ACTA GEOGRAPHICA SLOVENICA

GEOGRAFSKI ZBORNIK

2019
59
2

ACTA GEOGRAPHICA SLOVENICA
GEOGRAFSKI ZBORNIK
59-2 • 2019

Contents

Special issue – Franciscean cadaster as a source of studying landscape changes

ISSN 1581-6613
9 771581 661010

ACTA GEOGRAPHICA SLOVENICA

59-2
2019

ISSN: 1581-6613
COBISS: 124775936
UDC/UDK: 91

Issued by/izdajatelj: Geografski inštitut Antona Melika ZRC SAZU
Published by/založnik: Založba ZRC
Co-published by/sozaložnik: Slovenska akademija znanosti in umetnosti

Address/Naslov: Geografski inštitut Antona Melika ZRC SAZU, Gosposka ulica 13, SI – 1000 Ljubljana, Slovenija

The papers are available on-line/prispevki so dostopni na medmrežju: http://ags.zrc-sazu.si (ISSN: 1581–8314)

Ordering/naročanje: Založba ZRC, Novi trg 2, p. p. 306, SI – 1001 Ljubljana, Slovenija; zalozba@zrc-sazu.si

Annual subscription/letna naročnina: 20 € for individuals/za posameznike, 28 € for institutions/za ustanove.
Single issue/cena posamezne številke: 12,50 € for individuals/za posameznike, 16 € for institutions/za ustanove.

Cartography/kartografija: Geografski inštitut Antona Melika ZRC SAZU
Translations/prevodi: DEKS, d. o. o.
DTP/prelom: SYNCOMP, d. o. o.
Printed by/tiskarna: Tiskarna Present, d. o. o.
Print run/naklada: 450 copies/izvodov

The journal is subsidized by the Slovenian Research Agency and is issued in the framework of the Geography of Slovenia core research programme (P6-0101)/revija izhaja s podporo Javne agencije za raziskovalno dejavnost Republike Slovenije in nastaja v okviru raziskovalnega programa Geografija Slovenije (P6-0101).

The journal is indexed also in/revija je vključena tudi v: SCIE – Science Citation Index Expanded, Scopus, JCR – Journal Citation Report/Science Edition, ERIH PLUS, GEOBASE Journals, Current geographical publications, EBSCOhost, Geoscience e-Journals, Georef, FRANCIS, SJR (SCImago Journal & Country Rank), OCLC WorldCat, Google scholar, and CrossRef.

Oblikovanje/Design by: Matjaž Vipotnik

Front cover photography: Exploration of the collapse dolines, such as the one at the Small Natural Bridge in Rakov Škocjan, has enabled a deeper understanding of karst processes in recent years (photograph: Matej Lipar).
Fotografija na naslovnici: Raziskave udornice, kot je ta pri Malem Naravnem mostu v Rakovem Škocjanu, so v zadnjih letih omogočile globlje razumevanje kraških procesov (fotografija: Matej Lipar).

Acta geographica Slovenica, 59-2, 2019, 7–26

THE USEFULNESS OF UNSUPERVISED CLASSIFICATION METHODS FOR LANDSCAPE TYPIFICATION: THE CASE OF SLOVENIA

Drago Perko, Rok Ciglič, Mauro Hrvatin

CORTYN, SHUTTERSTOCK.COM

The landscape classification of countries with a high landscape diversity, such as Slovenia, using a computer and various classification methods is a difficult task. The picture shows three landscape types: fertile flysch Mediterranean hills with the Bay of Koper (foreground), Mediterranean karst plateaus (middle), and high Dinaric karst plateaus (background).

DOI: https://doi.org/10.3986/AGS.7377
UDC: 911.5(497.4)
COBISS: 1.01

The usefulness of unsupervised classification methods for landscape typification: The case of Slovenia

ABSTRACT: Supervised and unsupervised classification methods can be a useful tool in determining various geographical spatial divisions, especially regionalizations and typifications. Because Slovenia is geographically very diverse, its divisions are a particularly significant and interesting research challenge. The main objective of this article is to determine the effectiveness of unsupervised classification methods, and therefore we compare the well-established landscape typology of Slovenia from 1996 with landscape typologies that were modeled using various unsupervised classification methods. Our results show that landscape typologies modeled using unsupervised classification methods deviate more from the original landscape typology of Slovenia than landscape typologies modeled using random and expert-supervised classification methods.

KEY WORDS: geography, geographic information system, modeling, classification, landscape typology, Slovenia

Uporabnost metod nenadzorovane klasifikacije za pokrajinsko tipizacijo na primeru Slovenije

POVZETEK: Metode nadzorovane in nenadzorovane klasifikacije so lahko koristno orodje pri določanju različnih geografskih prostorskih delitev, še posebej pri regionalizacijah in tipizacijah. Ker je Slovenija geografsko zelo raznolika, so njene delitve še posebej velik in zanimiv raziskovalni izziv. Glavni namen članka je ugotoviti učinkovitost metod nenadzorovane klasifikacije, zato primerjamo uveljavljeno pokrajinsko tipizacijo Slovenije iz leta 1996 s pokrajinskimi tipizacijami, ki smo jih modelirali z različnimi metodami nenadzorovane klasifikacije. Naši rezultati kažejo, da se pokrajinske tipizacije, modelirane z metodami nenadzorovane klasifikacije, izvirni pokrajinski tipizaciji Slovenije ne približajo tako dobro, kot pokrajinske tipizacije, modelirane z metodami naključne in ekspertne nadzorovane klasifikacije.

KLJUČNE BESEDE: geografija, geografski informacijski sistem, modeliranje, klasifikacija, pokrajinska tipizacija, Slovenija

Drago Perko, Rok Ciglič, Mauro Hrvatin
Research Center of the Slovenian Academy of Sciences and Arts, Anton Melik Geographical Institute
drago@zrc-sazu.si, rok.ciglic@zrc-sazu.si, mauro@zrc-sazu.si

This article was submitted for publication on February 26th, 2018.
Uredništvo je prejelo prispevek 26. februarja 2018

1 Introduction

Regionalizations and typifications are among the most complicated fields of research in geography (Hammond 1964; Dikau, Brabb and Mark 1991; Kladnik 1996; Brabyn 1998; McMahon, Wiken and Gauthier 2004; Gallant, Douglas and Hoffer 2005; Iwahashi and Pike 2006; Ellison 2010; Ciglič 2014). This is especially the case with the spatial divisions of Slovenia, which despite being a small country has some of the most diverse landscapes in the world (Melik 1935; Ilešič 1956; 1958; Gams 1983; Natek 1993; Gams, Kladnik and Orožen Adamič 1995; Gabrovec and Hrvatin 1998; Gams 1998; Perko 1998; Plut 1999; Špes et al. 2002; Kladnik, Perko and Urbanc 2009; Perko and Hrvatin 2009; Hrvatin and Perko 2012; Ciglič and Perko 2012; 2013). This is most likely why Slovenian geographers have only produced four true landscape typologies to date. The first one, which included eighteen landscape types, was produced by Anton Melik in 1946 (Melik 1946), the second was produced by Drago Perko in 1996 and included nine landscape types (Kladnik 1996; Perko 1998; Perko 2001; Perko 2007), the third one, produced in 2002 by Metka Špes et al., contained thirteen types (Špes et al. 2002), and the fourth was produced in 2014 by Drago Perko et al. and included twenty-four landscape types (Perko, Hrvatin and Ciglič 2015).

Through increasingly more precise and accessible digital spatial data, technological development has also introduced changes to geographical classifications, including regionalization and typification, with various models and geographic information systems becoming widely used and influential research methods (Demeritt and Wainwright 2005).

The term **classification** has several definitions (McGarigal, Cushman and Stafford 2000). Thus, for example, it can refer to any formal arrangement of data into a hierarchy of categories or distribution into classes (Whittow 2000), or systematic assignment to classes or groups based on shared characteristics (Clark 1998). Classifications can be roughly divided into supervised and unsupervised classifications. To date, Slovenian geographers have primarily modeled supervised classifications and compared them against the 1996 landscape typology of Slovenia, which comprises nine types combined into four groups and is most widely used and also included in Slovenian legislation. Landscape typology models were produced using various supervised classification methods based on the training samples of already classified or defined cells of individual landscape types. It was determined that the method selection had a strong impact on the results because modeled typologies matched the original ones by 51 to 75% (Ciglič 2014; Ciglič and Perko 2015). During the last testing of the original 1996 typology, its accuracy was estimated at 94% (Ciglič et al. 2017). The capacity and impact of methods were also determined using distorted classifications (Ciglič 2018), which allowed for additional evaluations of the original classification.

However, for the first time in Slovenian research history, this article deals with models of Slovenia's landscape typologies based on unsupervised classification methods and contain the same number of types (groups) as the original 1996 landscape typology. Unsupervised models are compared against the original typology and supervised models, which contributes to evaluating supervised and unsupervised classification methods as well as assessing the suitability of the original typology's concept and design. With regard to unsupervised classification, it is important to know that the characteristics of individual types (i.e., groups) are not known in advance, but that unsupervised classification methods define or identify them by themselves.

2 The importance of landscape classification

There is still no agreement on whether landscape classification entails searching for actual or merely abstract units (Gams 1984; Udo de Haes and Klijn 1994; Bailey 1996), but, in any case, a classification is a minor or major abstraction of differences between features (Natek and Žiberna 2004). It is natural for people to seek order and organization among phenomena (Haggett 2001), which also applies to spatial or landscape phenomena. Because landscapes constantly change, constant verification of landscape types is essential (Mücher et al. 2003), facilitating more economical use of natural resources and their replenishment. Landscape changes must be accompanied by ongoing adaptation of society's organization and operation to the environment (Plut 2005), which makes it possible to implement principles of sustainability in the economy, society, and environment. Landscape types are important because they relatively homogenously respond to human impact (Špes et al. 2002) and demand similar landscape planning.

