Landsat 5 and 7 Derived Surface Water Temperature for U.S. Lakes, Reservoirs, and Estuaries. *Int. J. Remote Sens.* **2018**, *39*, 7789–7805. [CrossRef]
50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
51. Ganesh, N.; Jain, P.; Choudhury, A.; Dutta, P.; Kalita, K.; Barsocchi, P. Random Forest Regression-Based Machine Learning Model for Accurate Estimation of Fluid Flow in Curved Pipes. *Processes* **2021**, *9*, 2095. [CrossRef]
52. Seo, D.; Kim, Y.; Eo, Y.; Park, W.; Park, H. Generation of Radiometric, Phenological Normalized Image Based on Random Forest Regression for Change Detection. *Remote Sens.* **2017**, *9*, 1163. [CrossRef]
53. Rouse, J.W., Jr.; Haas, R.H.; Deering, D.W.; Schell, J.A.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; Texas A&M University: College Station, TX, USA, 1974.
54. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
55. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [CrossRef]
56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2022.

57. Duguay-Tetzlaff, A.; Bento, V.; Göttsche, F.; Stöckli, R.; Martins, J.; Trigo, I.; Olesen, F.; Bojanowski, J.; da Camara, C.; Kunz, H. Meteosat Land Surface Temperature Climate Data Record: Achievable Accuracy and Potential Uncertainties. *Remote Sens.* **2015**, *7*, 13139–13156. [CrossRef]
58. Hijmans, R.J. *Terra: Spatial Data Analysis*, USA, 2022.
59. Skowron, R. *Zróżnicowanie i Zmienność Wybranych Elementów Reżimu Termicznego Wody w Jeziorach Na Niżu Polskim*; Uniwersytetu Mikołaja Kopernika: Toruń, Poland, 2011; ISBN 978-83-231-2695-9.
60. Barsi, J.; Schott, J.; Hook, S.; Raqueno, N.; Markham, B.; Radocinski, R. Landsat-8 Thermal Infrared Sensor (TIRS) Vicarious Radiometric Calibration. *Remote Sens.* **2014**, *6*, 11607–11626. [CrossRef]
61. Hashim, N.S.; Windupranata, W.; Sulaiman, S.A.H. Shallow-Water Bathymetry Estimation Using Single Band Algorithm and Green Band Algorithm. In Proceedings of the 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 8 August 2020; pp. 214–219.
62. Jia, X.; Willard, J.; Karpatne, A.; Read, J.S.; Zwart, J.A.; Steinbach, M.; Kumar, V. Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. *ACMIMS Trans. Data Sci.* **2021**, *2*, 1–26. [CrossRef]
63. Magee, M.R.; Wu, C.H.; Robertson, D.M.; Lathrop, R.C.; Hamilton, D.P. Trends and Abrupt Changes in 104 Years of Ice Cover and Water Temperature in a Dimictic Lake in Response to Air Temperature, Wind Speed, and Water Clarity Drivers. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 1681–1702. [CrossRef]
64. Binding, C.E.; Zastepa, A.; Zeng, C. The Impact of Phytoplankton Community Composition on Optical Properties and Satellite Observations of the 2017 Western Lake Erie Algal Bloom. *J. Gt. Lakes Res.* **2019**, *45*, 573–586. [CrossRef]
65. Damtew, Y.T.; Verbeiren, B.; Awoke, A.; Triest, L. Satellite Imageries and Field Data of Macrophytes Reveal a Regime Shift of a Tropical Lake (Lake Ziway, Ethiopia). *Water* **2021**, *13*, 396. [CrossRef]
66. Qing, S.; Runa, A.; Shun, B.; Zhao, W.; Bao, Y.; Hao, Y. Distinguishing and Mapping of Aquatic Vegetations and Yellow Algae Bloom with Landsat Satellite Data in a Complex Shallow Lake, China during 1986–2018. *Ecol. Indic.* **2020**, *112*, 106073. [CrossRef]
67. Hao, X.; Yang, Q.; Shi, X.; Liu, X.; Huang, W.; Chen, L.; Ma, Y. Fractal-Based Retrieval and Potential Driving Factors of Lake Ice Fractures of Chagan Lake, Northeast China Using Landsat Remote Sensing Images. *Remote Sens.* **2021**, *13*, 4233. [CrossRef]
68. Khanesar, M.A.; Branson, D.T. Prediction Interval Identification Using Interval Type-2 Fuzzy Logic Systems: Lake Water Level Prediction Using Remote Sensing Data. *IEEE Sens. J.* **2021**, *21*, 13815–13827. [CrossRef]
69. Shu, S.; Liu, H.; Beck, R.A.; Frappart, F.; Korhonen, J.; Lan, M.; Xu, M.; Yang, B.; Huang, Y. Evaluation of Historic and Operational Satellite Radar Altimetry Missions for Constructing Consistent Long-Term Lake Water Level Records. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1643–1670. [CrossRef]
70. Aguilar-Lome, J.; Soca-Flores, R.; Gómez, D. Evaluation of the Lake Titicaca's Surface Water Temperature Using LST MODIS Time Series (2000–2020). *J. S. Am. Earth Sci.* **2021**, *112*, 103609. [CrossRef]
71. Du, J.; Jacinthe, P.; Zhou, H.; Xiang, X.; Zhao, B.; Wang, M.; Song, K. Monitoring of Water Surface Temperature of Eurasian Large Lakes Using MODIS Land Surface Temperature Product. *Hydrol. Process.* **2020**, *34*, 3582–3595. [CrossRef]
72. Kishcha, P.; Starobinets, B.; Lechinsky, Y.; Alpert, P. Absence of Surface Water Temperature Trends in Lake Kinneret despite Present Atmospheric Warming: Comparisons with Dead Sea Trends. *Remote Sens.* **2021**, *13*, 3461. [CrossRef]
73. Moukomla, S.; Blanken, P. Remote Sensing of the North American Laurentian Great Lakes' Surface Temperature. *Remote Sens.* **2016**, *8*, 286. [CrossRef]
74. Wan, W.; Li, H.; Xie, H.; Hong, Y.; Long, D.; Zhao, L.; Han, Z.; Cui, Y.; Liu, B.; Wang, C.; et al. A Comprehensive Data Set of Lake Surface Water Temperature over the Tibetan Plateau Derived from MODIS LST Products 2001–2015. *Sci. Data* **2017**, *4*, 170095. [CrossRef]
75. Szkwarek, K.; Wochna, A. Wyznaczanie temperatury powierzchniowej jeziora Raduńskiego Górnego na podstawie zdjęć satelitarnych Landsat 8. *Tutor. Gedanensis* **2020**, *5*, 41–45.
76. Sharaf, N.; Fadel, A.; Bresciani, M.; Giardino, C.; Lemaire, B.J.; Slim, K.; Faour, G.; Vinçon-Leite, B. Lake Surface Temperature Retrieval from Landsat-8 and Retrospective Analysis in Karaoun Reservoir, Lebanon. *J. Appl. Remote Sens.* **2019**, *13*, 044505. [CrossRef]
77. Martinsen, K.T.; Andersen, M.R.; Sand-Jensen, K. Water Temperature Dynamics and the Prevalence of Daytime Stratification in Small Temperate Shallow Lakes. *Hydrobiologia* **2019**, *826*, 247–262. [CrossRef]
78. Tokuda, D.; Kim, H.; Yamazaki, D.; Oki, T. Development of a Global River Water Temperature Model Considering Fluvial Dynamics and Seasonal Freeze-Thaw Cycle. *Water Resour. Res.* **2019**, *55*, 1366–1383. [CrossRef]
79. Huang, Y.; Liu, H.; Hinkel, K.; Yu, B.; Beck, R.; Wu, J. Analysis of Thermal Structure of Arctic Lakes at Local and Regional Scales Using in Situ and Multidate Landsat-8 Data. *Water Resour. Res.* **2017**, *53*, 9642–9658. [CrossRef]
80. Wrzesiński, D.; Brychczyński, A. Zróżnicowanie reżimu odpływu rzek w północno-zachodniej Polsce. *Bad. Fizjogr.* **2014**, *65*, 261–274.
81. Cao, B.; Kang, L.; Yang, S. *Retrieval of Lake Water Temperature Based on LandSat TM Imagery: A Case Study in East Lake of Wuhan*; Tian, J., Ma, J., Eds.; SPIE: Bellingham, WA, USA, 2013; p. 892108.
82. Luciani, G.; Bresciani, M.; Biraghi, C.A.; Ghirardi, N.; Carrion, D.; Rogora, M.; Brovelli, M.A. Satellite Monitoring System of Subalpine Lakes with Open Source Software: The Case of SIMILE Project. *Balt. J. Mod. Comput.* **2021**, *9*, 135–144. [CrossRef]
83. Sener, S.; Sener, E. Estimation of Lake Water Temperature with ASTER and Landsat 8 OLI-TIRS Thermal Infrared Bands: A Case Study Beysehir Lake (Turkey). *Living Planet Symp.* **2016**, *740*, 256.

84. Aïtelghazi, A.; Rhinane, H.; Bensalmia, A.; Giuliani, G. Using the Landsat-7 Data to Study the Correlation between the Surface Temperature and Phytoplankton Turbidity Case Study: Al Massira Lake (Settat-Morocco). *Mater. Today Proc.* **2019**, *13*, 496–504. [CrossRef]
85. Chao Rodríguez, Y.; el Anjoumi, A.; Domínguez Gómez, J.A.; Rodríguez Pérez, D.; Rico, E. Using Landsat Image Time Series to Study a Small Water Body in Northern Spain. *Environ. Monit. Assess.* **2014**, *186*, 3511–3522. [CrossRef]
86. Choiński, A. *Katalog jezior Polski*; Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza: Poznań, Poland, 2006.

scientific reports

Check for updates

OPEN

Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland

Krzysztof Dyba

In recent years, automatic image classification methods have significantly progressed, notably black box algorithms such as machine learning and deep learning. Unfortunately, such efforts only focused on improving performance, rather than attempting to explain and interpret how classification models actually operate. This article compares three state-of-the-art algorithms incorporating random forests, gradient boosting and convolutional neural networks for geomorphological mapping. It also attempts to explain how the most effective classifier makes decisions by evaluating which of the geomorphometric variables are most important for automatic mapping and how they affect the classification results using one of the explainable artificial intelligence techniques, namely accumulated local effects (ALE). This method allows us to understand the relationship between predictors and the model's outcome. For these purposes, eight sheets of the digital geomorphological map of Poland on the scale of 1:100,000 were used as the reference material. The classification results were validated using the holdout method and cross-validation for individual sheets representing different morphogenetic zones. The terrain elevation entropy, absolute elevation, aggregated median elevation and standard deviation of elevation had the greatest impact on the classification results among the 15 geomorphometric variables considered. The ALE analysis was conducted for the XGBoost classifier, which achieved the highest accuracy of 92.8%, ahead of Random Forests at 84% and LightGBM at 73.7% and U-Net at 59.8%. We conclude that automatic classification can support geomorphological mapping only if the geomorphological characteristics in the predicted area are similar to those in the training dataset. The ALE plots allow us to analyze the relationship between geomorphometric variables and landform membership, which helps clarify their role in the classification process.

Geomorphology is a scientific discipline that studies landforms, their features and the processes that shape them[1]. One of the key aspects of geomorphology is the mapping process, which involves identifying landforms and determining their spatial distribution in the context of processes occurring on the Earth's surface[2]. Traditional and automatic mapping are two different approaches to mapping landforms based on their features, shape, and spatial distribution.

Traditional geomorphological mapping is based on fieldwork and manual interpretation of various data sources (for example, digital elevation models, topographic maps, aerial or satellite imagery), which requires a high level of expertise and experience. Therefore, this approach is time consuming and expensive. Another debatable issue is the repeatability of mapping results related to the subjective nature of interpretation, which can consequently lead to different divisions and ranges of landforms or soil units[3].

On the other hand, automatic geomorphological mapping can be more efficient and cheaper, and most importantly, can provide reproducible results by removing the aspect of subjectivity. Basically, three different approaches to automatic classification can be distinguished, i.e., the pixel-based[4–8], object-based[9,10] and pattern-based. The first two are currently used as state-of-the-art, but the last one is new and requires further research.

The pattern approach mainly relies on convolutional neural networks (CNNs), which involve a multi-step learning process using convolutional layers to create a feature map that extracts certain image patterns. CNNs

Applied Geoinformatics Research Unit, Adam Mickiewicz University, Bogumiła Krygowskiego 10, 61-680 Poznań, Poland. email: krzysztof.dyba@amu.edu.pl

have become very popular in computer vision due to their high efficiency in identifying low-level features and patterns, making them very effective for data classification[11,12].

Recent research on the application of convolutional neural networks in geomorphology includes the use of a multi-channel deep neural network architecture to classify landforms[13], a comparison of Random Forests and U-Net models to classify loess formations[14], a comparison between traditional and automated U-Net-based approaches[15], and classification using textural properties of the terrain[16].

So far, several initiatives have been undertaken to develop high-resolution digital geomorphological maps of selected areas in Poland based on traditional mapping, including Roztocze Upland[17], Pomeranian and Warmian–Masurian voivodeships[18], Mazovia[19], Carpathians[20], Narew National Park[21], Wielkopolska–Kujawy Lowland, Mysliborsk Lakeland and Szczecin Lowland[22], Podlasie[23], and Tykocin[24]. Nevertheless, the mentioned studies were conducted by independent research teams and are not unified, thus they have different catalogs of landforms, mapping principles and spatial scales.

However, research on automatic classification of the geomorphological landforms in Poland remains at an early stage. The first study compares unsupervised automatic classification with the traditional mapping for the Sudetes[25]. The second study also concerns unsupervised classification for the area of the Silesian Upland[26]. Another study on supervised classification was conducted by Janowski et al.[27], in which the authors compared machine learning algorithms for classifying glacial landforms in the Lubawa Upland and Gardno–Leba Plain areas using ground truth dataset. In a previous article co-written by the present author, we clustered the landforms of the entire country using an unsupervised method[28]. This means that we made no prior assumptions about geomorphological units. Finally, we separated 20 land surface types in the process of interpreting and labeling clusters.

The first objective of this article is to perform a supervised classification using machine learning based on the available sheets of the digital geomorphological map of Poland. Unlike the unsupervised approach, the catalog of geomorphological units is known in advance, but the problem is to map it as best as possible using an automatic classification method. The second objective is to interpret the classification decisions made by the model, in particular to explain which geomorphometric variables are most relevant and how they affect the classification results.

Materials and methods

We divided this section into several subsections to clearly present the extensively used materials and methods. Section "Digital geomorphological map" describes the digital geomorphological map of Poland. Section "Morphometric variables" provides information on geomorphometric explanatory variables and how they are processed. Section "Selection of a classification model" presents the machine learning and neural network models employed, while Section "Validation" presents the methods and metrics for their validation. Section "Model explanation" describes the method to explain the classifier's decision. Finally, Section "Software" contains technical information about the software used.

