AUTHOR QUERY FORM

ELSEVIER	**Journal:** ECOIND **Article Number:** 1606	**Please e-mail or fax your responses and any corrections to:** **E-mail:** corrections.esch@elsevier.thomsondigital.com **Fax:** +353 6170 9272

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
	Reference(s) given here were noted in the reference list but are missing from the text – please position each reference in the text or delete it from the list.
	The reference given here is cited in the text but is missing from the reference list – please make the list complete or remove the reference from the text: "Saura and Martinez-Millán (2001)".
Q1	Please confirm that given names and surnames have been identified correctly.
Q2	Reference "Saura and Martinez-Millán (2001)" is cited in the text but not provided in the reference list. Please provide it in the reference list or delete these citations from the text.
Q3	Uncited references: This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.
Q4	Please supply the name of the city of publication for reference "Forman and Godron (1986)".
	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

G Model
ARTICLE IN PRESS

Ecological Indicators xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Highlights

Ecological Indicators xxx (2013) xxx–xxx

Factors biasing the correlation structure of patch level landscape metrics

Szilárd Szabó*, Zoltán Túri, Sándor Márton

- We assessed the correlation structure of 13 patch level landscape metrics with PCA.
- We applied several combinations of landscape types, resolutions and variable sets to reveal the influencing factors on correlations.
- Outcomes indicate the relevance of variable sets and smaller importance on cell size and landscape types (including patch size and configuration).
- Control measurements showed the reliability of the results and revealed the high variability of core area metrics.

Ecological Indicators xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Factors biasing the correlation structure of patch level landscape metrics

Q1 Szilárd Szabó[a,*], Zoltán Túri[a], Sándor Márton[b]

[a] Department of Physical Geography and Geoinformation Systems, University of Debrecen, Egyetem tér. 1, 4032 Debrecen, Hungary
[b] Department of Sociology and Social Politics, University of Debrecen, Egyetem tér. 1, 4032 Debrecen, Hungary

ARTICLE INFO

Article history:
Received 17 February 2013
Received in revised form 25 June 2013
Accepted 30 June 2013

Keywords:
Spatial metrics
Correlation structure
Multivariate analysis
PCA
Generalization

ABSTRACT

Landscape metrics are in varying correlations with each other. Several authors have revealed their correlation structure and determined sets of metrics which can be used in landscape analysis. We assumed that correlation structure is not stable and is biased by several factors, thus, selection based on the correlation can vary by case studies. In this study we dealt with 13 patch level landscape metrics using three landscape types, consisting of 9 subregions with 7 and 14 land cover classes, applying 5 different cell sizes. In each step of the analysis other factors that can bias the results were controlled, or the analyses were carried out separately. In accordance with our aims, we uncovered the factor structure of the metrics in different situations, with the parameters which might possibly bias the results. Results showed that cell size, landscape types and number of land cover classes had a greater or lesser effect on cross-correlations. However, the greatest effect was experienced when variables were changed slightly (i.e. two metrics were replaced with two new ones). A comparison of factor structure was conducted with the coefficient of congruence, rank order based on factor loadings, and biplots. According to our findings, congruence values are not reliable in all cases, while ranks and biplots were not sensitive to the changes in circumstances. Possible outcomes were tested with calculations of 3 test areas (a large landscape from NE-Hungary and two countries). Results can be relevant for landscape ecologists dealing with many variables and multivariate techniques.

1. Introduction

Landscape metrics are the quantitative tools of landscape analysis, giving a clear, reproducible methodology to quantify the features of habitat patches and their spatial distribution, with a direct connection to ecological observations and processes (Forman and Godron, 1986; Waltz, 2011). Several landscape indices have been successfully integrated into ecological studies (Kupfer, 2012; Schindler et al., 2013). In general, the simplest metrics, such as patch size, perimeter–area ratio, distance from nearest habitat patches or total number of species, are widely used (e.g. Magura et al., 2001; Szilassi et al., 2010). In the practice of landscape planning, metrics of connectivity and fragmentation are applied (Jaeger et al., 2008; Girvetz et al., 2008; Penn-Bressel, 2005; Stone, 2007).