Spatial classification following the natural characteristics of a landscape forms the basis for optimal spatial organization. Environmental issues also relate more to natural borders than administrative ones (Bailey 1996; Olson et al. 2001), which is why spatial classifications based on natural factors are becoming increasingly common and are replacing political classifications (Bernert et al. 1997). For instance, NUTS3 regions, which in the Mediterranean often include rural and urban coastal areas (Hazeu et al. 2011), may vary greatly in terms of their natural and social characteristics.

Various disciplines use classifications adapted to the content of their research and their needs (e.g., climate, vegetation, and soil classifications), but often their work would be made easier through uniform landscape classifications (Brabyn 2018), especially at different spatial levels (McMahon, Wiken and Gauthier 2004). A landscape classification at the highest spatial level can be used to outline borders for general purposes and various disciplines, whereas spatial units or landscape types at a lower level can serve as a starting point for more specific purposes (Bailey 1996) or classification within individual disciplines. This requires understanding of the relations between a landscape as a whole and its components (e.g., land use and biodiversity), which is vital for managing the environment (landscape) and its resources (Jongman et al. 2006).

Landscape classification is also important for preserving the natural and cultural landscape; inventorying, evaluating, and monitoring the current situation; managing, planning, and conducting measurements; exploring scenarios; sampling; transferring models into the physical environment; presenting landscape diversity; analyzing environmental pressures; and so on (Runhaar and Udo de Haes 1994; Bailey 1996; Bunce et al. 1996; Bernert et al.1997; Bastian 2000; Mücher et al. 2003; Loveland and Merchant 2004; Romportl and Chuman 2012).

Based on all the above, it is not surprising that in some places classifications even have a formal character, being part of official documents or even legislation in specific countries. In 1996, the EU introduced the Pan-European Biological and Landscape Diversity Strategy (Pan-European … 1996), and in 2000 it adopted the European Landscape Convention (The European … 2018). For example, in Slovenia the landscape typology of the country (Perko 1998) is used in defining land quality assessment criteria (Pravilnik o določanju … 2008).

3 Classification methods

Classification entails combining similar units based on logical criteria (Dodge 2008). As part of geographic information systems, units are defined with p data layers or, in other words, they have p dimensions. Classifications can be made based on a single criterion or factor (a monothetic approach) or several factors (a polythetic approach). The former involves a one-dimensional and the latter a multidimensional data space (Loveland and Merchant 2004).

Many classification methods are used (McGarigal, Cushman and Stafford 2000; Rogerson 2006; Abonyi and Feil 2007; Dodge 2008; Warner and Campagna 2009) and, in terms of results, each of them more or less imposes a specific structure and leads to a specific solution. Therefore, it is best to compare the results of various methods (Ferligoj 1989; McGarigal, Cushman and Stafford 2000; Theodoridis and Koutroumbas 2006; Ciglič 2018).

Classification methods are divided into soft and hard or relative and absolute, but most often a distinction is made between supervised and unsupervised classifications (Warner and Campagna 2009). With supervised methods, known values of training cells are available for classification, whereas that is not the case with unsupervised ones (Theodoridis and Koutroumbas 2006); both are, however, at least partially shaped by an individual's subjective judgment and knowledge (Warner and Campagna 2009).

As part of supervised classification, specific examples of units are selected from individual groups, which should have the most typical values possible, and based on these examples rules are designed for assigning all the remaining units to the types defined in advance. In contrast, as part of unsupervised classification, units can be categorized based on their characteristics or values even without any prior information about the units (Ferligoj 1989; Oštir 2006), which is a certain advantage compared to supervised classifications. The aim of unsupervised classification in groups is to achieve the maximum internal homogeneity and the minimum external isolation of groups (Ferligoj 1989) or to minimize variance within the group and maximize variance between groups (Rogerson 2006).

Unsupervised classification methods have several weaknesses (McGarigal, Cushman and Stafford 2000), such as sensitivity to outliers, which these methods often assign to a separate class, and great dependency of results on the initial groups defined and their number.

The unsupervised classification procedure can be divided into five basic steps (Ferligoj 1989; Theodoridis and Koutroumbas 2006):

- Step 1: selecting the units;
- Step 2: selecting the variables;
- Step 3: computing the similarities (and differences) between units;
- Step 4: selecting and carrying out the classification method; and
- Step 5: assessing the final classification, which can only be done by an expert in the relevant field.

Various relatively detailed digital data on natural geographical factors are available for Slovenia (Ciglič et al. 2016). Among these, over forty variables or data layers were selected and then normalized to values from 0 to 100, and adjusted to a uniform resolution of 200 m (the resolution of the least accurate layer), which means that Slovenia was divided into 506,450 units or cells. This completed the step 1 (selecting the units).

Step 2 entailed an assessment of all data layers in terms of their usefulness for landscape typology modeling (Ciglič 2012, 2013, 2014; Ciglič and Perko 2017). The following three criteria were applied: correlation between data layers, correlation between data layers and available landscape typologies, and suitability of data layers in terms of the classification level or scale (the scales were determined at which an individual data layer was still sufficiently diverse to be suitable for classification).

Eliminating less important data layers can reduce the time and costs of implementation, while also simplifying the understanding of modeling procedures (Jiang et al. 2008; Tirelli and Passani 2011). Based on the criteria mentioned above, the following four data layers were selected for landscape typology modeling: elevation, slope, rock permeability, and precipitation regime (the ratio between summer and fall precipitation).

Steps 3 and 4 already depend on the individual methods selected. Several unsupervised classification methods and their versions and settings in various software were tested for modeling the typologies of Slovenia. Four methods from the *TerrSet* software (the former Idrisi) were selected for presentation in this article: the histogram peak analysis method, the iterative self-organizing unsupervised classifier method, the *k*-means method, and the iterative self-organizing data analysis method.

3.1 Histogram peak analysis

The histogram peak analysis method is based on frequency distribution and classifies cells in groups using a multidimensional histogram (Richards and Jia 2006). It first looks for peaks in a multidimensional histogram (i.e., the values with the highest frequency) and then assigns every cell to its nearest peak, forming groups. The areas between the peaks (i.e., valleys) have values with the lowest frequency, creating boundaries between groups (Eastman 2016).

In the *TerrSet* program, this method is available as part of the CLUSTER module (TerrSet … 2015), in which the user can set the following parameters: number of data layers, number of classes for each data layer or gray levels, cutoffs for excluding extreme values or the saturation percentage, generalization level (broad or fine) for identifying peaks, and clustering rule, through which the user can drop less significant clusters, set the maximum number of clusters, or retain all clusters.

The following settings were selected: gray levels = 6, saturation percentage = 1%, generalization level = fine, and clustering rule = maximum nine groups.

3.2 The iterative self-organizing unsupervised classifier method

In *TerrSet*, the iterative self-organizing unsupervised classifier method is available as part of the ISOCLUST module (TerrSet … 2015), which in fact uses three other modules for this method. With the CLUSTER module it first derives the initial clusters (or seeds) using a multidimensional histogram (like with the histogram peak analysis; see Section 3.1), after which it employs the modules MAKESIG and MAXLIKE to automatically define the training sites and perform a supervised classification. It repeats the procedure several times using new training cells from individual clusters. Because of the efficiency of the seeding step, very few iterations are usually required to achieve a stable cluster.

The user first selects the data layers and then specifies the number of iterations, the desired number of clusters, and the minimum sample size per class.

The following settings were selected: number of iterations = 99, number of clusters desired = 9, and minimum sample size per class = 40.

3.3 *K*-means

Classification following the *k*-means method is based on distances between cells in multidimensional space. The final classification strongly depends on the definition of the number of clusters (*k)* and the initialization of centroids (centers) – that is, central cells around which other cells gather. First, the user specifies the number of clusters and the methods (rules) for initializing the cluster centroids, after which new, more appropriate centroids are calculated. This procedure is repeated until the new centroids are the same as the centroids from the previous iteration (Ferligoj 1989; Richards and Jia 2006). Initial centroids can also be selected randomly or dispersed evenly between cells (Ferligoj 1989).

In *TerrSet*, this method is available as part of the KMEANS module (TerrSet … 2015), in which different rules for initializing the centroids can be selected:

- The random partition rule randomly assigns each cell to one of the *k* clusters and then determines the initial centroids;
- The random seed rule randomly selects *k* cells as the initial centroids and then assigns each cell to one of the *k* clusters according to the minimum-distance rule; and
- The diagonal axis rule systematically sorts *k* centroids from the *n*-dimensional space from the minimum to the maximum value of *n* data layers.

In addition to specifying the rule for initializing the cluster centroids, the user selects the maximum number of output clusters and the option to merge (overly) small clusters, which do not exceed the selected percentage of the entire image cells, with larger ones. The user also specifies two stopping criteria to terminate the clustering process. With the first one, the process is terminated if during the last iteration the percentage of migrating cells is less than a specified percentage of all cells, and, with the second criterion, the process stops when a specified number of iterations has been reached.

The following settings were selected for this study: maximum number of output clusters = 9, cluster centroid initialization rule = random seed, merge clusters with proportions less than or equal to 1%, and stopping criteria = percentage of migration cells (pixels) less than or equal to 1% and maximum iterations 999.