Digital geomorphological map

The digital geomorphological map of Poland on the scale of 1:100,000 is a vector map showing the forms of relief and the genesis of the Earth's surface alongside information about its formation[29]. The color scheme is based on the Gustavsson et al.[30] concept with modifications. Eight available sheets with a total area of 9072 km^2 were used as a reference dataset (Fig. 1). Currently, it is the only such detailed and up-to-date source on a national scale

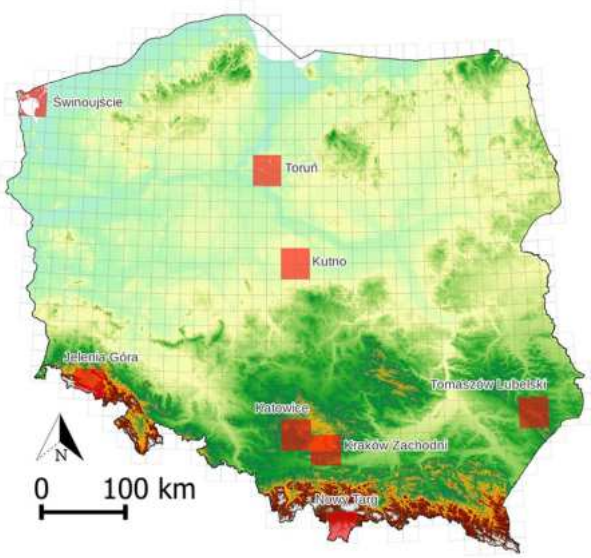

Figure 1. Sheet coverage of the digital geomorphological map of Poland.

with uniform principles of development. The landforms presented are from all morphogenetic zones, including the coastal area (Świnoujście), the young and old glacial areas (Toruń and Kutno), the upland areas (Katowice, Kraków Zachodni and Tomaszów Lubelski), and the areas of young and old mountains (Jelenia Góra and Nowy Targ). In the technical manual there are 77 surface divisions in 10 morphogenetic groups; however, only 54 divisions can be found on the available sheets. The landforms are listed in Supplementary Fig. S1.

The representativeness of the morphological forms is strongly unbalanced; for instance, the slope surface landform accounts for more than 23% of the total dataset, while the other 43 landforms represent less than 15% (Fig. 2). This issue is a major problem in automatic classification methods. This means that the algorithm is unable to learn how to correctly classify forms that are a significant minority (permille) in the dataset. To address this problem, we reduced the size of the 14 largest classes to 150,000 observations using the data under-sampling procedure[31] and removed the two least numerous classes (beach and dune plain). The second issue relates to missing values (*NA*) that result from areas not covered by mapping or water surfaces (Fig. 2). In the case of machine learning algorithms, typically missing values can be omitted (they will not be included in the training set), while neural networks use them in the learning process, and then they are masked (excluded). The final dataset consisted of over 3.3 million observations (pixels).

Machine learning algorithms require a discrete representation of data, for this reason we rasterized vector maps to a resolution of 30 m in the Polish geodetic coordinate system 1992 (EPSG: 2180). For this purpose, we created a classification table that contained the original category names encoded as text and their corresponding IDs in numerical form. We coded missing data (*NA*), water reservoirs, and areas not surveyed with a value of 0.

Morphometric variables

As the main data source, we used a digital terrain model with a resolution of 30 m adapted from Digital Terrain Elevation Data Level 2 (Fig. 3). The data has been smoothed and resampled, so the artifacts (noise strips) seen in the original do not appear[32]. Then we generated a number of derivative products based on it.

More than 100 different geomorphometric variables can be found in popular applications for geomorphometric analysis. It is impossible to include all of them for technical reasons (hardware limitations, processing time) and analytical reasons (some are strongly correlated). Therefore, we considered the 15 most commonly used and made a final selection of the most important features for classification using model performance metrics, visual inspection and model information gain (Fig. 4). Eventually, we reduced their number to 9 (Table 1), which increased the performance of the classifier (i.e., faster training and prediction, and reduced memory consumption) and, most importantly, simplified the structure of the model, thus making its decisions easier to interpret. We initially tested absolute elevation, multidirectional hillshade[33], slope[34], topographic position index[35], multi-scale topographic position index, tangential and profile curvatures[34], convergence index with radius[36], terrain surface texture[6], terrain surface convexity[6], topographic openness[37], aggregated elevation, local standard deviation, and textural features including contrast, energy and entropy[38].

The aggregated elevation was calculated using a statistic (in this case, the median) from neighboring pixels at a lower spatial resolution (500 and 1000 m respectively), and then the aggregated cell was divided into smaller blocks corresponding to a resolution of 30 m. If it was possible to set the analysis radius, we set it to 16 pixels (representing an area of about 0.7 km^2). Additionally, we removed the variables above a linear correlation of 0.9 because they essentially convey the same information except for aggregated elevation and entropy (they are perfectly correlated with absolute elevation but contain information on a larger spatial scale), which allows for

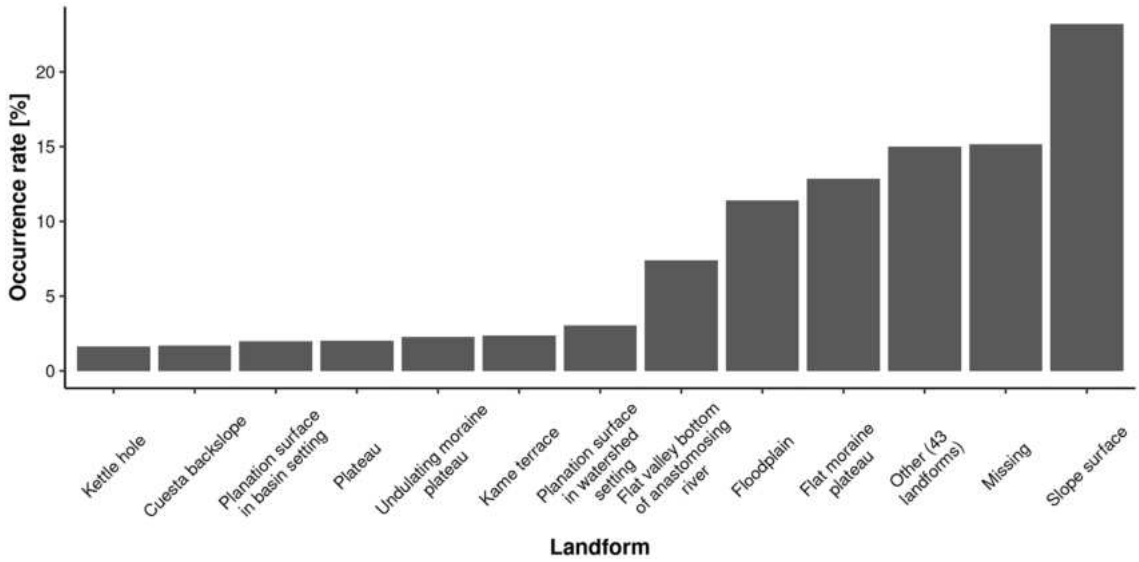

Figure 2. Distribution of the geomorphological forms from the used sheets. All forms with a total area of less than 166 km^2 (i.e., the 80th percentile) are combined into one category in the figure: Other.

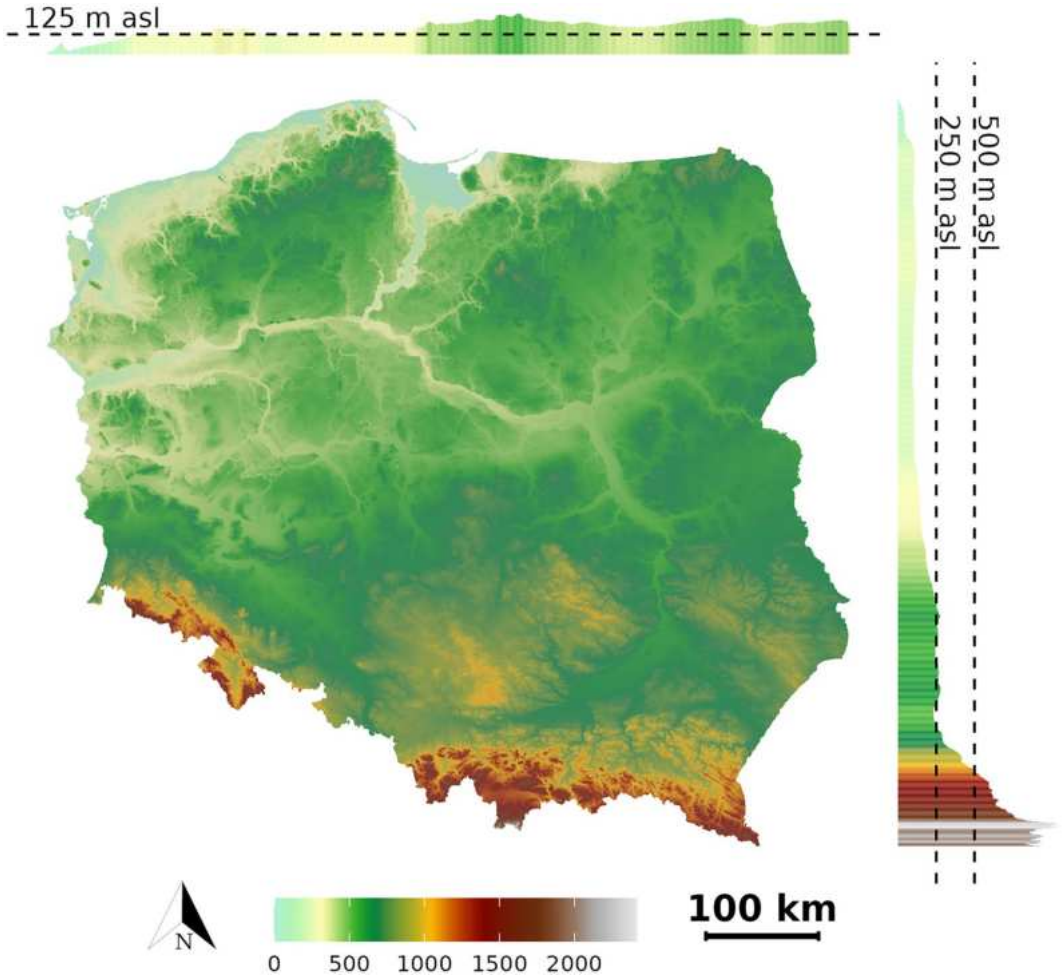

Figure 3. Elevation map of Poland with hillshading. Histograms with average elevation values calculated for latitudes and longitudes are seen on the sides.

mapping morphological objects with continuity. The geomorphometric variables used in this study are presented in Fig. 5.

Selection of a classification model

We compared the three most popular models based on machine learning—Random Forests[39] and gradient boosting including XGBoost[40] and LightGBM[41], and the convolutional neural network (CNN) model—U-Net[42]. The main difference between random forests and gradient boosting is that the former reduces the variance of a large number of complex models with low bias (the models are built independently and parallelly), while the gradient boosting reduces the bias of a large number of simple models with low variance (the models depend on each other because each is based on all previous small models with the appropriate weight, hence the name "boosting"). Both XGBoost and LightGBM models are based on gradient boosting, but the former uses an "exact" algorithm, while the latter uses an "approximate" algorithm (observations with similar values are aggregated into bins). This acts as a compromise between performance and accuracy of trained models.

In contrast, convolutional neural networks are primarily dedicated to computer vision, whereas machine learning models focus on modeling tabular data. They consist of multiple layers that can extract vital image features (such as edges) and reduce the spatial resolution while retaining the important information. Thus, it can be expected that the recognition of spatial patterns will be independent of the shift (i.e., the model will be able to recognize the same pattern in a different place) and the spatial hierarchy of objects will be considered (for example, the first layer of the network will learn to recognize small local patterns, and the next layer will aggregate them into larger structures). In the case of machine learning, this is not directly possible, and the data must be prepared in an appropriate way (feature engineering).

Model hyperparameters tuning

Machine learning models require a predeterminating of the hyperparameters such as maximum tree depth, number of leaves (nodes), learning rate, etc. to be effective. This procedure is called tuning. In order to find the most effective hyperparameters, we used a random search procedure, which involves defining a search space

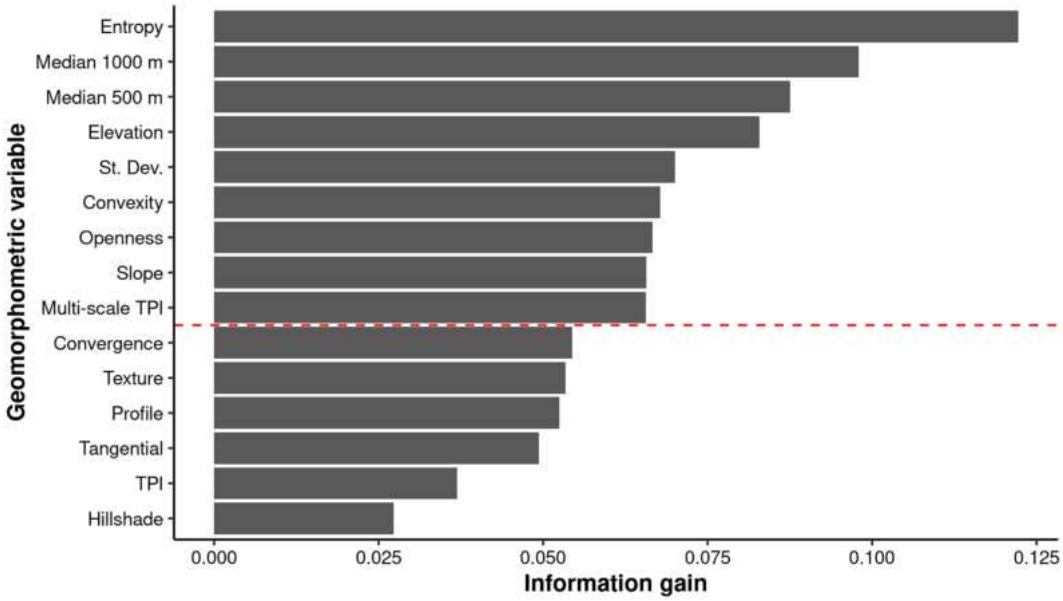

Figure 4. Importance of the geomorphometric variables for the geomorphological classification using the XGBoost model. The higher the value, the greater the suitability. The red dashed line indicates discarded low-significance variables.