In addition, we should mention that indices have been criticized for being redundant (i.e. strong correlation), for having map scales which do not match the scale of processes, for a lack of clear recommendations regarding usage and for inconsistent correlation with ecological processes, as well as for producing contradictory results (Cale and Hobbs, 1994; Darmstad, 2009; Haines-Young and Chopping, 1996; Li and Wu, 2004; Tischendorf, 2001). Another criticism is that all analyses will produce numerical results and the ecological functionality for most of the metrics has not been proved (Baldwin et al., 2004; Turner, 2005). Furthermore, pixel size, map scale and map extent also alter the results (Saura and Martinez-Millán, 2001; Wu et al., 2000).

The first software that was able to derive landscape metrics in bulk was FRAGSTATS (McGarigal and Marks, 1995) and this had a significant influence on landscape analysis. Researchers started to deal with landscape indices in hundreds of papers (e.g. Hargis et al., 1998; Kareiva and Wennergren, 1995). The redundancy of the metrics was obvious from the beginning, but the new metrics were easier to interpret, or had some additional meaning, or simply correlated with others in spite of measuring different aspects of the landscape. Instead of preferring one index, several authors recommended revealing the correlation structure of the metrics through factor analysis and chose the relevant non-correlated indexes. McGarigal and McCombs (1995) and Riitters et al. (1995) were the first to determine the statistical relationships between the metrics with multivariate methods. They, and other authors (e.g. Griffith et al., 2000; Cushman et al., 2008; Schindler et al., 2008; Skånes

* Corresponding author. Tel.: +36 52 512900/22326; fax: +36 52 512945.
E-mail addresses: szabo.szilard@science.unideb.hu, szaboszilard.geo@gmail.com (S. Szabó).

1470-160X/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.ecolind.2013.06.030

and Bunce, 1997; Ueemaa et al., 2011), used principal component analysis (PCA) to reduce the number of indices, providing a methodology to choose the most meaningful metrics. A different evaluation was conducted by Baranyai et al. (2011): they used an ordinal clustering algorithm and non-metric multidimensional scaling (NMDS) to reveal the relations between 14 connectivity measures.

Multivariate techniques such as PCA, NMDS or cluster analysis are effective tools to reduce the number of variables, but results are not consistent. As in other areas of the environment where the environmental variables are not constant (Leitao and Ahern, 2002), results are influenced by the scale, dominant patch size, minimum mapping unit, number of land use classes, cell size of the raster coverages, etc. Therefore, only the methodology can be constant, and findings should often be handled as case studies. As we have described, many authors have dealt with the question of correlation or redundancy but there has been no research on correlation dependencies. It was merely supposed the correlations of landscape metrics can change with the input parameters.

In the present work we dealt with the correlation stability of the indices, focusing on their correlation structure. We assumed that both correlation structure, and consequently the principal components as well, changes with the properties of input data. If the changes are not significant, landscape metric selection can be based on correlation techniques; however if this is not case, this kind of selection produces different results that cannot be extrapolated. Our aim was to provide a justification for this assumption; accordingly, we tested the effects of resolution, the number of land cover classes, different sets of variables and map extent. We provided a method to control the changes.

2. Methods

2.1. Study sites

Nine study areas next to each other were selected along the River Tisza. Over the past 20,000 years the river has changed its channel frequently in the Great Hungarian Plain (Marosi and Szilárd, 1969). These changes produced significant shifts in the direction of the riverbed. The river widened its floodplain, eroding the original Pleistocene sand dunes. In the Holocene, three of the selected study areas were floodplains, three areas were sandy islands without inundation, as their surface was higher than the flood level, and three were loess terrains (formed in the former floodplain of the river; Gábris and Túri, 2008). According to the different origins of landscape evolution, the landscape pattern was different, in spite of their close location (Fig. 1).

The boundaries of the study areas were determined using the edges of habitat patches, considering natural or artificial borders (e.g. roads, channels). In this way we were able to avoid the splitting of habitat patches, which can cause skewed results when calculating areas and shape indices.