3.4 Iterative self-organizing data analysis

The iterative self-organizing data analysis (ISODATA) is an improved *k*-means method. It can merge clusters, just like the *k*-means technique, but it can also split them. It determines the initial centroids and clusters the same way as the *k*-means method, after which it calculates the standard deviation within each cluster and the distances between cluster centroids. It splits a cluster into two if the standard deviation is higher than the one specified by the user, or it merges two clusters if the distance between two cluster centroids is smaller than the one specified by the user. It repeats the process with new clusters and new centroids. The process is terminated if the standard deviation and distances between centroids no longer make it possible to merge or split clusters, if during the last iteration the percentage of migrating pixels is less than that specified by the user or if the process reaches the number of iterations specified.

In *TerrSet*, this method is available as part of the ISODATA module (TerrSet … 2015). The following settings were specified for this study: initial number of clusters = 9, maximum number of output clusters = 9, cluster centroid initialization rule = random seed, stopping criteria = percentage of migration cells (pixels) less than or equal to 1% and maximum iterations 999, minimum cluster size = 100, standard deviation within a cluster for splitting = 25, Euclidean distance between clusters for merging = 12.5, and maximum number of pairs to merge within an iteration = 2.

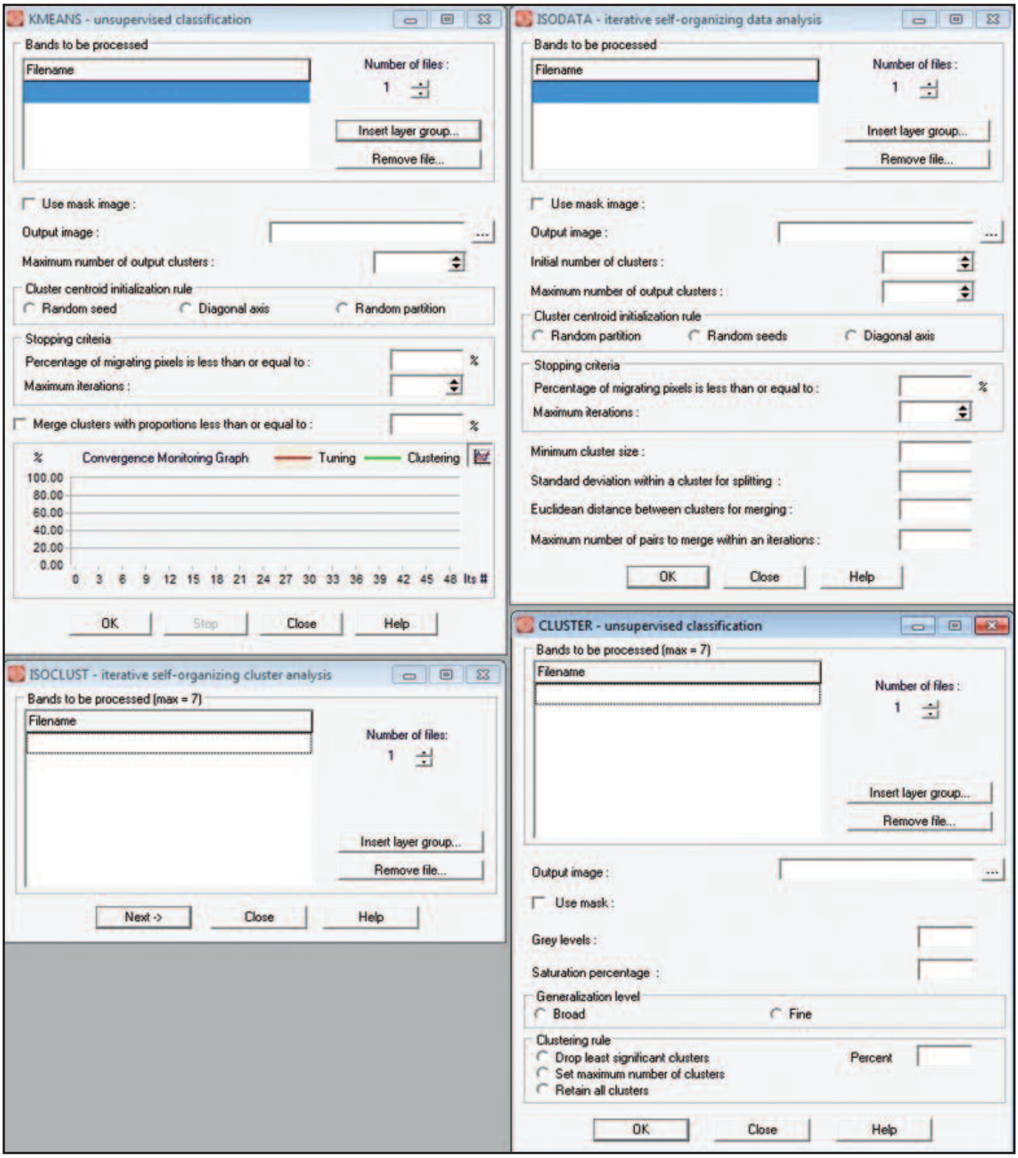

Figure 1: CLUSTER, ISOCLUST, KMEANS, and ISODATA modules in the TerrSet program.

4 Landscape typology modeling using unsupervised classification methods

Each of the four methods presented was used to produce typologies with various numbers of classes. This article describes in detail modeled landscape typologies with nine classes, which were compared against the best-established landscape typology of Slovenia of 1996, which contains nine landscape types combined into four landscape type groups.

The process of creating the 1996 landscape typology started in 1995. Perko (1998) entered the following four data layers into the geographic information system: surface elevation, surface inclination, lithology, and vegetation types. The inclination and elevation data were based on a 100 m digital elevation model, and the lithology and vegetation data were obtained through digitization of a 1:250,000 lithological map with thirty-seven basic units (Verbič 1998) and a vegetation map with sixty-two basic units (Zupančič et al. 1998) converted to a 100 m raster grid. All four layers were then generalized and simplified into seven classes.

Perko overlaid (intersected) all four layers. Altogether 2,401 different combinations were theoretically possible. Perko filtered the intersected layer three times using the modus inside of a moving 11 × 11 cell square window, obtaining forty-eight larger and spatially separate homogenous cores with the same combination of elevation, inclination, lithology, and vegetation. He printed the cores on a 1:250,000 map and, with the help of experts for individual parts of Slovenia, manually plotted the boundaries, mostly in morphological boundaries and larger watercourses. In the end, he combined these forty-eight manually delineated landscape units into nine landscape types, which he merged further into four landscape type groups.

The nine landscape types are: Alpine mountains, Alpine hills, Alpine plains, Pannonian low hills, Pannonian plains, Dinaric plateaus, Dinaric lowlands, Mediterranean low hills, and Mediterranean plateaus.

The four landscape type groups are Alpine landscapes, Pannonian landscapes, Dinaric landscapes, and Mediterranean landscapes.

This was the first partly computerized typology of Slovenia. Its research bases were first presented in 1998 (Perko 1998). It has been published in all major geographical works on Slovenia issued after Slovenia's independence. Since 2008, it has also been part of Slovenian tax legislation and has been used for rating agricultural land according to the Rules on Determining and Administering Land Rating.

As already described in Chapter 3 the CLUSTER, ISOCLUST, KMEANS, and ISODATA modules in the *TerrSet* software were used for modeling. The modeled typologies were compared against the original 1996 typology in terms of the correlation coefficient between the modeled typologies and the original typology, the cluster density in the modeled typologies by landscape type of the original typology, and the ratio between the actual and theoretical cluster frequency in the modeled typologies by landscape type of the original typology.

The first indicator was **cluster density** by landscape type of the original typology – that is, the number of cells in an individual cluster per 100 cells of a specific landscape type. The maximum value of density possible is 100, which is when all the cells of an individual cluster lie within a specific landscape type, and the minimum value is 0, when a specific landscape type does not contain even a single cell of this cluster.

If the cells of all nine clusters were evenly distributed across the nine landscape types of the original typology in the modeled typologies, the density of all clusters would be 11. A cluster with a density of at least twice as much (i.e., at least 22) was defined as a typical representative of this landscape type. If a modeled typology was completely the same as the original typology, each of the nine clusters would have a density of 100 in only one landscape type, whereas the density in the remaining eight landscape types would be 0. If, for instance, Cluster 7 of a modeled typology had thirty-three cells per one hundred cells of Dinaric plateaus, Cluster 7 would be a good representative or approximation of the Dinaric plateaus in the original typology.

The indicator **ratio between the actual and theoretical cluster frequency** in the modeled typologies by landscape type of the original typology relies on contingency tables, in which the rows and columns represent clusters and landscape types, and the cells contain the actual frequency (number) of cells in individual clusters by individual landscape type. Clusters that had their actual frequency in a specific landscape type at least twice the theoretical frequency were defined as typical representatives of this landscape type.

For example, in Table 1, at the intersection of Cluster 9, which was specified using the ISODATA method, and the Alpine mountains landscape type the actual frequency – that is, the number of cells of Cluster 9 that lie within the Alpine mountains (their total is 27,975) – is provided. The theoretical frequency of that table cell is 4,811 and equals the total of all cells of Cluster 9 in the last column (31,835) and all cells of the Alpine mountains in the last row (76,533) divided by the number of all cells (506,450). Because the ratio between the actual and theoretical frequency (5.82) is greater than 2, Cluster 9 is a good representative of the Alpine mountains.

The **correlation coefficient** between the modeled typologies and the original typology also relies on contingency tables. Cramer's *V* was selected for the study; it has lower values than the Pearson correlation coefficient, but it does not depend on the size of the tables and therefore allows comparisons between tables with different numbers of columns and rows or between typologies with different numbers of classes.

Table 1: Example of arranging cells of a nine-cluster typology modeled using the ISODATA module by nine landscape types of the original typology.