#	Variable	Range	Mean value	Unit
1	Absolute elevation	− 0.03, 2483	171 ± 129.2	m
2	Slope	0, 76.1	1.75 ± 3.08	deg
3	Local standard deviation 1000 m	0, 262.7	5 ± 8.3	m
4	Multi-scale topographic position index	− 53.1, 44.6	0 ± 0.7	m
5	Terrain surface convexity	0, 88.6	48.7 ± 7.5	–
6	Entropy	7, 940318	52,441 ± 44,508	–
7	Topographic openness	0.61, 1.7	1.55 ± 0.03	–
8	Median elevation 500 m	0, 2335	171 ± 128.9	m
9	Median elevation 1000 m	0, 2238	170 ± 128.5	m

Table 1. Morphometric variables used in this study.

(combination) of hyperparameters and random sampling. Then, using the drawn combination, the model is trained, and its performance is evaluated on an independent dataset.

We considered the following hyperparameters in this procedure: *eta* (step size shrinkage used to prevent overfitting), *max_depth* (maximal tree depth), *nrounds* (number of iterations), *subsample* (subsample ratio of the training data) in XGBoost; *num.trees* (number of trees to grow), *mtry* (number of variables randomly sampled as candidates at each node split), *min.node.size* (minimal node size), *max.depth* in Random Forests; *learning_rate* (performs the same function as *eta* in XGBoost), *max_bin* (maximum number of bins that feature values will be bucketed in), *num_leaves* (maximum number of leaves in one tree), *nrounds*, *max_depth, bagging_fraction* (performs the same function as *subsample* in XGBoost*), feature_fraction* (ratio of variables randomly sampled for each tree) in LightGBM. Appropriate selection of these hyperparameters prevents the model from overfitting the training data.

It should be emphasized that the applied machine learning algorithms do not use the pixel neighborhood, so information about the shape and continuity of geomorphological forms is not included. In fact, information about the values of the geomorphometric variables is only used for individual pixels. We used feature engineering, to address this problem, which is based on three elements: 1) calculation of the geomorphometric variables in the local window (if it was possible); 2) use of selected geomorphometric variables at lower spatial resolution to detect larger landforms; 3) use image textural features. See Section "Morphometric variables" for more details.

Convolutional neural network

We evaluated the convolutional neural networks using the U-Net model in Tensorflow[43]. This architecture consists of two main components, i.e., a contracting path (encoder) and an expansive path (decoder). The former

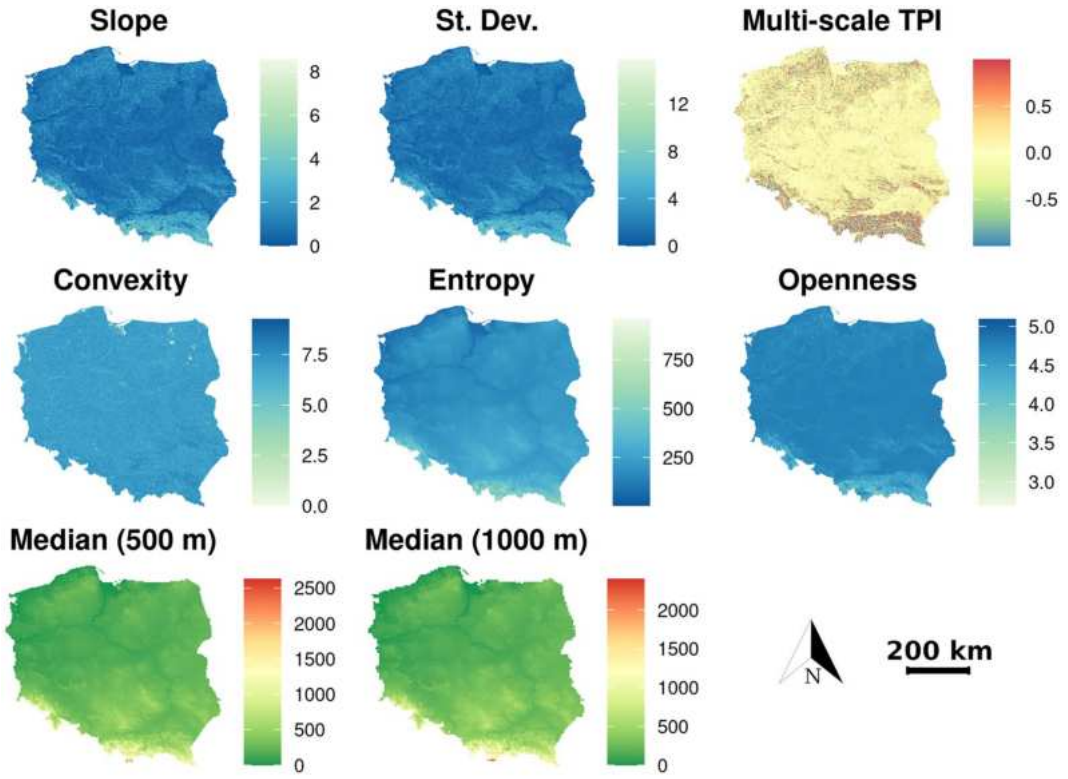

Figure 5. Geomorphometric variables used in this study. The absolute elevation is shown in Fig. 3 with the topographic color scale. The aggregated by median elevations of 500 and 1000 m look almost identical, but actually represent different spatial scales. The slope, standard deviation, convexity, entropy, openness variables are scaled by the square root, and the multi-scale TPI by the sine to better represent spatial variability.

progressively reduces the spatial resolution of the input image while increasing the number of features. The expansive path is the inverse of the contracting path and involves upsampling operations to restore the spatial resolution and reduce the number of features. The final layer consists of a convolutional layer with a softmax activation function, producing pixel-wise class predictions. The detailed architecture is shown in the original article by Ronneberger et al. in Fig. 1[42].

Several tile sizes were tested as input, i.e., 32×32, 64×64, 128×128 and 256×256 pixels. Finally, the most satisfactory results were obtained for blocks of 128×128 pixels due to the compromise between capturing spatial patterns by the model and the number of missing values in the tiles. To solve the problem of a large percentage of missing data, we removed those blocks for which the number of missing values was more than 70%, ultimately resulting in a total of 685 raster blocks. In order to increase the amount of input data, data augmentation was applied by flipping images in the vertical and horizontal planes. Adam's algorithm was used as the optimization function[44]. It should be noted that deep learning models have hundreds of thousands of parameters for tuning and, therefore, require much more input data compared to machine learning models. In this study, U-Net was used only as a reference method and its lower performance is expected compared to the other models tested.

Validation

We used holdout validation to validate the results—30% of the randomly selected input dataset was used as a test set to calculate the models' performance metrics, i.e., accuracy, Cohen's kappa coefficient (κ) and Matthews correlation coefficient (φ). The former provides overly optimistic results for unbalanced datasets, but the second and third are corrected for this effect and offer more reliable results. However, because we under-sampled classes to balance our dataset, the difference between these metrics is insignificant. Moreover, we used fivefold cross-validation to test the accuracy of the most efficient classifier (i.e., XGBoost) in this study for individual areas. Note that non-spatial validation can produce somewhat biased results[45,46], and in order to evaluate the performance completely independently, new geomorphological sheets (i.e., those that have not been used to train the models) should be used.

Model explanation

The models used in this study are black box models. This means that the predictions and decisions they generate are not interpretable in a simple way. In other words, the high complexity of the algorithms causes difficulties in explaining how it actually works[47,48]. In order to understand which geomorphometric features the model uses to make decisions, we used the XGBoost gain metric (Fig. 4), which determines the improvement in model performance by adding a specific feature to the decision tree. Moreover, in addition to examining which variables

are most useful for mapping, we also evaluated the interactions between the classification results and each geomorphometric variable using accumulated local effects plots[49].

The accumulated local effect (ALE) is a machine learning interpretability method that allows gain insights into the model's behavior, identifying how features affect predictions. The ALE method is similar to the partial dependence plot[50], but is faster and more robust (i.e., it enables an analysis of the correlated variables). The former focuses on local effects that are calculated in small windows, while the PDP calculates average values. The resulting ALE plot shows how the model prediction changes as the particular feature value increases (assuming that the other features are fixed), enabling an examination of the relationship between a feature and the model's prediction. In practice, this helps identify interactions that are not evident by simply assessing the significance of the features like using XGBoost gain metric. To the best of the author's knowledge, this method has not been previously used to explain the decision making of the classification models in geomorphological mapping.

Software

The geomorphometric variables were generated in GRASS GIS 7.8.0[51] with default function parameters. The data analysis and machine learning parts were completed in R[52], while the neural networks were used in Python. In particular, the *stars* package was utilized for processing the raster data[53], and *sf* for the vector data[53]. Statistical metrics were implemented in the *yardstick*[54] package. The *ranger* package was used to train Random Forests[55]. Accumulated local effects plots were generated by the *ALEPlot* package[49].

The development of the models was very time-consuming. It took nearly two weeks of continuous computations to train all 1180 models on an AMD Ryzen 9 5900X with 128 GB RAM. The models were trained in parallel on 12 physical CPU cores. The largest part of the trained models were models based on LightGBM algorithm, because of the largest number of hyperparameters to be tuned compared to other machine learning models.

Results

Classification

As a result of the evaluation on an independent test dataset, the XGBoost model proved to be the best with an accuracy of 92.8%. It was followed by Random Forests with an accuracy of 80.4% and LightGBM with an accuracy of 73.7%. The worst performance was achieved by the U-Net model due to insufficient training data (Table 2). During model tuning, the highest performance of the classifiers was obtained for the following hyperparameters:

1. XGBoost: $eta = 0.2$; $max\_depth = 20$; $nrounds = 150$; $subsample = 0.6$
2. Random Forests: $num.trees = 1000$; $mtry = 5$; $min.node.size = 1$; $max.depth = 20$
3. LightGBM: $learning\_rate = 0.05$; $max\_bin = 2048$; $num\_leaves = 70$; $nrounds = 150$; $max\_depth = 15$; $bagging\_fraction = 1$; $feature\_fraction = 1$

Further performance of the models can be improved by using larger values for the *max_depth* and *max_bin* hyperparameters, but this actually results in overfitting on the test dataset.

The potential application of model fusion may be intriguing. This technique typically results in an overall improvement in classification performance using aggregated results from several different models. However, this is provided that all models offer similar and high prediction efficiency, which is not case in this study. Ultimately, this would reduce the quality of the prediction and, moreover, it would become impossible to explain the performance of the combined models.

From this point on, only the XGBoost classifier is subjected to further analysis because it achieved the best result compared to the other models. The analysis of models with lower performance is unjustified, especially in the context of explaining how geomorphometric variables influence landform classifications (i.e., misclassifications mean misinterpretations). In the evaluation of the predicted landforms on individual sheets, XGBoost recorded the best accuracy for Jelenia Góra at 96.2%, and the lowest for Tomaszów Lubelski at 88% (Table 3). The average accuracy value using cross-validation was over 93%, while the Kappa coefficient and Matthews correlation coefficient values were slightly lower. This demonstrates the high potential application, provided that the predicted landforms and distributions of geomorphometric variables are similar in both the test and training datasets. However, it is not possible to conclude that there is a strong correlation between the number of landforms and model accuracy—classification performance is rather related to the representativeness of the forms and the complexity of the spatial patterns associated with the geomorphological characteristics of the areas. Examples of the classifier's application are shown in Fig. 6. It is noteworthy that small landforms appear on the

Model	Accuracy	Kappa coefficient	Matthews correlation coefficient
Random Forests	0.840	0.830	0.830
XGBoost	0.928	0.917	0.917
LightGBM	0.737	0.724	0.725
U-Net	0.598	0.576	0.576

Table 2. Evaluation of the model classification performance. The higher the metric values, the better the model.

Sheet	Mesoregion	Number of landforms	Accuracy	Kappa coefficient
Świnoujście	Szczecin Coastland	19	0.961	0.957
Toruń	Chełmno–Dobrzyń Lakeland, Toruń–Eberswalde Ice Marginal Valley	25	0.913	0.908
Kutno	Central Masovia Lowland, Southern Wielkopolska Lowland	15	0.958	0.952
Jelenia Góra	Western Sudety, Mountains	19	0.962	0.958
Tomaszów Lubelski	Roztocze Upland, Sandomierz Basin	15	0.880	0.866
Katowice	Silesia Upland, Woźniki–Wieluń Upland	13	0.943	0.936
Kraków	Krakow–Częstochowa Upland, Kraków Gate	17	0.949	0.945
Nowy Targ	Orawa–Podhale Basin, Tatra Range	16	0.940	0.934

Table 3. Performance of the XGBoost models for individual sheets based on fivefold cross-validation. The physico-geographical mesoregions are defined based on Solon et al.[56] classification.

predicted rasters that are not visible on the geomorphological map. This may be due to the higher spatial resolution of the geomorphometric variables compared to the reference map (not necessarily the prediction errors).

Model explanation

We first assessed the overall relevance of each geomorphometric variable for mapping. Among the tested variables, entropy, median elevation and absolute elevation turned out to be the most significant for classification. Next ranked were standard deviation, convexity and openness, slope and multi-scale TPI. The least useful for mapping were texture, profile and tangential curvatures, TPI, and hillshade (Fig. 4). The last group of variables with the lowest importance was excluded from the final classification since they do not actually improve mapping results, but significantly increase processing time, require additional memory and cause greater model complexity.

In the next step, we deepened the analysis of the relationship between the used geomorphometric variables, and the probability of the landform specified with accumulated local effects. As an example, we chose the four well-representative landforms, for instance: (a) proluvial plain; (b) plateau; (c) rock wall/rock slope; (d) depositional scree slope (Fig. 7). All other geomorphological landforms are presented in Supplementary Fig. S2.

Figure 7 shows how the probability of landform affiliation changes depending on geomorphometric variables. The first landform proluvial plain (Fig. 7a) is an extensive sandy flat surface created as a result of the periglacial-fluvial accumulation process. It is noticeable, in this case, the greatest impact on the detection of this form is the openness feature, whose high values (above 1.55) indicate an open and flat surface. The other geomorphometric variables are not very significant. The second example is the plateau (Fig. 7b), which is usually characterized by an irregular surface and explicit hillsides. In this case, low values of two features, i.e., entropy (below 85,000) and openness (below 1.58), reduce the probability of classifying this form, while an increase in the value of the slope increases this probability (in particular, a slope above 10°). The last example is rock wall/rock slope (Fig. 7c) and depositional scree slope (Fig. 7d). The former is a very steep or rugged fragment of the surface with a high slope, in which the process of weathering and falling rock materials occurs, creating an accumulated rubble slope at the foot of the slope. The latter usually takes the form of a mound or heap that is composed of rock rubble from a rock wall/rock slope. To detect the rock wall/rock slope, the slope, standard deviation and multi-scale TPI variables are important, high values of which increase the probability of classifying this landform. However, in the case of the depositional scree slope, high values of slope (above 30°) and standard deviation (above 150 m) reduce the probability of classifying this landform. The probability of classifying this landform by the model increases with high entropy (above 750,000) and low values of openness (below 1.3), which is probably related to the size of the rock material that creates irregular (undulated) surfaces. These examples demonstrate the convergence of classification decisions made by the model and geomorphological knowledge.