The study areas had different characteristics and their utilization was exploited taking this into account. As the area is a plain and, following water regulation in the 19th century, the whole area became available for agricultural production (Table 1), the dominant land use type was consequently arable land (generally above 50%). There were only a low percentage of areas of natural vegetation (generally below 10%). In the case of sand dunes #2 residential areas are mainly recreational gardens with small houses in a rural environment, arable lands being the second largest land use type.

We defined our nomenclature in the following way: landscape types (floodplains, loess based terrains and sand dunes), subregions (all landscape types were divided into three parts according to Fig. 1) and the smallest units were the land cover patches (the landscape metrics were calculated using these).

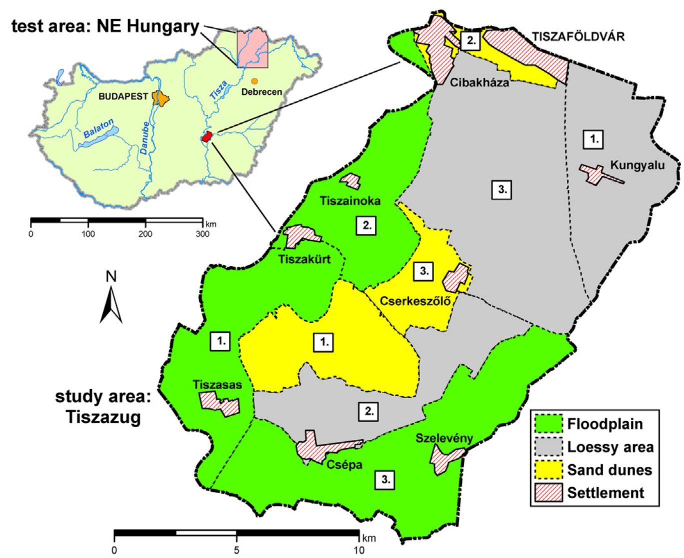

Fig. 1. Location of the study areas and subregions.

2.2. Land use data and landscape metrics

We vectorized all the identifiable habitat patches using digital ortophotos from the year 2005 (0.5 m resolution) in GIS environment (with ArcGIS, ESRI, 2008) applying visual interpretation. The minimum mapping unit was 0.0025 ha. We applied the nomenclature (generally the second level; in some cases – e.g. forests – the third level) of the CLC database in order to use a uniform system and to avoid having too many, and overspecified, land use/land cover (LULC) classes. Altogether there were 14 LULC classes that can be interpreted in the statistical analysis: residential area, industrial area, mine/dump/construction site, artificial green area, arable land, vineyard/orchard, grassland, coniferous forest, deciduous forest, mixed forests, shrub, wetland, water body. We reduced the number of classes, as, given their similarity, these can be aggregated into seven categories: artificial surface (residential and industrial areas, mines), forest (mixed, coniferous, deciduous forests), arable land, orchard, grassland, shrub and water. If we do not differentiate between mixed, coniferous and deciduous forests we can simply use the term 'forest'. In many cases when we have to use historical maps or old aerial photos for large areas, there is no way of distinguishing forest types; we can only recognize that there was a forest there. Shrubby areas and wetlands, and, additionally, agricultural and mixed agricultural areas, cannot be distinguished without knowing the area (and can hardly be recognized in old black and white aerial photos).

We converted our vector overlays to raster format and processed them in FRAGSTATS 3.4 (McGarigal and Marks, 1995). We applied 5, 10, 25, 50 and 100 m cell sizes for raster layers for each study area and calculated landscape indices. 13 patch level metrics were calculated.

According to our aims, we chose patch level metrics: we aimed to identify patches based on their individual spatial characteristics. Identification supposes the existence of the uniqueness of the patches from a given point of view.

Landscape metrics were the following (for a detailed description see McGarigal and Marks, 1995):

- Area and edge metrics: Area (AREA), Perimeter (PERIM);
- Shape related metrics: Perimeter Area ratio (PARA), Radius of Gyration (GYRATE), Shape index (SHAPE), Related Circumscribing Circle (CIRCLE), Contiguity Index (CONTIG), Perimeter-Area Fractal Dimension (PAFRAC);

Please cite this article in press as: Szabó, S., et al., Factors biasing the correlation structure of patch level landscape metrics. Ecol. Indicat. (2013), http://dx.doi.org/10.1016/j.ecolind.2013.06.030

Table 1
Main features of the study sites.