ISODATA module	Alpine mountains	Alpine hills	Alpine plains	Pannonian low hills	Pannonian plains	Dinaric plateaus	Dinaric lowlands	Medi-terranean low hills	Medi-terranean plateaus	Total
Cluster 1	5,288	34,123	2,382	31,609	1,264	3,747	5,448	0	0	83,861
Cluster 2	250	6,543	2,435	30,360	25,360	0	0	0	0	64,948
Cluster 3	5,762	29,497	143	593	11	78	88	0	0	36,172
Cluster 4	8,365	1,395	224	0	0	33,590	9,724	5,654	13,885	72,837
Cluster 5	4,516	10,161	888	0	0	12,956	6,570	18,477	2,738	56,306
Cluster 6	14,932	8,997	115	5	0	11,493	366	663	183	36,754
Cluster 7	2,379	4,470	13,832	3,349	5,545	20,106	20,861	1,681	27	72,250
Cluster 8	7,066	20,840	447	8,803	156	9,845	4,330	0	0	51,487
Cluster 9	27,975	452	15	0	0	3,375	0	15	3	31,835
Total	76,533	116,478	20,481	74,719	32,336	95,190	47,387	26,490	16,836	506,450

4.1 Modeling using nine clusters

A graphic presentation of cluster distribution according to the four selected unsupervised classification methods using nine clusters is provided in Figure 2. The original typology with nine landscape types and the original typology with four landscape type groups are added for comparison.

The **correlation coefficients** between the original classification and the classifications produced using the four modules presented above are as follows: 0.4489 for the CLUSTER classification, 0.4159 for the ISOCLUST classification (the lowest among all four classifications), 0.4609 for the KMEANS classification (the highest among all classifications), and 0.4518 for the ISODATA classification.

The **density** of cells in an individual cluster by nine landscape types (Table 2) shows how well the clusters spatially match the landscape types. Clusters with a density of 22 or more are good representatives or approximations of a specific landscape type.

In the CLUSTER classification, Cluster 1 is a good approximation of the Alpine hills and the Pannonian low hills, Cluster 2 is a good approximation of the Alpine plains, Pannonian plains, Pannonian low hills, and Dinaric lowlands, Cluster 3 of the Alpine mountains and Dinaric plateaus, Cluster 5 of the Pannonian plains, and Clusters 6 and 9 of the Mediterranean low hills and Mediterranean plateaus, whereas Clusters 4, 7, and 8 are not a good approximation of even a single landscape type. Cluster 4 demonstrates the highest cell density (18) in the Alpine hills, Cluster 7 in the Alpine mountains (18), and Cluster 8 in the Dinaric plateaus (15). Thus, Cluster 1 has a density over 22 in two landscape types, Cluster 2 in four types, Cluster 3 in two types, Cluster 5 in one type, and both Clusters 6 and 9 in two types. All landscape types are represented: five types in one cluster and four types in two clusters, which means that these clusters are a good spatial match with two landscape types, not just one.

In the ISOCLUST classification, Cluster 1 is a good approximation of the Alpine hills, Cluster 2 of the Dinaric plateaus, Cluster 3 of the Dinaric plateaus, Dinaric lowlands, and Mediterranean plateaus, Cluster 4 of the Mediterranean low hills, Cluster 7 of the Pannonian low hills and Pannonian plains, Cluster 8 of the Alpine plains and Pannonian plains, and Cluster 9 of the Alpine hills and Pannonian low hills, whereas Clusters 5 and 6 are not a good approximation of even a single landscape type. Cluster 5 has the highest density (17) in the Alpine hills and Cluster 6 in the Pannonian low hills (20). Thus, Cluster 1 has a density over 22 in one landscape type of the original typology, Cluster 2 the same, Cluster 3 in four types, and Clusters 7, 8 and 9 in two types. All landscape types are represented: six in one cluster and three in two clusters.

In the KMEANS classification, Cluster 1 is a good approximation of the Alpine hills and Pannonian low hills, Cluster 3 of the Alpine plains and Dinaric lowlands, Cluster 4 of the Alpine mountains, Cluster 6 of the Alpine hills, Cluster 7 of the Mediterranean low hills, Cluster 8 of the Pannonian low hills and Pannonian plains, and Cluster 9 of the Dinaric plateaus, Dinaric lowlands, and Mediterranean plateaus,

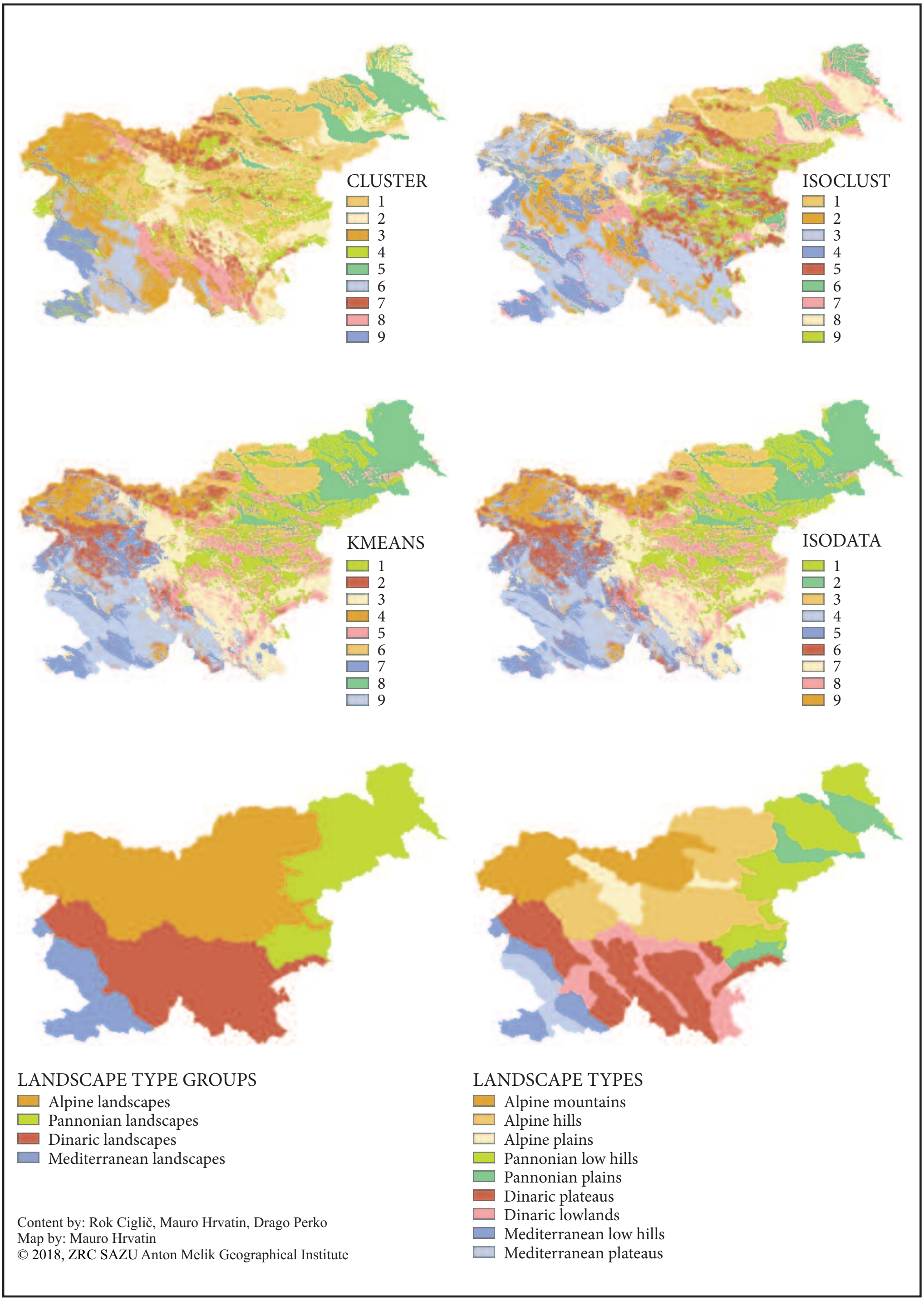

Figure 2: Nine-cluster typologies modeled using the unsupervised classification methods (in the key the clusters are ordered based on the percentage of cells in an individual cluster: the one with the highest percentage is at the top and the one with the lowest percentage is at the bottom).

whereas Clusters 2 and 5 are not a good approximation of even a single landscape type. Cluster 2 displays the highest density (21) in the Alpine mountains and Cluster 5 in the Alpine hills (18). Thus, Cluster 1 has a density above 22 in two landscape types of the original typology, as does Cluster 3, Clusters 4, 6, and 7 in one type, Cluster 8 in two types, and Cluster 9 in three types. All landscape types are represented: six in one cluster and three in two clusters.

In the ISODATA classification, Cluster 1 is a good approximation of the Alpine hills and Pannonian low hills, Cluster 2 of the Pannonian low hills and Pannonian plains, Cluster 3 of the Alpine hills, Cluster 4 of the Dinaric plateaus and Mediterranean plateaus, Cluster 5 of the Mediterranean low hills, Cluster 7 of the Alpine plains and Dinaric lowlands, and Cluster 9 of the Alpine mountains, whereas Clusters 6 and 8 are not a good approximation of even a single landscape type. Cluster 6 has the highest density (20) in the Alpine mountains and Cluster 8 in the Alpine hills (18). Thus, Cluster 1 has a density of over 22 in two landscape types, Cluster 2 the same, Cluster 3 in one type, Cluster 4 in two types, Cluster 5 in one type, Cluster 7 in two types, and Cluster 9 in one. All landscape types are represented: seven types in one cluster and two in two clusters.

The indicator **ratio between the actual and theoretical cluster frequency** in the modeled typologies by landscape type of the original typology (Table 3) shows which clusters are good representatives of a specific landscape type. Good representatives are the clusters whose actual frequency is at least twice their theoretical frequency.