We also considered how the landform area represented in the dataset relates to the variability of the impact on the classification probability calculated from accumulated local effects. For this purpose, we defined the amplitude as the difference between the influence that increases the probability of being classified in a given class (the maximum is 1) and the influence that reduces this probability (the minimum is − 1), so the maximum amplitude can be 2. We noticed a positive relationship between this amplitude and the area of the landforms (Fig. 8). This means that it is easier for the model to provide a classification decision when the sample is larger. The largest amplitude occurs for the elevation, entropy and standard deviation, and this is consistent with the variable importance results from the XGBoost model. Moreover, it should be emphasized that the range of impact between sheets is different. This is because the sheets present areas of varying morphogenesis with different levels of geomorphological features (marking); therefore, some geomorphometric variables are more (or less) useful for characterizing the forms occurring there. In practice, this means that if the slope surface class can be easily classified on the Toruń sheet (i.e., a young glacial area) using the slope variable, it may be impossible on the Nowy Targ sheet (i.e., a young mountains area) due to the completely different structure and characteristics of the Tatra Mountains range.

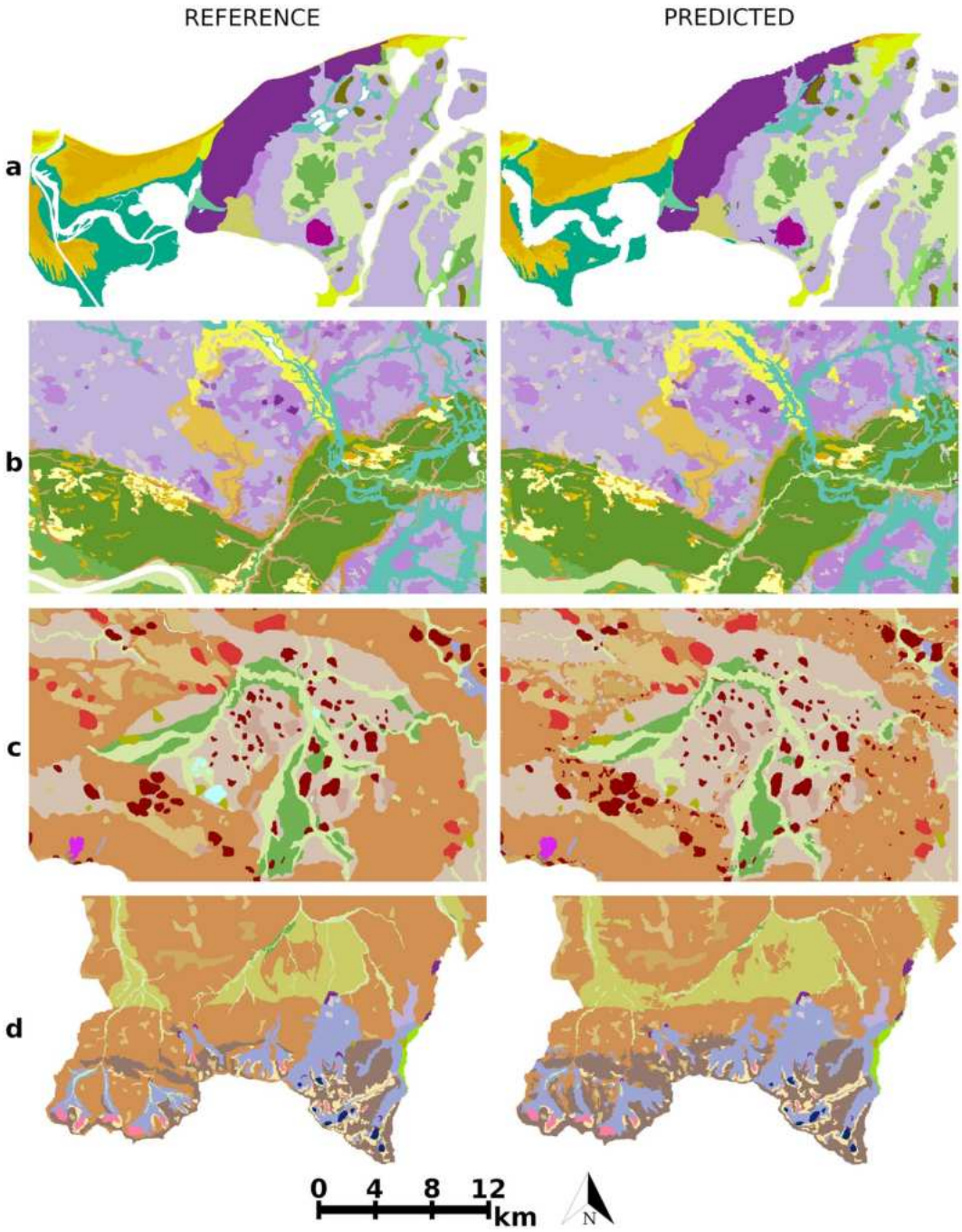

Figure 6. Comparison of the reference data (left) with predicted landforms (right) in the area of: (**a**) Wolin Island; (**b**) Chełmno-Dobrzyń Lakeland; (**c**) Jelenia Góra Basin; (**d**) Tatra Mountains. The predicted raster was smoothed with a modal filter of 5 pixels and landforms smaller than 21 pixels (~ 18,000 m^2) were removed using a sieve filter. The legend is available in Supplementary Fig. S1.

Discussion

As demonstrated, the accumulated local effect plot is a valuable tool for interpreting the classification decisions made by the machine learning model. Surprisingly, to the best of the author's knowledge, this technique has not yet been used in geomorphological mapping. Fundamentally, it allows us to check why the classifier has distinguished a given landform, which is especially important in the case of incorrect classifications (we can interpret on the basis of explanatory variables what causes the error). Moreover, this method can be used even in traditional mapping; if a geomorphologist is not sure about recognizing a certain landform in the field, he can assist with ALE plots. Its main advantage is the relative ease of interpretation since it provides a clear visualization of how each geomorphometric feature influences the model's predictions. In this study, it was used to interpret the classification decisions based on the gradient boosting model (XGBoost), but actually it can be applied to a

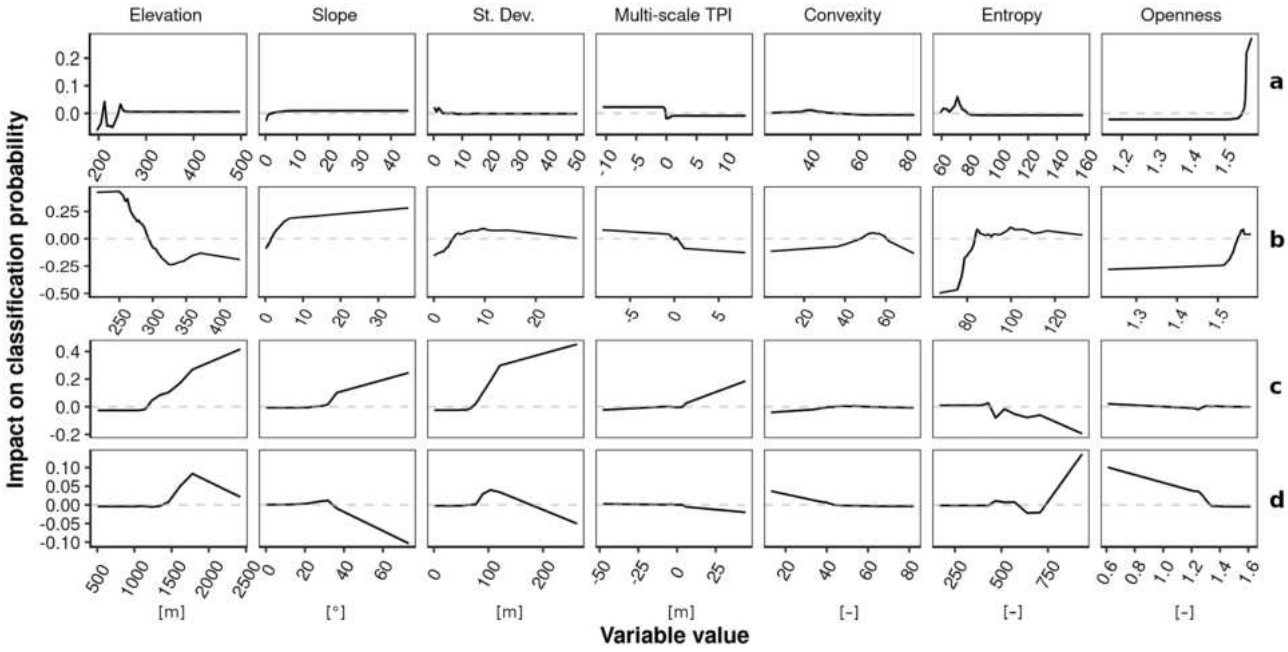

Figure 7. Accumulated local effects plot showing how geomorphometric variables affect the probability of classifying: (**a**) proluvial plain; (**b**) plateau; (**c**) rock wall/rock slope; (**d**) depositional scree slope. Entropy is expressed in thousands.

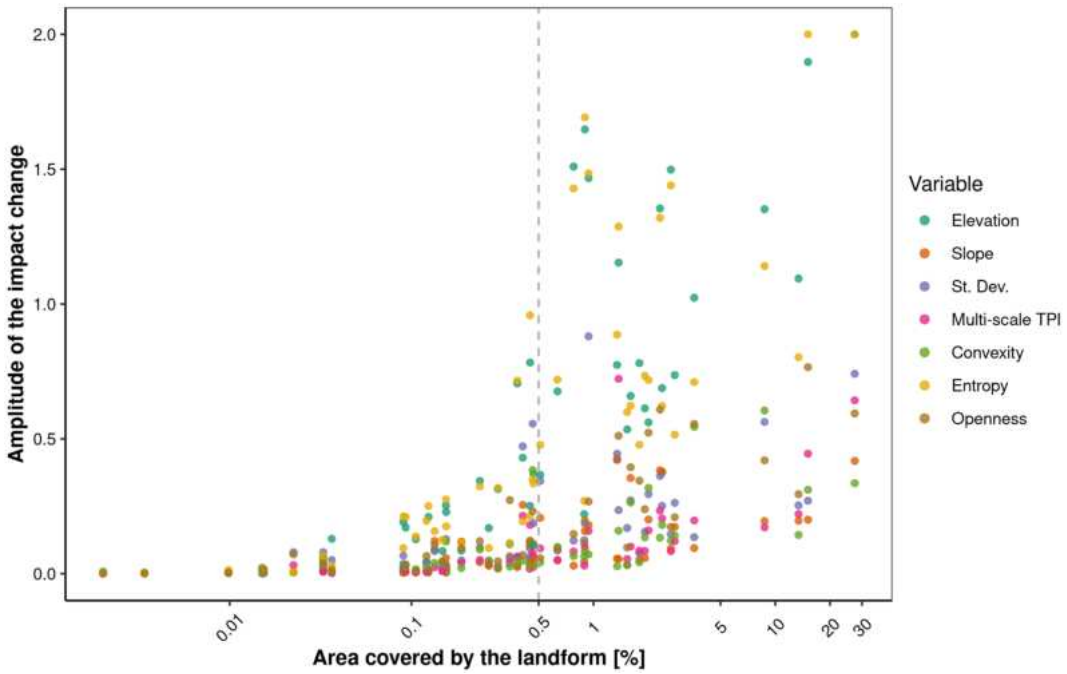

Figure 8. Amplitude of the impact change on the landform classification depending on its area and geomorphometric variables. Each set of points represents a different landform. 0.5% is the threshold value (marked with a dashed gray line) at which there is a significant increase in amplitude. The values on the X-axis are presented on a logarithmic scale. The total area is 9072 km^2.

wide range of different models (for example, linear models, tree-based models, neural networks). While ALE is certainly a useful tool, it also has some drawbacks. The main limitations are related to small datasets, low feature variability and the sensitivity of the model itself. The quality of the input dataset and the accuracy of the model should be carefully considered before conclusions are drawn.

Comparing the obtained results for convolutional neural networks, we can see differences in the accuracy of the classifiers between those provided by Du et al.[13]: 83–98%, Li et al.[14]: 78–87%, Meij et al.[15]: 44–94%, Xu

et al.[16]: 84% and 70%, and in this study (59.8%). These differences are due to two reasons. The first is that in the mentioned studies, classifications were carried out only for a few geomorphological units, while in this study, 54 different landforms were classified. Naturally, this means obtaining such high-performance values is more complicated. The second reason is the selection of the research area. The areas chosen by the cited authors are very diverse and relatively easy to distinguish, which does not entirely illustrate the scale of the problem. The largest challenge is the analysis of the areas of complex genesis with poorly marked geomorphological features. In this case, the lower efficiency of the classifier is expected for the Polish sheets used, which consist mainly of lowlands shaped by the glacial and denudative processes.

In this study, U-Net was used as a well-established reference model representing the convolutional neural network approach. Although it is widely used for image segmentation, it has some limitations related to the optimization of a huge number of parameters. Since its publication, an improved version has been proposed by Dinh et al.[57], namely U-Lite, which requires fewer parameters (but still hundreds of thousands). There are also alternative CNN architectures with relatively fewer parameters, for example LeNet-5, requiring 60,000 parameters[58] or its improved version (3DLeNet) recently proposed by Fırat et al.[59] for classifying hyperspectral images. However, simplifying the architecture and reducing the number of parameters can make the model unable to recognize complex spatial patterns and structures, and therefore its effectiveness will still be low. The better performance of tree-based algorithms compared to CNNs in the study can be explained by the fact that they can perform better when handling few data observations. In a digital soil mapping experiment utilizing Random Forests, Bouslihim et al.[60] showed that it could provide good performance by selecting only a few of the most relevant explanatory variables.

While this article focuses on the mapping of existing digital geomorphological maps using automatic classification on a regional scale, a further question arises: Do the methods and dataset used allow extrapolation of results for the entire country? In order to answer the question, we attempted to use the XGBoost classifier, which was trained on a large sample of over 3 million observations and has previously provided promising results (Table 2). Based on the experiment, we conclude that at this point the results are unsatisfactory and do not meet mapping standards. The main limitation in this case is the insufficient amount of reference materials, as they constitute approximately 3% of the country's coverage (over 9,000 km^2), thus causing the trained model to be unable to recognize the same landforms in areas with different geomorphological characteristics. This issue was discussed more extensively by Bouasria et al.[61] in the context of spatial extrapolation, where authors concluded that increasing the size of the spatial extent of the survey reduces the accuracy of the model. Therefore, we recommend further work to increase coverage by digital geomorphological maps at a scale of 1:100,000.