Study sites	Area (ha)	Number of patches	Mean patch size (ha)	Largest patch area (ha)	Dominating land cover type (%)
Floodplain #1	2241	218	10.28	792.28	Plough land (59%)
Floodplain #2	2575	530	4.86	445.28	Plough land (58%)
Floodplain #3	4455	457	9.75	771.77	Plough land (68%)
Sand dunes #1	2138	585	3.65	640.30	Orchard (50%)
Sand dunes #2	970	1006	0.96	167.35	Residential (28%)
Sand dunes #3	1107	726	1.52	140.16	Plough land (42%)
Loess based terrain #1	2838	223	12.72	1013.67	Plough land (77%)
Loess based terrain #2	2542	325	7.82	1000.14	Plough land (79%)
Loess based terrain #3	5518	203	27.18	2041.84	Plough land (89%)

- Core area metrics: Core Area (CORE), Number of Core Areas (NCORE), Core Area Index (CAI);
- Aggregation metrics: Euclidean Nearest-Neighbour (ENN), Proximity index (PROX).

We applied 11 metrics as a set and 2 metrics were used in the analysis to detect the effects of differing variables.

2.3. Data analysis

To reveal the correlation structure we conducted principal component analysis (PCA) with Varimax rotation (in this case with principal components, PCs). PCs do not correlate, but within the PCs the correlation of variables is maximal. Variables were transformed with the formula $\log(k+1)$ due to the different dimensions of the metrics' magnitude and in order to improve normality. This method has been applied in several previous studies (e.g. Leitao and Ahern, 2002; Schindler et al., 2008). Principal components were retained when eigenvalues exceeded 1 according to Kaiser's criteria. We carried out the analysis with the PCA in all variations of landscape types, resolutions and land cover classes (Fig. 2). Communalities were controlled (we excluded low values when this was needed), Kaiser–Meyer–Olkin values were accepted above 0.6, and Bartlett's tests were significant ($p < 0.05$).

Comparison of the structure matrix was carried out with the coefficient of congruence (r_c). According to MacCallum et al. (1999) congruence values were qualified as "excellent" when $r_c > 0.98$, "good" between 0.98 and 0.92, "borderline" between 0.92 and 0.82, "poor" between 0.82 and 0.62 and "terrible" when values stayed below 0.68. Congruence (r_c) was found to be better than the Pearson correlation when correlating factors, since r_c estimated the correlation between the factors themselves, while Pearson r took into account two column vectors of factor loadings (Aluja-Fabregat et al., 2000). For the graphical interpretation of the eigenvalues of PCs, biplot diagrams were applied. Biplots were calculated from 20% of the whole dataset to ensure the visibility of the results. Reducing the data did not influence the diagrams, but made it possible to see the lines of the variables.

Statistical analyses were carried out in SPSS17 (SPSS Inc., 2007) and PAST (Hammer et al., 2001) software. The coefficient of congruence was calculated with Invariance (Watkins, 2005).

2.4. Test for extrapolation

It is important to judge if the results can be extrapolated, i.e. to establish whether our findings of correlation structure can be generalized or are only true in this small area. Accordingly, we processed 3 further areas: a CLC50 map of a 3470 km^2 study area in North-eastern Hungary, the CLC2000 map of Hungary and Portugal. Table 2 showed the main characteristics of the digital layers including our primary test area (Tiszazug). We calculated the same landscape metrics for all layers, then produced correlograms of the variables with R (corrgram package, Wright, 2012). Correlograms indicated the connections with colours (the darker the colour, the greater the correlation), with hashes (right hash: positive, left hash: negative correlation); pie charts showed the magnitude of connections (Kabacoff, 2011). In addition, we extracted the ranges for each variable (landscape metric) pair of the correlation matrices concerning each test area. Ranges were determined and evaluated according to Fig. 3.