In the CLUSTER classification, Cluster 1 is a good approximation of the Alpine hills (a ratio of 2.37), Cluster 2 of the Alpine plains (3.93) and Dinaric lowlands (2.45), Cluster 3 of the Alpine mountains (3.08), Cluster 5 of the Pannonian plains (as much as 8.99), Cluster 6 of the Dinaric lowlands (3.03), Mediterranean low hills (3.66), and Mediterranean plateaus (5.83), Cluster 7 of the Alpine mountains (3.09), Cluster 8 of the Dinaric plateaus (3.02) and Dinaric lowlands (2.78), and Cluster 9 of the Mediterranean low hills (as much as 9.78) and Mediterranean plateaus (even 10.79), and Cluster 4 has a ratio below 2 with all the landscape types. Its ratio is the highest with the Alpine hills (1.44). Three landscape types with a ratio above 2 appear in one cluster, four in two clusters, the Dinaric lowlands even in three clusters, and the Pannonian low hills in none. The ratio of the Pannonian low hills is the highest in Cluster 5 (i.e., 1.79, which is close to the threshold value of 2).

In the ISOCLUST classification, Cluster 1 is a good approximation of the Alpine hills (a ratio of 3.47), Cluster 2 of the Alpine mountains (2.08) and Dinaric plateaus (2.40), Cluster 3 of the Dinaric plateaus (2.35) and Mediterranean plateaus (3.29), Cluster 4 of the Mediterranean low hills (6.01), Cluster 5 of the Alpine hills (2.11), Cluster 6 of the Pannonian low hills (4.96), Cluster 7 of the Pannonian low hills (2.65) and Pannonian plains (2.73), Cluster 8 of the Alpine plains (4.88) and Pannonian plains (6.45), and Cluster 9 of the Pannonian low hills (2.94). Four landscape types with a ratio above 2 appear in one cluster, three in two clusters, and the Pannonian hills even in three clusters.

In the KMEANS classification, Cluster 1 is a good approximation of the Pannonian low hills (a ratio of 2.58), Cluster 2 of the Alpine mountains (2.98), Cluster 3 of the Alpine plains (4.75) and Dinaric lowlands (3.13), Cluster 4 of the Alpine mountains (5.81), Cluster 6 of the Alpine hills (3.55), Cluster 7 of the Mediterranean low hills (2.65), Cluster 8 of the Pannonian low hills (3.12) and Pannonian plains (6.21), and Cluster 9 of the Dinaric plateaus and Mediterranean plateaus, and Cluster 5 has a ratio below 2 with all the landscape types. Its ratio is the highest with the Alpine hills (1.69). Seven landscape types with a ratio above 2 appear in one cluster and two in two clusters.

In the ISODATA classification, Cluster 1 is a good approximation of the Pannonian low hills (a ratio of 2.55), Cluster 2 of the Pannonian low hills (3.17) and Pannonian plains (6.12), Cluster 3 of the Alpine hills (3.55), Cluster 4 of the Dinaric plateaus (2.45) and Mediterranean plateaus (5.73), Cluster 5 of the Mediterranean low hills (6.27), Cluster 6 of the Alpine mountains (2.69), Cluster 7 of the Alpine plains (4.73) and Dinaric lowlands (3.09), and Cluster 9 of the Alpine mountains. Cluster 8 has a ratio below 2 with all the landscape types. Its ratio is the highest with the Alpine hills (1.76). Seven landscape types with a ratio above 2 appear in one cluster and two in two clusters.

Taking into account all three indicators presented, the typology modeled using the iterative self-organizing data analysis method (the ISODATA module) and the typology modeled using the *k*-means method (the KMEANS module) are the closest to the original typology of nine landscape types.

The KMEANS typology has a slightly higher correlation coefficient than the ISODATA typology but, on the other hand, the ISODATA typology displays a density above the threshold value of 22 in two clusters simultaneously only with two landscape types compared to three in the KMEANS typology.

The typology modeled using the iterative self-organizing unsupervised classifier method (the ISOCLUST module) matches the original typology the least.

Table 2: Cell density of an individual cluster in the modeled typologies by nine landscape types of the original typology (green numbers indicate a density of 22 or more).

Modules	Alpine mountains	Alpine hills	Alpine plains	Pannonian low hills	Pannonian plains	Dinaric plateaus	Dinaric lowlands	Mediterranean low hills	Mediterranean plateaus	Total
CLUSTER 1	17.79	62.21	13.76	42.76	3.86	4.16	14.76	0.03	0.00	26.27
CLUSTER 2	1.25	6.12	63.91	26.86	28.21	13.01	39.76	2.60	0.27	16.26
CLUSTER 3	47.27	3.55	0.06	0.00	0.00	34.81	1.63	8.47	7.97	15.37
CLUSTER 4	10.25	18.15	4.85	16.62	0.58	13.17	8.11	17.97	1.75	12.64
CLUSTER 5	0.47	3.87	6.07	13.44	67.34	0.00	0.00	0.00	0.00	7.49
CLUSTER 6	1.11	0.58	0.00	0.00	0.00	10.01	19.91	24.05	38.33	6.58
CLUSTER 7	17.77	4.68	3.22	0.32	0.01	8.50	2.27	0.16	0.00	5.76
CLUSTER 8	2.02	0.82	8.12	0.00	0.00	14.74	13.56	0.13	0.31	4.88
CLUSTER 9	2.07	0.02	0.00	0.00	0.00	1.59	0.00	46.58	51.37	4.76
ISOCLUST 1	8.25	29.96	1.43	0.81	0.03	1.42	0.39	0.04	0.11	8.63
ISOCLUST 2	19.67	5.15	0.88	0.01	0.00	22.71	6.51	0.22	11.25	9.46
ISOCLUST 3	30.42	3.31	16.48	0.14	0.09	54.97	42.06	10.07	76.94	23.40
ISOCLUST 4	20.63	11.06	2.32	0.04	0.00	8.69	6.64	71.91	5.43	11.96
ISOCLUST 5	3.70	17.18	2.65	9.28	0.64	5.23	12.08	0.00	0.00	8.14
ISOCLUST 6	1.21	1.96	0.97	20.32	1.66	0.70	0.12	0.31	4.80	4.10
ISOCLUST 7	0.16	4.19	13.36	25.94	26.69	1.52	18.32	13.40	0.79	9.79
ISOCLUST 8	13.51	5.00	50.69	1.94	66.97	1.53	0.87	4.05	0.10	10.39
ISOCLUST 9	2.45	22.18	11.23	41.53	3.91	3.23	13.00	0.00	0.58	14.15
KMEANS 1	6.73	29.41	11.18	42.73	3.91	3.91	11.29	0.00	0.00	16.58
KMEANS 2	21.05	6.55	0.64	0.10	0.00	11.58	0.87	1.06	0.98	7.07
KMEANS 3	3.36	3.53	67.39	3.75	17.05	20.48	44.39	9.21	0.43	14.19
KMEANS 4	35.81	0.39	0.08	0.00	0.00	3.50	0.00	0.04	0.02	6.17
KMEANS 5	7.93	18.11	2.47	13.21	0.58	11.81	11.00	0.00	0.00	10.70
KMEANS 6	7.26	23.70	0.41	0.79	0.03	0.00	0.00	0.00	0.00	6.68
KMEANS 7	6.99	10.95	5.06	0.00	0.00	7.20	8.59	71.25	6.52	9.88
KMEANS 8	0.35	5.50	11.94	39.43	78.43	0.00	0.00	0.00	0.00	12.62
KMEANS 9	10.52	1.86	0.82	0.00	0.00	41.52	23.86	18.44	92.05	16.11
ISODATA 1	6.91	29.30	11.63	42.30	3.91	3.94	11.50	0.00	0.00	16.56
ISODATA 2	0.33	5.62	11.89	40.63	78.43	0.00	0.00	0.00	0.00	12.82
ISODATA 3	7.53	25.32	0.70	0.79	0.03	0.08	0.19	0.00	0.00	7.14
ISODATA 4	10.93	1.20	1.09	0.00	0.00	35.29	20.52	21.34	82.47	14.38
ISODATA 5	5.90	8.72	4.34	0.00	0.00	13.61	13.86	69.75	16.26	11.12
ISODATA 6	19.51	7.72	0.56	0.01	0.00	12.07	0.77	2.50	1.09	7.26
ISODATA 7	3.11	3.84	67.54	4.48	17.15	21.12	44.02	6.35	0.16	14.27
ISODATA 8	9.23	17.89	2.18	11.78	0.48	10.34	9.14	0.00	0.00	10.17
ISODATA 9	36.55	0.39	0.07	0.00	0.00	3.55	0.00	0.06	0.02	6.29

Table 3: Ratio between the actual and theoretical cluster frequency in the modeled typologies by nine landscape types of the original typology (green numbers indicate ratios of 2 or more).