Conclusions

In this article, we evaluated the potential of applying machine learning models and convolutional neural network to automatic geomorphological mapping and examined the usefulness and impact of selected geomorphometric variables on the results of landform classifications. Based on the results of this study, we can conclude that supervised learning methods are effective for mapping known sheets (Fig. 6), but ineffective for extrapolation to new areas, especially when the catalog of landforms is very extensive. Therefore, at this point, we can state that automatic methods cannot replace the traditional approach, but they can support mapping if the geomorphological characteristics in the predicted area are similar to those in the training dataset.

We used diagnostic techniques based on the analysis of the importance of geomorphometric variables to indicate the most useful variables for geomorphological mapping, and accumulated local effects plots to precisely examine how their values influence the model's classification decisions. This made the obscure and complicated classification mechanisms of the black box model more explicit and open to human interpretation.

The topic of automatic mapping remains unsolved, and further research is required. Further work should primarily focus on developing better geomorphometric variables for machine learning models and improving the architecture of the convolutional neural network to detect rarer landforms. In addition, future work should also address the issue of spatial validation at the model training and testing stages. Regarding the issue of explaining the decisions made by classification models, it would be useful to check the differences and similarities in the method inspired by game theory, i.e., *shapley additive explanation* proposed by Lundberg et al.[62].

Data availability

The programming scripts used for this analysis and to generate the figures are available in the following GitHub repository: https://github.com/kadyb/geomorph_classification. The reference geomorphological maps are available from the Head Office of Geodesy and Cartography in Poland, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Received: 25 November 2023; Accepted: 1 March 2024
Published online: 05 March 2024

References

1. Oldroyd, D. R. & Grapes, R. H. Contributions to the history of geomorphology and quaternary geology: An introduction. *Geol. Soc.* **301**(1), 1–17. https://doi.org/10.1144/SP301.1 (2008).
2. Knight, J., Mitchell, W. A. & Rose, J. Geomorphological field mapping. In *Developments in Earth Surface Processes*, Vol. 15, 151–187 (Elsevier, 2011). https://doi.org/10.1016/B978-0-444-53446-0.00006-9.
3. Bazaglia Filho, O. *et al.* Comparison between detailed digital and conventional soil maps of an area with complex geology. *Rev. Bras. Ciênc. Solo* **37**(5), 1136–1148. https://doi.org/10.1590/S0100-06832013000500003 (2013).
4. Dikau, R., Brabb, E. E. & Mark, R. M. Landform classification of New Mexico by computer. Report 91–634 (1991). https://doi.org/10.3133/ofr91634.

5. Irvin, B. J., Ventura, S. J. & Slater, B. K. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. *Geoderma* **77**(2–4), 137–154. https://doi.org/10.1016/S0016-7061(97)00019-0 (1997).
6. Iwahashi, J. & Pike, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* **86**(3–4), 409–440. https://doi.org/10.1016/j.geomorph.2006.09.012 (2007).
7. Jasiewicz, J. & Stepinski, T. F. 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 (2013).
8. MacMillan, R. A., Pettapiece, W. W., Nolan, S. C. & Goddard, T. W. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets Syst.* **113**(1), 81–109. https://doi.org/10.1016/S0165-0114(99)00014-7 (2000).
9. Drăguţ, L. & Blaschke, T. Automated classification of landform elements using object-based image analysis. *Geomorphology* **81**(3–4), 330–344. https://doi.org/10.1016/j.geomorph.2006.04.013 (2006).
10. Drăguţ, L. & Eisank, C. Automated object-based classification of topography from SRTM data. *Geomorphology* **141–142**, 21–33. https://doi.org/10.1016/j.geomorph.2011.12.001 (2012).
11. Chollet, F. *Deep Learning with Python* (Manning Publications Co, 2017).
12. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436–444. https://doi.org/10.1038/nature14539 (2015).
13. Du, L. *et al.* 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 (2019).
14. Li, S., Xiong, L., Tang, G. & Strobl, J. 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 (2020).
15. Meij, W. M. *et al.* Comparing geomorphological maps made manually and by deep learning. *Earth Surf. Process. Landf.* **47**(4), 1089–1107. https://doi.org/10.1002/esp.5305 (2022).
16. Xu, Y., Zhu, H., Hu, C., Liu, H. & Cheng, Y. Deep learning of DEM image texture for landform classification in the Shandong area, China. *Front. Earth Sci.* **16**(2), 352–367. https://doi.org/10.1007/s11707-021-0884-y (2022).
17. Buraczyński, J. Development of the relief of Roztocze Upland (with electronic geomorphological map 1:50 000, elaborated by J. Buraczyński and Ł.Chabudziński). *Landf. Anal.* **27**, 67–89. https://doi.org/10.12657/landfana.027.006 (2014).
18. Marcinkowska, A., Ochtyra, A., Olędzki, J. R., Wołk-Musiał, E. & Zagajewski, B. Mapa geomorfologiczna województw pomorskiego i warmińsko-mazurskiego z wykorzystaniem metod geoinformatycznych. *Teledetekcja Śr.* **49**, 43–79 (2013).
19. Napiórkowska, M. Cyfrowa mapa geomorfologiczna Mazowsza. *Teledetekcja śr.* **45**, 23–39 (2011).
20. Borzuchowski, J. & Olędzki, J. R. Cyfrowa mapa geomorfologiczna Karpat. *Teledetekcja Śr.* **46**, 52–71 (2011).
21. Wołk-Musiał, E. & Gatkowska, M. Mapa geomorfologiczna narwiańskiego Parku Narodowego wraz z otuliną w skali 1: 25000. *Teledetekcja śr.* **45**, 40–50 (2011).
22. Dmowska, A., Gudowicz, J. & Zwoliński, Z. Cyfrowa adaptacja analogowych map geomorfologicznych. *Landf. Anal.* **12**, 35–47 (2010).
23. Krzemiński, J. Przeglądowa Mapa Geomorfologiczna Podlasia, w skali 1: 300 000. *Teledetekcja śr.* **42**, 43–58 (2009).
24. Wołk-Musiał, W., Lewiński, S. & Zagajewski, B. Geomorphological map (Tykocin sheet). Methodology. In *Observing Our Environment from Space* 385–390 (CRC Press, 2002).
25. Wieczorek, M. & Migoń, P. 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 (2014).
26. Szypuła, B. & Wieczorek, M. Geomorphometric relief classification with the k-median method in the Silesian Upland, southern Poland. *Front. Earth Sci.* **14**(1), 152–170. https://doi.org/10.1007/s11707-019-0765-9 (2020).
27. Janowski, L., Tylmann, K., Trzcinska, K., Rudowski, S. & Tegowski, J. Exploration of glacial landforms by object-based image analysis and spectral parameters of digital elevation model. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–17. https://doi.org/10.1109/TGRS.2021.3091771 (2022).
28. Dyba, K. & Jasiewicz, J. Toward geomorphometry of plains: Country-level unsupervised classification of low-relief areas (Poland). *Geomorphology* **413**, 108373. https://doi.org/10.1016/j.geomorph.2022.108373 (2022).
29. Rączkowska, Z. & Zwoliński, Z. Digital geomorphological map of Poland. *Geogr. Pol.* **88**(2), 205–210. https://doi.org/10.7163/GPol.0025 (2015).
30. Gustavsson, M., Kolstrup, E. & Seijmonsbergen, A. C. A new symbol-and-GIS based detailed geomorphological mapping system: Renewal of a scientific discipline for understanding landscape development. *Geomorphology* **77**(1–2), 90–111. https://doi.org/10.1016/j.geomorph.2006.01.026 (2006).
31. Yap, B. W., Rani, K. A., Rahman, K. A., Fong, S., Khairudin, Z. & Abdullah, N. N. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), in Lecture Notes in Electrical Engineering*, Vol. 285 (eds Herawan, T., Deris, M. M. & Abawajy, J.) 13–22 (Springer Singapore, 2014). https://doi.org/10.1007/978-981-4585-18-7_2.
32. Jasiewicz, J., Netzel, P. & Stepinski, T. F. Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology* **221**, 104–112. https://doi.org/10.1016/j.geomorph.2014.06.011 (2014).
33. Mark, R. K. Multidirectional, oblique-weighted, shaded-relief image of the Island of Hawaii. Report 92–422 (1992). https://doi.org/10.3133/ofr92422.
34. Zevenbergen, L. W. & Thorne, C. R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* **12**(1), 47–56. https://doi.org/10.1002/esp.3290120107 (1987).
35. Guisan, A., Weiss, S. B. & Weiss, A. D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* **143**(1), 107–122. https://doi.org/10.1023/A:1009841519580 (1999).
36. Köthe, R. & Lehmeier, F. *SARA—system zur automatischen relief-analyse* (User Manual, 1996).
37. Yokoyama, R., Shirasawa, M. & Pike, R. J. Visualizing topography by openness: A new application of image processing to digital elevation models. *Photogramm. Eng. Remote Sens.* **68**(3), 257–266 (2002).
38. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314 (1973).
39. Breiman, L. Random forests. *Mach. Learn.* **45**(1), 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
40. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016). https://doi.org/10.1145/2939672.2939785.
41. Ke, G. *et al.* LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17 3149–3157 (Curran Associates Inc., 2017).
42. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention: MICCAI 2015*, *in Lecture Notes in Computer Science* Vol. 9351 (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
43. Abadi, M. *et al.* TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, in OSDI'16, 265–283 (USENIX Association, 2016).
44. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv, 29, 2017. Accessed Mar 27, 2023. [Online]. http://arxiv.org/abs/1412.6980.

45. Ploton, P. *et al.* Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**(1), 4540. https://doi.org/10.1038/s41467-020-18321-y (2020).
46. Meyer, H. & Pebesma, E. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nat. Commun.* **13**(1), 2208. https://doi.org/10.1038/s41467-022-29838-9 (2022).
47. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* 2nd edn. (Christoph Molnar, 2022).
48. Biecek, P. & Burzykowski, T. *Explanatory model analysis: Explore, explain, and examine predictive models* 1st edn. In *Chapman & Hall/CRC Data Science Series* (CRC Press, 2021).
49. Apley, D. W. & Zhu, J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**(4), 1059–1086. https://doi.org/10.1111/rssb.12377 (2020).
50. Greenwell, B. M., Boehmke, B. C. & McCarthy, A. J. A simple and effective model-based variable importance measure. arXiv, May 12, 2018. Accessed Nov 05, 2023. [Online]. http://arxiv.org/abs/1805.04755.
51. Neteler, M., Bowman, M. H., Landa, M. & Metz, M. GRASS GIS: A multi-purpose open source GIS. *Environ. Model. Softw.* **31**, 124–130. https://doi.org/10.1016/j.envsoft.2011.11.014 (2012).
52. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022). [Online]. https://www.R-project.org/.
53. Pebesma, E. J. & Bivand, R. *Spatial data science: With applications in R* 1st edn. In *Chapman & Hall/CRC Press the R Series* (CRC Press, 2023).
54. Kuhn, M., Vaughan, D. & Hvitfeldt, E. yardstick: Tidy characterizations of model performance (2022). [Online]. https://CRAN.R-project.org/package=yardstick.
55. Wright, M. N. & Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Soft.* https://doi.org/10.18637/jss.v077.i01 (2017).
56. Solon, J. *et al.* Physico-geographical mesoregions of Poland: Verification and adjustment of boundaries on the basis of contemporary spatial data. *Geogr. Pol.* **91**(2), 143–170. https://doi.org/10.7163/GPol.0115 (2018).
57. Dinh, B.-D., Nguyen, T.-T., Tran, T.-T. & Pham, V.-T. 1M parameters are enough? A lightweight CNN-based model for medical image segmentation. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* , 1279–1284 (IEEE, 2023). https://doi.org/10.1109/APSIPAASC58517.2023.10317244.
58. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324. https://doi.org/10.1109/5.726791 (1998).
59. Fırat, H., Asker, M. E., Bayindir, M. İ & Hanbay, D. Spatial-spectral classification of hyperspectral remote sensing images using 3D CNN based LeNet-5 architecture. *Infrared Phys. Technol.* **127**, 104470. https://doi.org/10.1016/j.infrared.2022.104470 (2022).
60. Bouslihim, Y. *et al.* The effect of covariates on Soil Organic Matter and pH variability: A digital soil mapping approach using random forest model. *Ann. GIS* https://doi.org/10.1080/19475683.2024.2309868 (2024).
61. Bouasria, A., Bouslihim, Y., Gupta, S., Taghizadeh-Mehrjardi, R. & Hengl, T. Predictive performance of machine learning model with varying sampling designs, sample sizes, and spatial extents. *Ecol. Inform.* **78**, 102294. https://doi.org/10.1016/j.ecoinf.2023.102294 (2023).
62. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. arXiv, Mar 06, 2019. Accessed Sep 21, 2023. [Online]. http://arxiv.org/abs/1802.03888.

Acknowledgements

This research was funded in part by the 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 digital geomorphological maps were used in this study under license number *DIO.7211.342.2021_PL_N* issued by the Head Office of Geodesy and Cartography in Poland. The author especially thanks his thesis and grant supervisor Professor Jarosław Jasiewicz. The author also thanks Professors Zbigniew Zwoliński and Małgorzata Mazurek for consulting on the project's results in the context of geomorphology mapping.

Author contributions

K.D.: conceptualization, methodology, investigation, software, validation, formal analysis, data curation, visualization, writing-original draft, writing—review and editing, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-56066-6.