3. Results

Analysis of 11 patch level landscape metrics revealed the correlation structure of the dataset. Although correlations of the variables were distinct in varying measures, the structure of the PCs showed similar factor loadings in most cases. Coefficients of

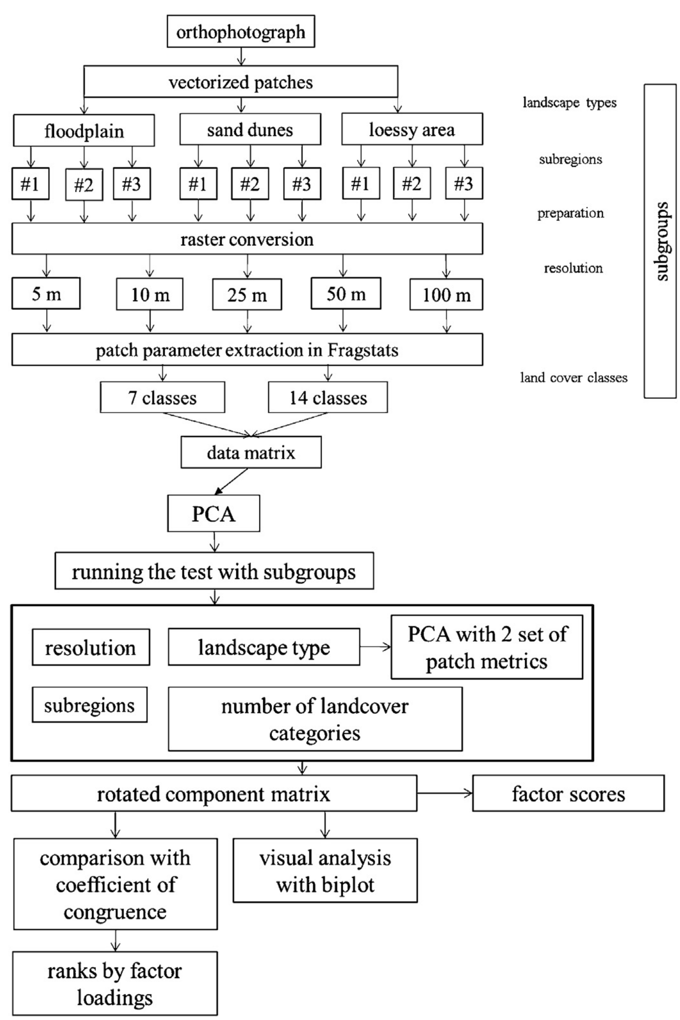

Fig. 2. Schematic outline of the procedures applied in the analysis.

Table 2
Metadata and some selected relevant data of the study areas.

	Tiszazug	NE Hungary	Hungary	Portugal
Data type	Vectorized ortophotos	CLC50	CLC2000	CLC2000
Minimum mapping unit (ha)	0.0025	4	25	25
Cell size (m)	5	10	100	100
Area (km^2)[a]	243	3418	93,027	89,405
Number of patches	4273	4595	39,244	31,473

[a] Calculated from CLC2000.

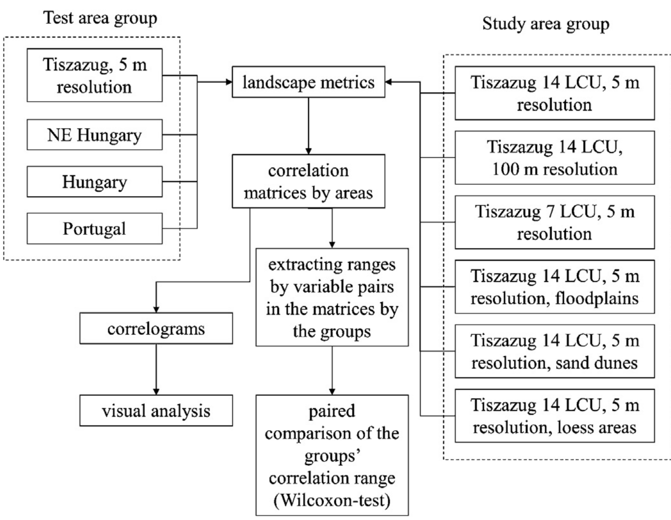

Fig. 3. Test of generalization of the results.

congruence values were mainly above 0.98, showing excellent similarity between component matrices.