Modules	Alpine mountains	Alpine hills	Alpine plains	Pannonian low hills	Pannonian plains	Dinaric plateaus	Dinaric lowlands	Medi-terranean low hills	Medi-terranean plateaus	Total
CLUSTER 1	0.68	2.37	0.52	1.63	0.15	0.16	0.56	0.00	0.00	1.00
CLUSTER 2	0.08	0.38	3.93	1.65	1.74	0.80	2.45	0.16	0.02	1.00
CLUSTER 3	3.08	0.23	0.00	0.00	0.00	2.27	0.11	0.55	0.52	1.00
CLUSTER 4	0.81	1.44	0.38	1.31	0.05	1.04	0.64	1.42	0.14	1.00
CLUSTER 5	0.06	0.52	0.81	1.79	8.99	0.00	0.00	0.00	0.00	1.00
CLUSTER 6	0.17	0.09	0.00	0.00	0.00	1.52	3.03	3.66	5.83	1.00
CLUSTER 7	3.09	0.81	0.56	0.06	0.00	1.48	0.39	0.03	0.00	1.00
CLUSTER 8	0.41	0.17	1.67	0.00	0.00	3.02	2.78	0.03	0.06	1.00
CLUSTER 9	0.44	0.00	0.00	0.00	0.00	0.33	0.00	9.78	10.79	1.00
ISOCLUST 1	0.96	3.47	0.17	0.09	0.00	0.17	0.05	0.00	0.01	1.00
ISOCLUST 2	2.08	0.55	0.09	0.00	0.00	2.40	0.69	0.02	1.19	1.00
ISOCLUST 3	1.30	0.14	0.70	0.01	0.00	2.35	1.80	0.43	3.29	1.00
ISOCLUST 4	1.73	0.92	0.19	0.00	0.00	0.73	0.56	6.01	0.45	1.00
ISOCLUST 5	0.45	2.11	0.33	1.14	0.08	0.64	1.48	0.00	0.00	1.00
ISOCLUST 6	0.30	0.48	0.24	4.96	0.41	0.17	0.03	0.07	1.17	1.00
ISOCLUST 7	0.02	0.43	1.37	2.65	2.73	0.16	1.87	1.37	0.08	1.00
ISOCLUST 8	1.30	0.48	4.88	0.19	6.45	0.15	0.08	0.39	0.01	1.00
ISOCLUST 9	0.17	1.57	0.79	2.94	0.28	0.23	0.92	0.00	0.04	1.00
KMEANS 1	0.41	1.77	0.67	2.58	0.24	0.24	0.68	0.00	0.00	1.00
KMEANS 2	2.98	0.93	0.09	0.01	0.00	1.64	0.12	0.15	0.14	1.00
KMEANS 3	0.24	0.25	4.75	0.26	1.20	1.44	3.13	0.65	0.03	1.00
KMEANS 4	5.81	0.06	0.01	0.00	0.00	0.57	0.00	0.01	0.00	1.00
KMEANS 5	0.74	1.69	0.23	1.23	0.05	1.10	1.03	0.00	0.00	1.00
KMEANS 6	1.09	3.55	0.06	0.12	0.01	0.00	0.00	0.00	0.00	1.00
KMEANS 7	0.71	1.11	0.51	0.00	0.00	0.73	0.87	7.21	0.66	1.00
KMEANS 8	0.03	0.44	0.95	3.12	6.21	0.00	0.00	0.00	0.00	1.00
KMEANS 9	0.65	0.12	0.05	0.00	0.00	2.58	1.48	1.14	5.71	1.00
ISODATA 1	0.42	1.77	0.70	2.55	0.24	0.24	0.69	0.00	0.00	1.00
ISODATA 2	0.03	0.44	0.93	3.17	6.12	0.00	0.00	0.00	0.00	1.00
ISODATA 3	1.05	3.55	0.10	0.11	0.00	0.01	0.03	0.00	0.00	1.00
ISODATA 4	0.76	0.08	0.08	0.00	0.00	2.45	1.43	1.48	5.73	1.00
ISODATA 5	0.53	0.78	0.39	0.00	0.00	1.22	1.25	6.27	1.46	1.00
ISODATA 6	2.69	1.06	0.08	0.00	0.00	1.66	0.11	0.34	0.15	1.00
ISODATA 7	0.22	0.27	4.73	0.31	1.20	1.48	3.09	0.44	0.01	1.00
ISODATA 8	0.91	1.76	0.21	1.16	0.05	1.02	0.90	0.00	0.00	1.00
ISODATA 9	5.82	0.06	0.01	0.00	0.00	0.56	0.00	0.01	0.00	1.00

4.2 Modeling using four clusters

A graphic presentation of cluster distribution according to the four selected unsupervised classification methods using four clusters is provided in Figure 3. The original typology with four landscape type groups and the original typology with nine landscape types are added for comparison.

The four-cluster typologies modeled were also compared against the original typology. Taking into account all three indicators presented, the typology modeled using the iterative self-organizing data analysis method (the ISODATA module) is the closest to the original typology with four landscape type groups and the original typology with nine landscape types, and the typology modeled using the iterative self-organizing unsupervised classifier method (the ISOCLUST module) match the original typologies the least.

It is interesting that the correlation coefficients between the modeled typologies and the original typology with nine landscape types are approximately one-third higher than with the original typology with four landscape type groups.

5 Quality of the typologies modeled

A comparison between the nine-cluster typologies modeled showed that the ISODATA and KMEANS typologies are the closest to the original typology with nine landscape types and can therefore be regarded as the best approximations. This applies to the modeled typologies as a whole, but the question was whether this also applies to individual landscape types or the other two modules may prove to be more effective in assigning cells to the cluster that is the best approximation of a specific landscape type (e.g., the Alpine mountains).

The **Herfindahl–Hirschman Index** was used to calculate the concentration of nine landscape types of the original typology by nine clusters of the modeled typologies and vice versa (Tables 4 and 5). The value of the index ranges from 0 to 1. In the case presented, it had value 1 if all the cells of a specific landscape type were in only one cluster and it had value 0 if the cells of a specific landscape type were evenly distributed across all clusters. The higher the index value, the better a cluster of a specific module is an approximation of a specific landscape type.

Every landscape type has the highest concentration index with one module (marked green in Table 4) and the lowest with another (marked red in Table 4). In the CLUSTER typology, two landscape types – the Alpine mountains and the Alpine hills – are the most concentrated across clusters (i.e., they have the highest concentration index), and as many as four types are concentrated the least: the Dinaric plateaus, Dinaric lowlands, Mediterranean low hills, and Mediterranean plateaus. In the ISOCLUST typology, only the Dinaric plateaus are the most concentrated across clusters, whereas as many as five are concentrated the least: the Alpine mountains, Alpine hills, Alpine plains, Pannonian low hills, and Pannonian plains. In the KMEANS typology, the three types most concentrated across clusters are the Dinaric lowlands, Mediterranean low hills, and Mediterranean plateaus, whereas no type displayed a particularly low concentration. Similarly, in the ISODATA typology, three types were most concentrated across clusters (i.e., the Alpine plains, Pannonian low hills, and Pannonian plains), and no type displayed a particularly low concentration. The values of the concentration indexes show that the CLUSTER module is the best for defining Alpine landscape types, the KMEANS module for determining Mediterranean landscape types, and the ISODATA module for defining Pannonian landscape types (Table 4). Differences between the typologies modeled are also evident from the opposite perspective – that is, in terms of cluster concentration by landscape type (Table 5). Among the nine clusters used, the CLUSTER module proved to be the best for three clusters and the worst for three clusters, the ISOCLUST module was the best for one cluster, the KMEANS module was the best for two clusters and the worst for two clusters, and the ISODATA module was the best for three clusters and the worst for as many as four (Table 5).

This means that individual landscape types can be determined more effectively by other modules than the one used for a specific modeled typology; for example, the CLUSTER module for Alpine mountains and Alpine hills. The concentration index for Alpine mountains is 41% higher with the CLUSTER module than the KMEAN module, and even 93% higher for Alpine hills.

The quality of the modeled typologies is also indicated by the degree of their similarity with the original typology of 1996. According to the three indicators based on which the modeled typologies were