Correspondence and requests for materials should be addressed to K.D.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2024

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
0169-555X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

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:

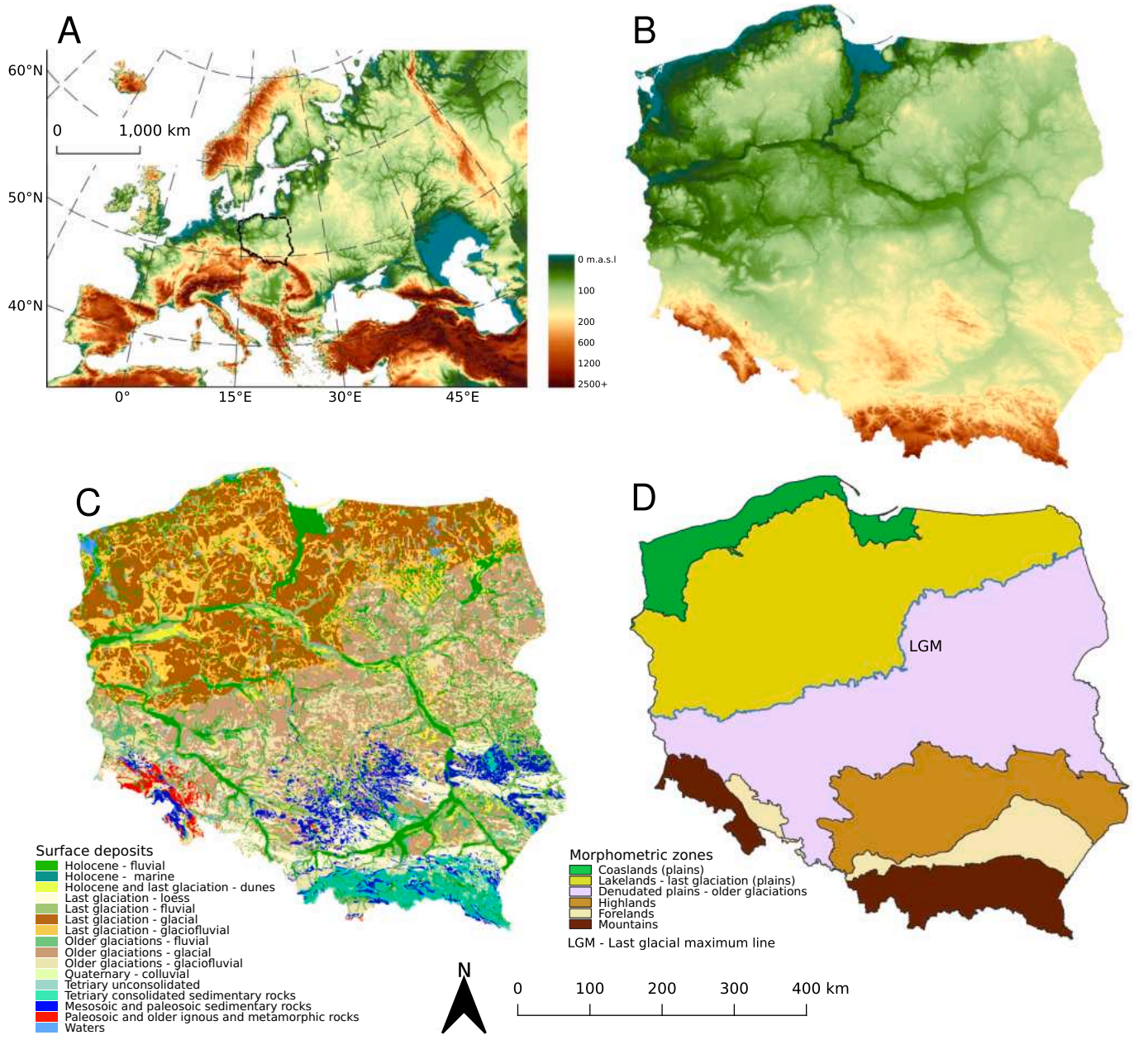

Fig. 1. A) Location of Polish borders superimposed on the Europe relief; B) Relief of Poland; C) Geology of Poland - origin of surface sediments, based on Geological Map of Poland 1:500,000 (Marks et al., 2006), simplified; D) Main morphogenetic zones (after Gilewska (1991)), boundaries adopted to physio-geographic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

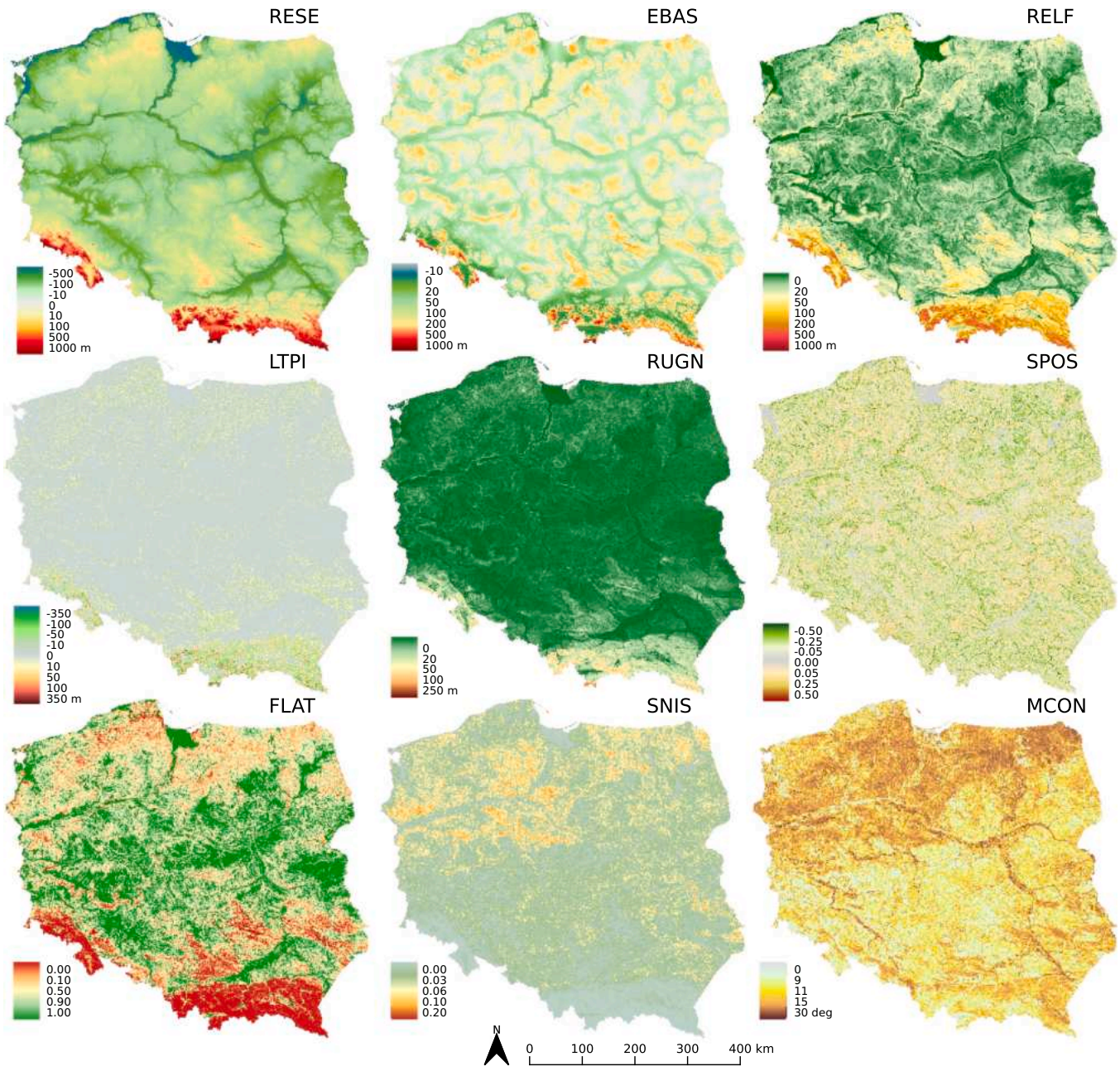

Fig. 2. Nine morphometric variables (ELEV is presented in Fig. 1), used in the analysis.

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

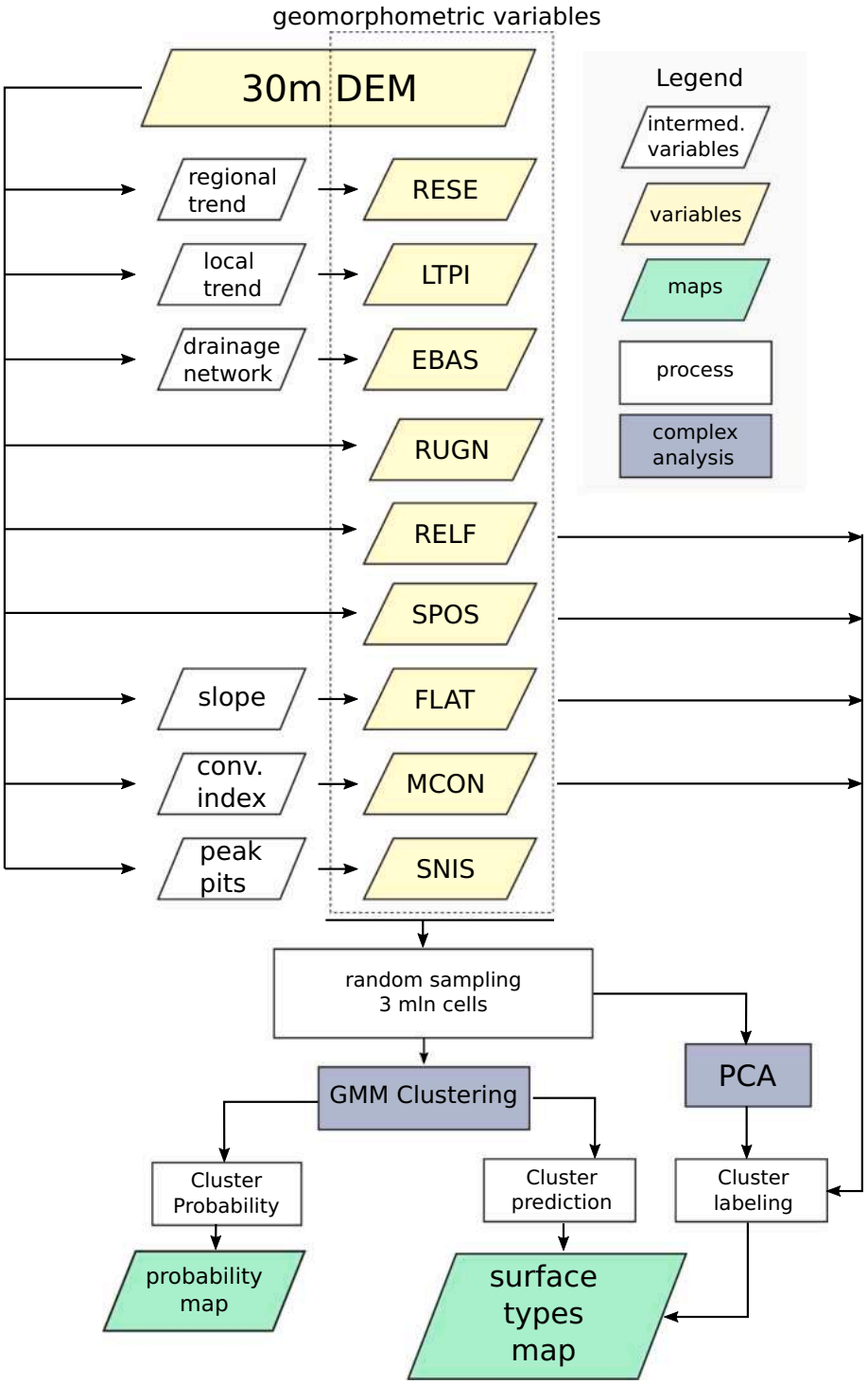

Fig. 3. Graphical summary of the data processing. For variable abbreviation, see Table 1.

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

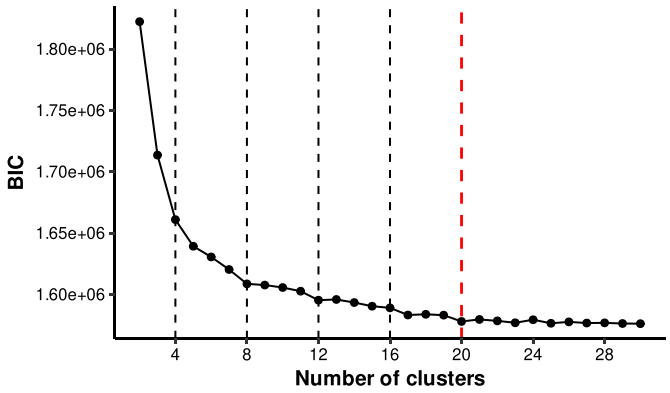

Fig. 4. The changes of Bayesian Information Criterion between 2 and 30 clusters. The red dashed line shows the final number of clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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:

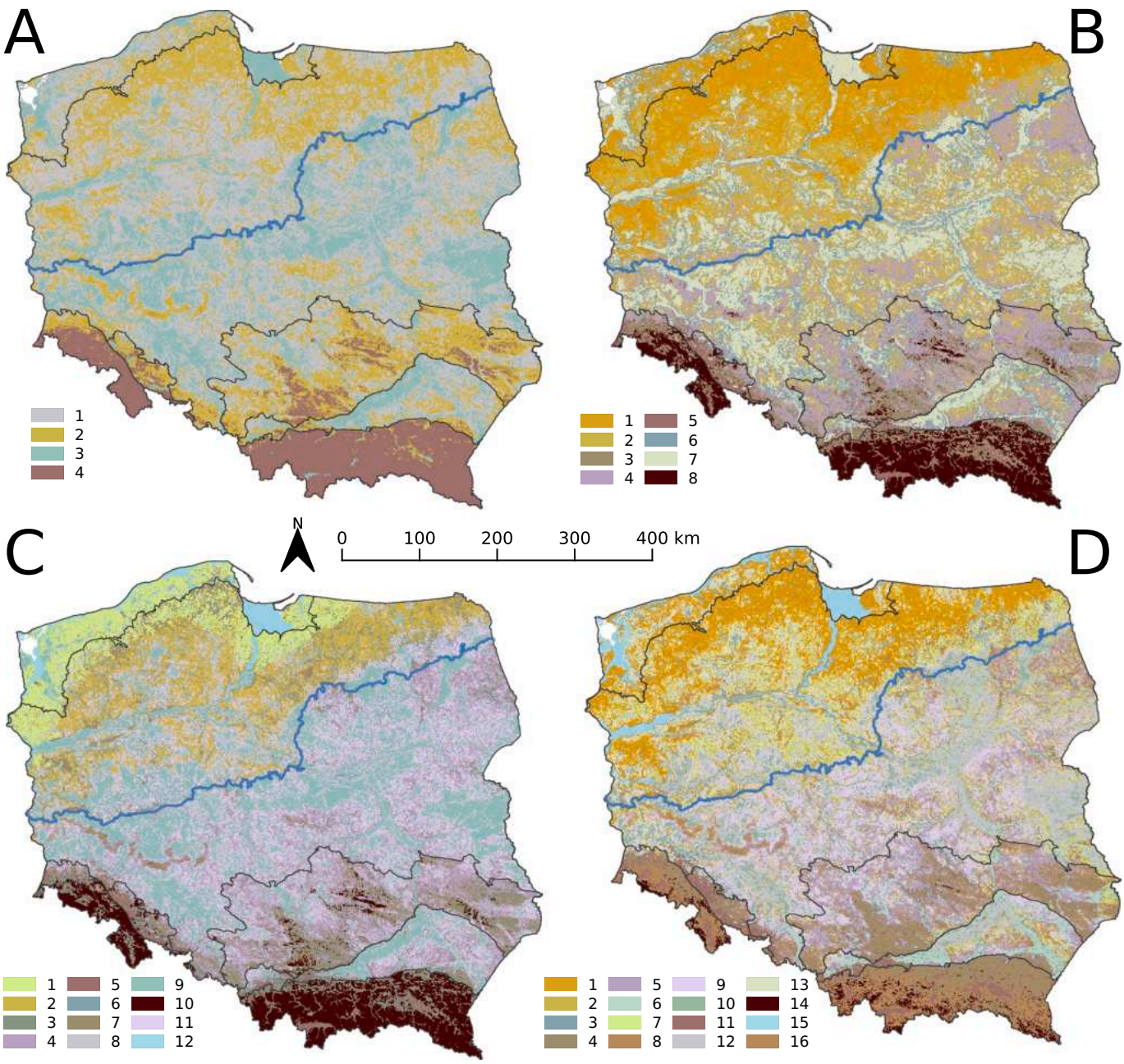

Fig. 5. Result of clustering with: A) four; B) eight; C) twelve; D) sixteen number of classes. Colors are meaningless, however partially relates to the symbology presented in Fig. 6. Blue line marks extend of Last Glacial Maximum (LGM). (For interpretation of the references to color in this figure legend,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