3.1. Effect of cell size on correlation structure

Cell size had a lesser effect on the correlation structure than was predicted, considering the changes in the values: resolution caused 20–30% changes in the value of the metrics, as a consequence of the fact that above 25 m cells several patches were merged into one larger patch due to the coarser resolution. Changes followed almost the same trend, especially in the first PCs (Fig. 4).

Overall, relations among the spatial metrics were in a stable structure, moderately altered by the applied cell sizes (Table 3): similarities never decreased below the "borderline" level. Between the 5–10, the 25–50, and 25–100 m categories similarity was "excellent" ($r_c > 0.98$) for each of the three PCs. All the other pairs in the comparisons had smaller r_c values, indicating differences.

Based on the r_c values we found the solutions of the 5 m and 100 m cell size which had one of the largest differences (considering the three PCs together), and analyzed the component matrix by creating ranks. The result (Table 4) differed from the table of congruencies as there was more relevant variation in the rank orders between 5 and 25 m than between 5 and 100 m PCA solutions. Although similarities were almost the same ($r_c > 0.98$) in the case of PC1, the order of the variables differed from the third metric in the rank. Subtracted factor loadings showed small variances, and had increasing tendencies: 0.03-0.09-0.11-0.10 (differences in absolute values between 5-10, 5-25, 5-50 and 5-100 m PC1s, respectively) on average. Both negative and positive differences occurred, and some variables changed their signs (AREA, PROX), showing the effect of cell size on them.

However, elements of the PCs never mixed; thus, although the factor loadings acquired some small changes, the factor structure remained permanent. PC1 and PC2 contained mainly shape metrics, with area and perimeter.

Furthermore, we analyzed the factor structure graphically, using biplots. In the multidimensional space of PC1 and PC2 we can observe the same tendencies of the variables (Figs. 5 and 6).

In case of the biplot of the 5 m PCA solution (Fig. 5) PROX had the largest variance and was in high negative correlation with ENN. PERIM had the second largest variance, and together with all the other metrics, was in strong negative correlation with PARA; in addition, it had no correlation with PROX and ENN. PERIM, GYRATE, AREA and CORE correlated strongly with each other, while SHAPE, CONTIG and CIRCLE made up another group correlating slightly

Table 3
Coefficient of congruence in case of various cell sizes ("excellent" similarities are highlighted in bold).

Cell size (m)	PC1	PC2	PC3
5–10	**0.998**	**0.998**	**0.998**
5–25	**0.985**	0.965	0.971
5–50	0.976	0.929	0.939
5–100	**0.98**	0.93	0.824
10–25	**0.991**	0.977	0.978
10–50	**0.982**	0.946	0.943
10–100	**0.983**	0.941	0.853
25–50	**0.997**	**0.988**	**0.988**
25–100	**0.991**	**0.98**	**0.988**
50–100	**0.982**	**0.984**	0.906

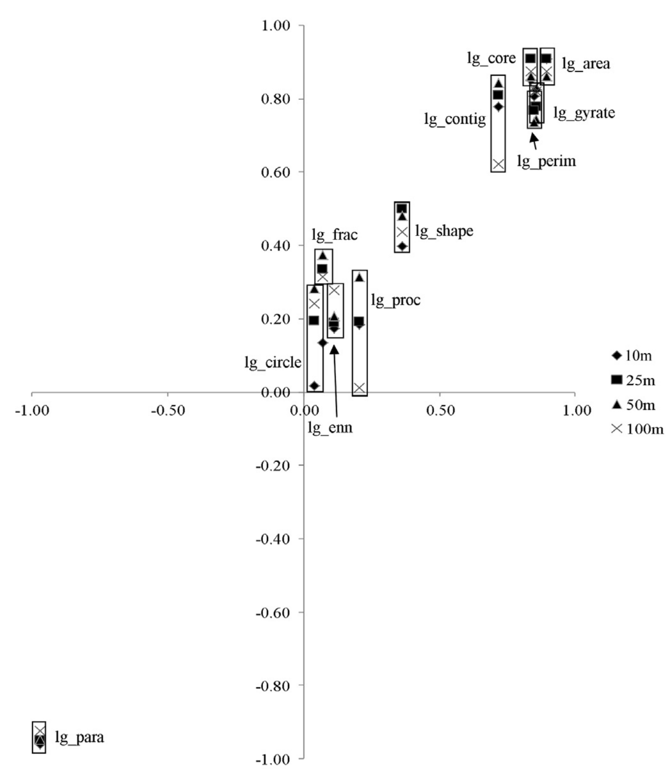

Fig. 4. Diagram of PC1s of the 10-25-50-100 m cells size PCA solutions against the PC1 of 5 m cell size.