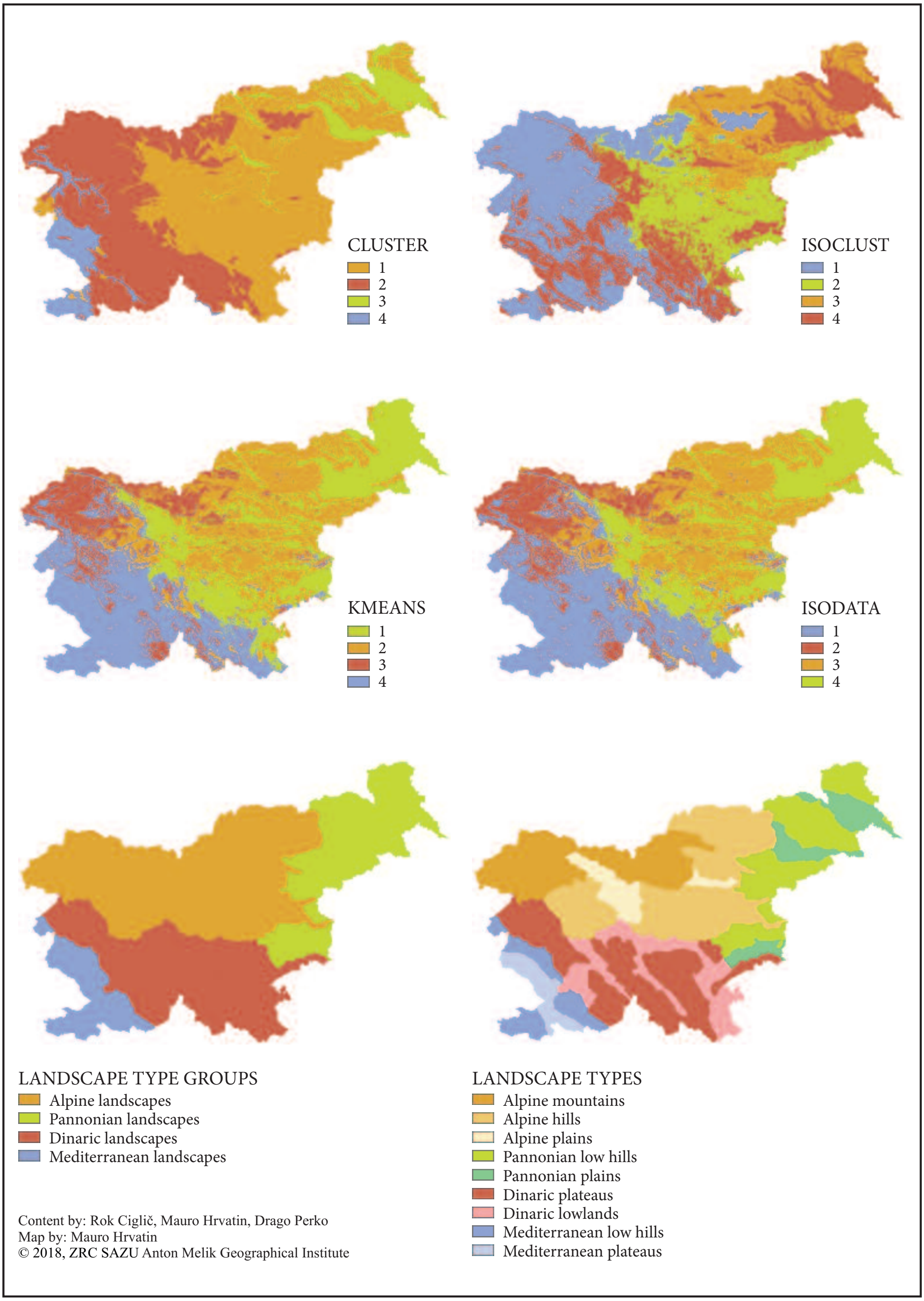

Figure 3: Four-cluster typologies modeled using the unsupervised classification method (in the key the clusters are ordered based on the percentage of cells in an individual cluster: the one with the highest percentage is at the top and the one with the lowest percentage is at the bottom).

compared against the original one, the best modeled typologies are those that have the highest correlation coefficients, very high densities by individual cells of the contingency table and at the same time very low densities by other cells of the contingency table, and very high ratios by individual cells of the contingency table and at the same time very low ratios by other cells in that table.

If the double density of cells in an individual cluster of modeled typologies by nine landscape types of the original typology (Table 2) is taken into account as a criterion and the frequencies in the cells of the contingency table that meet this condition are added up (values density of 20 or more are taken into account), the ISOCLUST module has 66% of all cells »properly« classified, the corresponding share with the KMEANS and ISODATA modules is 65%, and with the CLUSTER module it is 59%. Based on the criterion of the double ratio between the actual and theoretical cluster frequency of modeled typologies by nine landscape types of the original typology (Table 3), the ISOCLUST module has 56% of cells »properly« classified, compared to 53% in the KMEANS module and 51% in the ISODATA module. The differences between the modules are very small. In terms of density, nearly two-thirds of the cells are »properly« classified, compared to only just over half in terms of the frequency ratio.

The degree to which the typologies modeled using **unsupervised classification** methods are less correlated with the original typology of 1996 compared to the typologies modeled using **supervised classification** methods was assessed by comparing Cramer's *V* (similar to how the modeled typologies were compared against the original one). The supervised classification models were designed using four supervised classification methods: a decision tree, *k*-nearest neighbors, maximum likelihood, and minimum distance. With all four the same four data layers were used as with the unsupervised classifications. The models were designed using two sets of **training cells**. For the first set, the cells were selected **randomly** and for the second set **expert** sampling was used, which means that the most representative areas were selected based on the researcher's judgment. Thus, with all methods, random supervised classifications were distinguished from the expert ones, which yielded eight modeled typologies. Typologies using supervised classification methods were modeled in 2013 (Ciglič 2014).

At first glance, it may seem unusual that the typologies modeled using a random sample generally had a higher degree of correlation with the original typology (the correlation coefficient ranging from 0.5023

Table 4: Concentration indexes of landscape types of the original typology by clusters of modeled typologies (colors indicate the modeled typology or module where an individual landscape type is concentrated the most (green) or the least (red)).

Modules	Alpine mountains	Alpine hills	Alpine plains	Pannonian low hills	Pannonian plains	Dinaric plateaus	Dinaric lowlands	Mediterranean low hills	Mediterranean plateaus
CLUSTER	0.4589	0.5978	0.6092	0.4618	0.6902	0.3098	0.3885	0.4789	0.5871
ISOCLUST	0.3179	0.2962	0.4803	0.4488	0.6795	0.5353	0.3963	0.7001	0.7491
KMEANS	0.3263	0.3098	0.6479	0.5259	0.7755	0.3947	0.4439	0.7029	0.9128
ISODATA	0.3285	0.3168	0.6497	0.5292	0.7757	0.3431	0.4317	0.6920	0.8186
Average	0.3579	0.3801	0.5968	0.4914	0.7302	0.3957	0.4151	0.6435	0.7669

Table 5: Cluster concentration indexes of modeled typologies by landscape type of the original typology (colors indicate the modeled typology or module where an individual cluster is concentrated the most (green) or the least (red)).

Modules	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
CLUSTER	0.5385	0.2778	0.5715	0.3326	0.5847	0.3768	0.4979	0.5700	0.5689
ISOCLUST	0.7859	0.4893	0.4308	0.3715	0.4642	0.7050	0.3698	0.4110	0.4951
KMEANS	0.4869	0.5102	0.3325	0.8686	0.3949	0.8086	0.3881	0.5512	0.4658
ISODATA	0.4834	0.5526	0.8075	0.4441	0.3433	0.4885	0.3351	0.4032	0.8698
Average	0.5737	0.4575	0.5356	0.5042	0.4468	0.5947	0.3977	0.4838	0.5999

with the minimum distance method to 0.7261 with the *k*-nearest neighbors method) than the ones for which an expert sample was used (0.5153 with the minimum distance method and 0.6020 with the *k*-nearest neighbors method). The reason is that in expert-sample modeling some of the most typical areas of individual types (based on the researcher's subjective judgment) were selected, and therefore the modeling did not cover the entire variability of a landscape type or at least not to the same extent as the random-sample modeling, where, given the more generalized classification rules, the sample was more evenly distributed across a landscape type. Hence, with expert-sample modeling one can speak of **over-fitting**. This is also proven by the analysis of testing the success rate of modeling the training cells, where (precisely to the contrary) expert sampling achieved significantly higher scores than random sampling (Ciglič 2014).

The correlation coefficients between the typologies modeled using unsupervised classification methods and the original typology range from 0.4159 with the ISOCLUSTER module to 0.4609 with the KMEANS module. This means they are approximately one-fifth lower than the correlation coefficients between the original typology and the typologies modeled using expert supervised classification methods, and a third lower than the correlation coefficients between the original typology and the typologies modeled using random supervised classification methods.

6 Conclusion

The typologies modeled using the unsupervised classification methods presented only roughly approximate the original typology of Slovenia and to a lesser degree than the supervised classification models (Ciglič 2014). This is expected because typologies modeled using supervised classification methods are based on a referential or training classification or, in this case, the 1996 typology of Slovenia. On the other hand, the statistically significant correlation between the original typology of 1996 and the typologies modeled using unsupervised classification methods (i.e., completely independently of already identified landscape types) confirms that the original landscape typology (at least in terms of the data layers that were taken into account) is fairly appropriate.

However, because unsupervised classification methods have certain advantages over supervised classification methods and because the number and diversity of data covering all of Slovenian territory is growing, it can be expected that the differences between the quality of supervised and unsupervised classifications will become smaller and that modeled typologies will better approximate the actual situation, thus effectively replacing typologies produced using traditional procedures, which are usually more time-consuming and therefore more expensive.

ACKNOWLEDGEMENT: The authors acknowledge financial support from the Slovenian Research Agency (program no. P6-0101: Geography of Slovenia and project no. L1-7542: Advancement of computationally intensive methods for efficient modern general-purpose statistical analysis and inference).

7 References

Abonyi, J., Feil, B. 2007: Cluster analysis for data mining and system identification. Basel, Boston, Berlin, Birkhäuser.

Bailey, R. G. 1996: Ecosystem geography. New York.

Bastian, O. 2000: Landscape classification in Saxony (Germany) – a tool for holistic regional planning. Landscape and urban planning 50, 1–3. DOI: https://doi.org/10.1016/S0169-2046(00)00086-4

Bernert, J. A., Eilers, J. M., Freemark, K. E., Ribic, C. 1997: A quantitative method for delineating regions: An example for the western Corn Belt Plains Ecoregion of the USA. Environmental Management 21-3. DOI: https://doi.org/10.1007/s002679900038

Brabyn, L. 1998: GIS analysis of macro landform. 10th colloquium of the Spatial Information Research Centre, University of Otago. Dunedin.

Brabyn, L. 2018: NZ Landscape classification version II. Internet: https://www.waikato.ac.nz/eri/about/people/larsb/New-Zealand-Landscape-Classification-v2.1.pdf (16. 12. 2018).