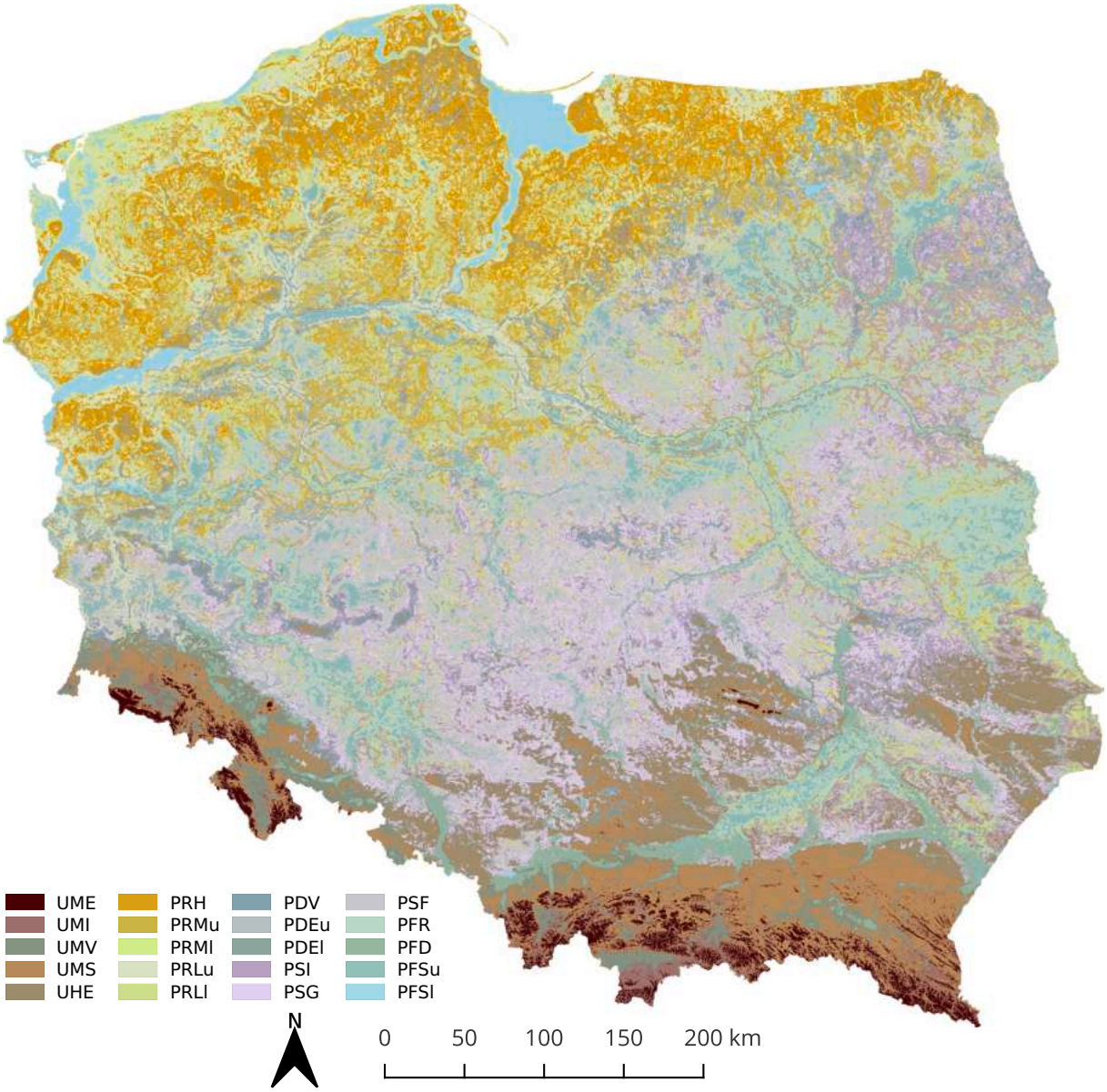

Fig. 6. Result of clustering with 20 clusters. See the Fig. 7 for detail classification and Table 2 for full land surface types names.

"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

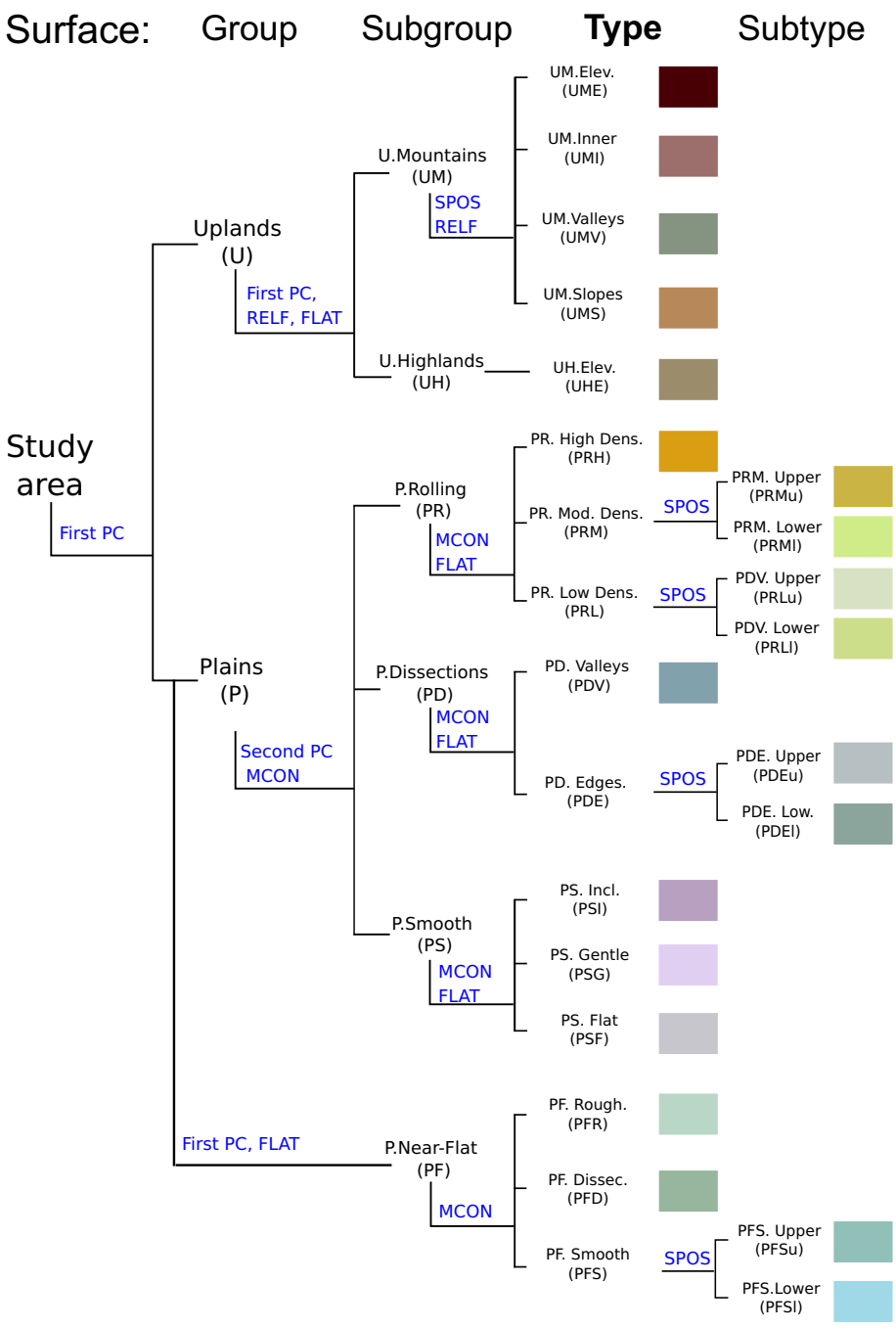

Fig. 7. Classification process (labeling) into land surface types Detail description of the process contains Subsection 3.2. See the *SEC:acronyms* section for full names.

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

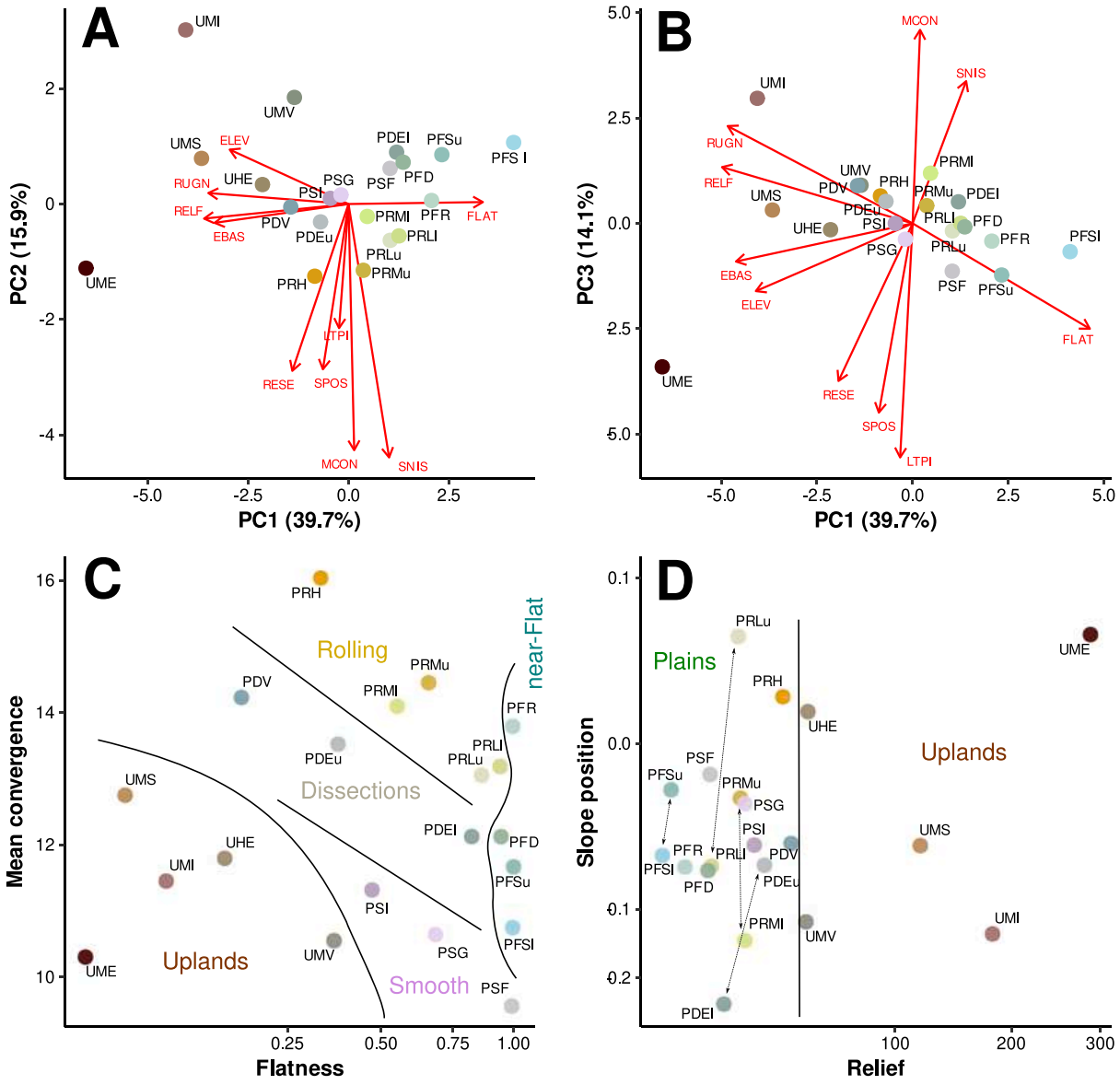

Fig. 8. Cluster distribution: A–B) against principal components (red arrows indicate variable's loadings); C–D) relief variables representing vertical variability (RELF and SPOS) and textural variables representing horizontal fluctuations (FLAT and MCON). Arrows on chart D show the position relation 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

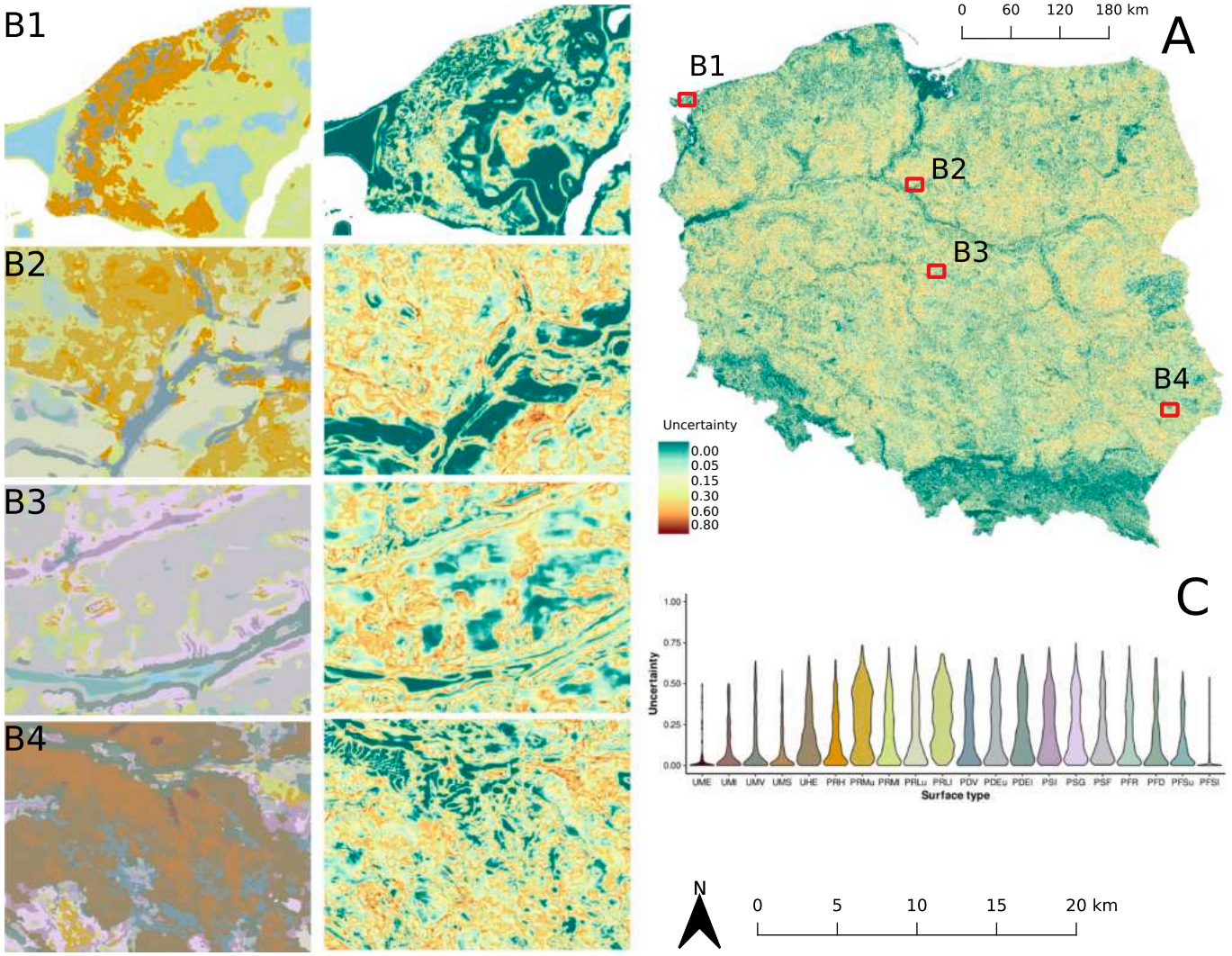

Fig. 9. The distribution of uncertainty: A) over entire study area; B) detailed, inside selected areas (location are marked on plate A); C) inside land surface types (violin plot).

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

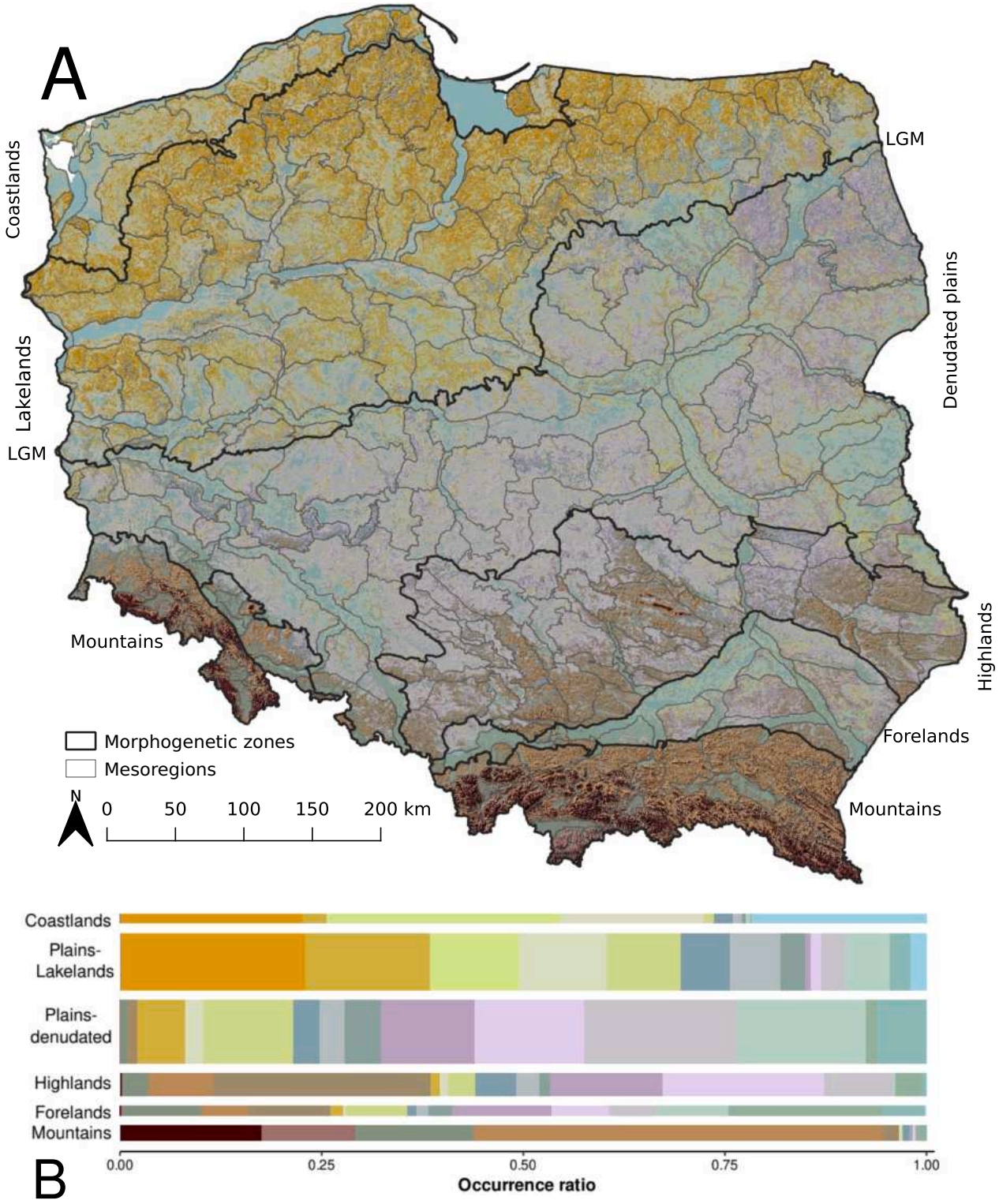

Fig. 10. Results of classification, compared with the extent of morphogenetic zones (see Fig. 1). A) Cluster distribution on the background of morphogenetic zones and physico-geographical mesoregions of Poland (Solon et al., 2018); B) Correspondence between morphogenetic zones and designated land surface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

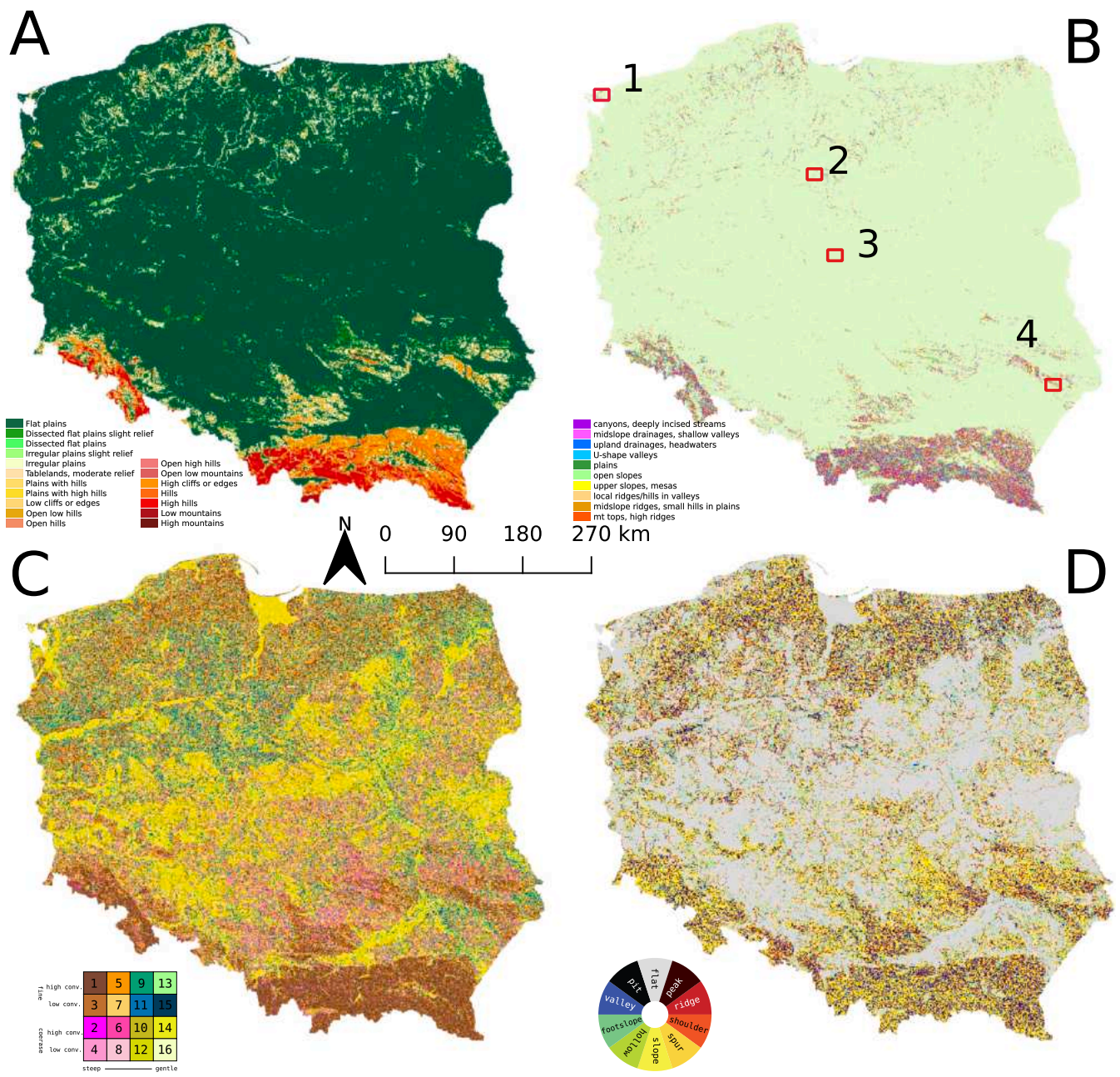

Fig. 11. Morphometric mosaics obtained with different methods. A) Hammond (1964) system; B) Topographic Position Index (Weiss, 2001); C) Iwahashi and Pike (2007); D) Geomorphons (Jasiewicz and Stepinski, 2013). Rectangles on Plate B marks location of testing areas presented on Fig. 12.

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).

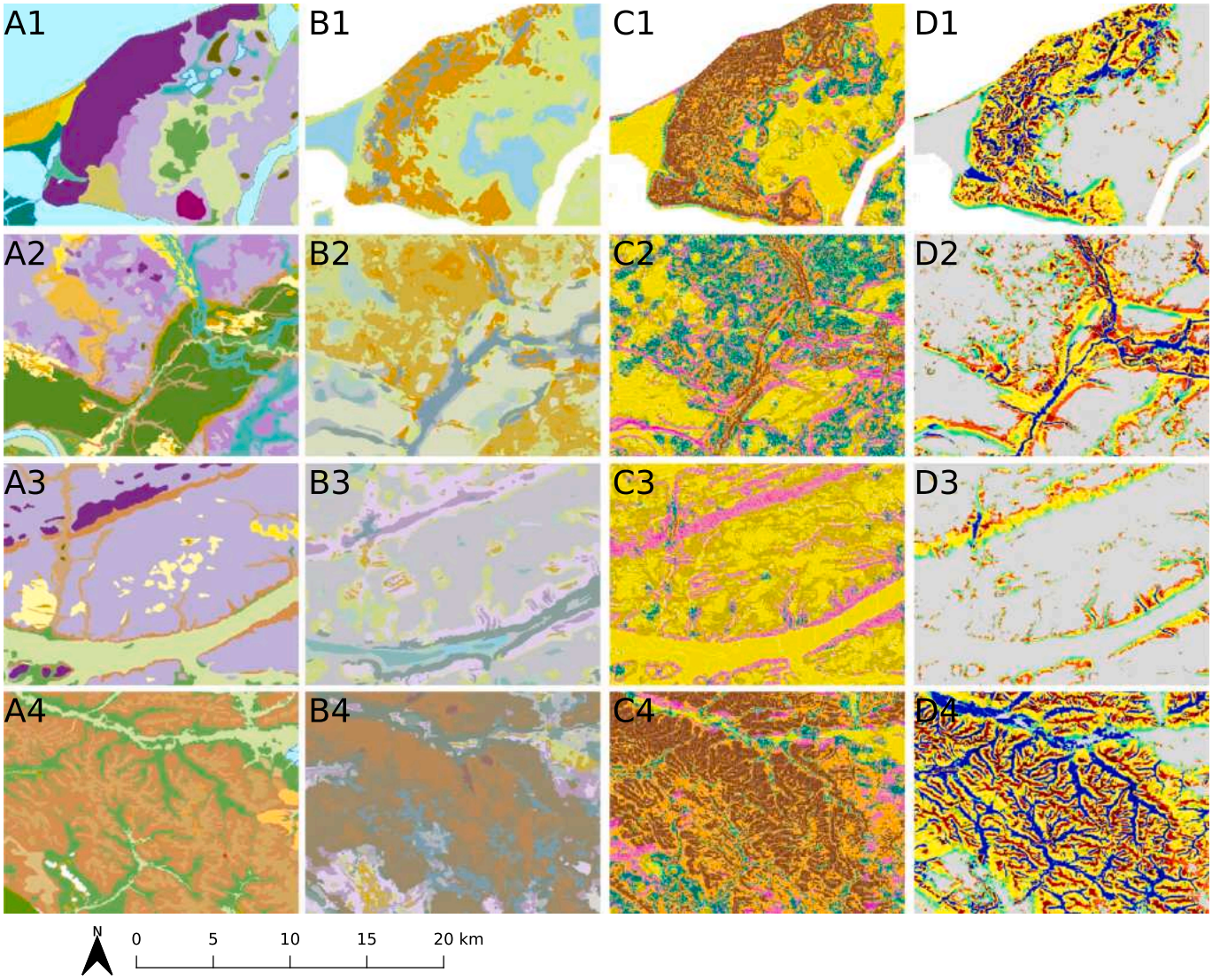

Fig. 12. Detailed comparison of expert-driven maps (Column A) with geomorphometric mosaics. Column B – our classification; Column C – Iwahashi and Pike (2007); column D – Geomorphons. Location of testing areas 1–4 is presented on Fig. 11, plate B. For expert-driven maps legend see Appendix A.

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 **Noise**
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.