Bunce, R. G. H, Barr, C. J., Clarke, R. T., Howard, D. C., Lane, A. M. J. 1996: Land classification for strategic ecological survey. Journal of environmental management 47-1. DOI: https://doi.org/10.1006/jema.1996.0034

Ciglič, R. 2012: Evaluation of digital data layers for establishing natural landscape types in Slovenia. Landscapes: Perception, knowledge, awareness, and action. Proceedings of the FG-SHU International Symposium on Geography. New York.
Ciglič, R. 2013: Uporaba geografskih informacijskih sistemov za določanje naravnih pokrajinskih tipov Slovenije. Ph.D. thesis, Univerza v Ljubljani, Filozofska fakulteta. Ljubljana.
Ciglič, R. 2014: Analiza naravnih pokrajinskih tipov Slovenije z GIS-om. Geografija Slovenije 28. Ljubljana.
Ciglič, R. 2018: Assessing the impact of input data incongruity in selected quantitative methods for modelling natural landscape typologies. Geografski vestnik 90-1. DOI: https://doi.org/10.3986/GV90107
Ciglič, R., Geršič, M., Perko, D., Zorn, M. (eds.) 2016: Digitalni podatki. GIS v Sloveniji 13. Ljubljana.
Ciglič, R., Perko, D. 2012: Slovenia in geographical typifications and regionalizations of Europe. Geografski vestnik 84-1.
Ciglič, R., Perko, D. 2013: Europe's landscape hotspots. Acta geographica Slovenica 53-1. DOI: https://doi.org/10.3986/AGS53106
Ciglič, R., Perko, D. 2015: Modelling as a method for evaluating natural landscape typology: The case of Slovenia. Landscape Analysis and Planning. Cham. DOI: https://doi.org/10.1007/978-3-319-13527-4
Ciglič, R., Perko, D. 2017: A method for evaluating raster data layers according to landscape classification scale. Ecological Informatics 39. DOI: https://doi.org/10.1016/j.ecoinf.2017.03.004
Ciglič, R., Perko, D., Hrvatin, M., Štrumbelj, E. 2017: Modeling and evaluating older landscape classifications with modern quantitative methods. From pattern and process to people and action. Ghent.
Clark, A. N. 1998: The Penguin dictionary of geography. London.
Demeritt, D., Wainwright, J. 2005: Models, modelling and geography. Questioning geography: fundamental Debates. Malden.
Dikau, R., Brabb, E. E., Mark, R. K. 1991: Landform classification of New Mexico by computer. Menlo Park.
Dodge, Y. 2008: The concise encyclopedia of statistics. New York.
Eastman, J. R. 2016: TerrSet. Geospatial Monitoring and Modeling System. Manual. Worcester.
Ellison, A. M. 2010: Repeatability and transparency in ecological research. Ecology 91-9. DOI: https://doi.org/10.1890/09-0032.1
Ferligoj, A. 1989: Razvrščanje v skupine. Teorija in uporaba v družboslovju. Ljubljana.
Gabrovec, M., Hrvatin, M. 1998: Površje. Geografski atlas Slovenije. Ljubljana.
Gallant, A. L., Douglas, D. B., Hoffer, R. M. 2005: Automated mapping of Hammond's landforms. IEEE geoscience and remote sensing letters 2-4. Piscataway.
Gams, I. 1983: Geografske značilnosti Slovenije. Ljubljana.
Gams, I. 1984: Metodologija geografske razčlenitve ozemlja. Geografski vestnik 56.
Gams, I. 1998: Pokrajinsko ekološka sestava Slovenije. Geografija Slovenije. Ljubljana.
Gams, I., Kladnik, D., Orožen Adamič, M. 1995: Naravnogeografske regije Slovenije. Krajevni leksikon Slovenije. Ljubljana.
Haggett, P. 2001: Geography: a global synthesis. Harlow.
Hammond, E. H. 1964: Analysis of properties in landform geography: An application to broadscale landform mapping. Annals of Association of American Geographers 54-1. DOI: https://doi.org/10.1111/j.1467-8306.1964.tb00470.x
Hazeu, G. W., Metzger, M. J., Mücher, C. A., Perez-Soba, M., Renetzeder, C., Andersen, E. 2011: European environmental stratifications and typologies: An overview. Agriculture, ecosystems and environment 142, 1–2. DOI: https://doi.org/10.1016/j.agee.2010.01.009
Hrvatin, M., Perko, D. 2012: Morphological typifications of Slovenia's surface using global classification methods. Geografski vestnik 84-1.
Ilešič, S. 1956: Slovenske pokrajine. Geografski obzornik 3-2.
Ilešič, S. 1958: Problemi geografske rajonizacije ob primeru Slovenije. Geografski vestnik 29-30.
Iwahashi, J., Pike, R. J. 2006: Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86, 3–4. DOI: https://doi.org/10.1016/j.geomorph.2006.09.012
Jiang, B., Ding, X., Ma, L., He, Y., Wang, T., Xie, W. 2008: A hybrid feature selection algorithm: combination of symmetrical uncertainty and genetic algorithms. Optimization and systems biology. Lijiang.

Jongman, R. H. G., Bunce, R. G. H., Metzger, M. J., Mücher, C. A., Howard, D. C., Mateus, V. L. 2006: Objectives and applications of a statistical environmental stratification of Europe. Landscape Ecology 21-3. DOI: https://doi.org/10.1007/s10980-005-6428-0

Kladnik, D. 1996: Naravnogeografske členitve Slovenije. Geografski vestnik 68.

Kladnik, D., Perko, D., Urbanc, M. 2009: Cultural landscapes in Slovenia from geographical perspective. Cultural landscape – Across disciplines. Kraków.

Loveland, T. R., Merchant, J. M. 2004: Ecoregions and ecoregionalization: geographical and ecological perspectives. Environmental Management 34-S1. DOI: https://doi.org/10.1007/s00267-003-5181-x

McGarigal, K., Cushman, S., Stafford, S. 2000: Multivariate statistics for wildlife and ecology research. New York.

McMahon, G., Wiken, E. B., Gauthier, D. A. 2004: Toward a scientifically rigorous basis for developing mapped ecological regions. Environmental Management 34-1. DOI: https://doi.org/10.1007/s00267-004-0170-2

Melik, A. 1935: Slovenija – geografski opis. Ljubljana.

Melik, A. 1946: Prirodnogospodarska sestava Slovenije. Geografski vestnik 18.

Mücher, C. A., Bunce, R. G. H., Jongman, R. H. G., Klijn, J. A., Koomen, A. J. M., Metzger, M. J., Wascher, D. M. 2003: Identification and characterisation of environments and landscapes in Europe. Alterra rapport 832. Wageningen.

Natek, K. 1993: Tipi površja v Sloveniji. Geografski obzornik 40-4.

Natek, K., Žiberna, I. 2004: Naravnogeografske regionalizacije Slovenije. Teorija in praksa regionalizacije Slovenije. Maribor.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnut, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001: Terrestrial ecoregions of the World: a new map of life on earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51-11. DOI: https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Oštir, K. 2006: Daljinsko zaznavanje. Ljubljana.

Pan-European Biological and Landscape diversity strategy, 1996. Internet: https://www.unibuc.ro/prof/patru-stupariu_i_g/docs/2014/noi/03_11_52_38paneurop_strategie.pdf (12. 12. 2018).

Perko, D. 1998: Regionalizacija Slovenije. Geografski zbornik 38.

Perko, D. 2001: Analiza površja Slovenije s stometrskim digitalnim modelom reliefa. Geografija Slovenije 3. Ljubljana.

Perko, D. 2007: Landscapes. Slovenia in focus. Ljubljana.

Perko, D., Hrvatin, M. 2009: Določanje morfoloških enot površja v Sloveniji s prirejeno Hammondovo metodo. Geografski vestnik 81-2.

Perko, D., Hrvatin M., Ciglič, R. 2015: A methodology for natural landscape typification of Slovenia. Acta geographica Slovenica 55-2. DOI: https://doi.org/10.3986/AGS.1938

Plut, D. 1999: Regionalizacija Slovenije po sonaravnih kriterijih. Geografski vestnik 71.

Plut, D. 2005: Teoretična in vsebinska zasnova trajnostno sonaravnega napredka. Dela 23. DOI: https://doi.org/10.4312/dela.23.59-113

Pravilnik o določanju in vodenju bonitete zemljišč. Uradni list Republike Slovenije 47/2008. Ljubljana.

Richards, J., Jia, X. 2006: Remote sensing digital image analysis: an introduction. Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-642-30062-2

Rogerson, P. A. 2006: Statistical methods for geography: a student guide. London, Thousand Oaks, New Delhi.

Romportl, D., Chuman, T. 2012: Proposal method of landscape typology in the Czech Republic. Journal of landscape ecology 5-3. DOI: https://doi.org/10.2478/v10285-012-0057-5

Runhaar, H. J., Udo de Haes, H. A. 1994: The use of site factors as classification characteristiccs for ecotopes. Ecosystem classification for environmental management. Dordrecht. DOI: https://doi.org/10.1007/978-94-017-1384-9_7

Špes, M., Cigale, D., Lampič, B., Natek, K., Plut, D., Smrekar, A. 2002: Študija ranljivosti okolja. Geographica Slovenica 35, 1–2. Ljubljana.

TerrSet Help System. Clark Labs. Worcester, 2015. Internet: https://clarklabssf.wpengine.com/wp-content/uploads/2015/07/TerrSet-Help-System.zip (12. 12. 2018).
Tirelli, T., Pessani, D. 2011: Importance of feature selection in decision-tree and artificial-neural-network ecological applications. *Alburnus alburnus alborella*: A practical example. Ecological Informatics 6-5. DOI: https://doi.org/10.1016/j.ecoinf.2010.11.001
The European Landscape Convention. Internet: https://www.coe.int/en/web/landscape/home (10. 9. 2018).
Theodoridis, S., Koutroumbas, K. 2006: Pattern recognition. San Diego.
Udo de Haes, H. A., Klijn, F. 1994: Environmental policy and ecosystem classification. Ecosystem classification for environmental management. Dordrecht. DOI: https://doi.org/10.1007/978-94-017-1384-9_1
Verbič, T. 1998: Kamnine. Geografski atlas Slovenije. Ljubljana.
Warner, T. A., Campagna, D. J. 2009: Remote sensing with Idrisi® Taiga. A beginner's guide. Hong Kong.
Whittow, J. B. 2000: Dictionary of physical geography. London.
Zupančič, M., Marinček, L., Puncer, I., Seliškar, A. 1998: Potencialno naravna vegetacija. Geografski atlas Slovenije. Ljubljana.