Table 4
Rank orders of some selected component matrix of PCA solutions.

PCs	5 m solution	25 m solution	100 m solution
PC1	PARA > AREA > PERIM > GYRATE > CORE > CONTIG	PARA > AREA > CORE > CONTIG > PERIM > GYRATE	PARA > AREA > CORE > GYRATE > PERIM > CONTIG
PC2	FRAC > CIRCLE > SHAPE	CIRCLE > FRAC > SHAPE	FRAC > CIRCLE > SHAPE
PC3	ENN > PROX	ENN > PROX	PROX > ENN

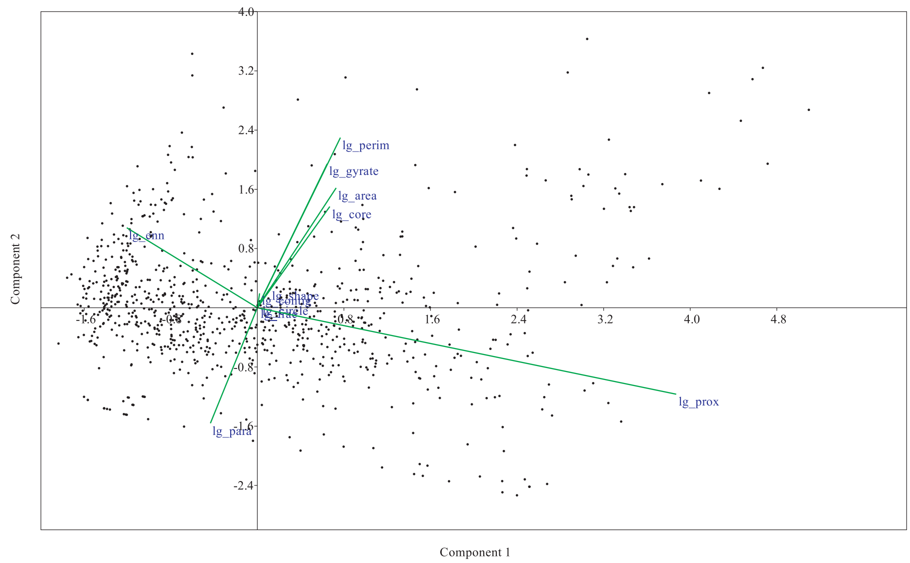

Fig. 5. Biplot of the landscape metrics in case of the dataset containing data of 5 m cell size.

with the group containing PERIM and PROX. PERIM and GYRATE, as well as AREA and CORE, were in strong correlation.

The biplot of the 100 m PCA solution (Fig. 6) showed similarities in general, but had some differences as directions were rotated (without changing the main relationships). In this solution, ENN had the largest variance, while PERIM and CORE together had the second largest variance.

3.2. *Effects of different landscapes on correlation structure*

Besides cell size, landscapes can bias the correlation structure of spatial metrics with their spatial pattern, land cover variability, patch sizes and patch shapes. However, the correlation structure was similar at the "excellent" level; all r_c values were above 0.97 except for the PC3 of sand dunes – loess terrains (which was 0.93).

Rank orders of the variables within the component matrix were identical in each landscape type. Furthermore, ranks were the same as the ranks of the 5 m PCA solution in Table 4. Differences between the PC loading pairs of the landscapes (e.g. $PC1_{floodplain}$ – $PC1_{sand\ dunes}$) were 0.008–0.026.

Biplot diagrams of landscape types showed similar structure without relevant differences compared to the cell sizes. ENN and PROX were in strong negative correlation in all cases; PROX had the largest variance in these cases, too. These metrics did not correlate

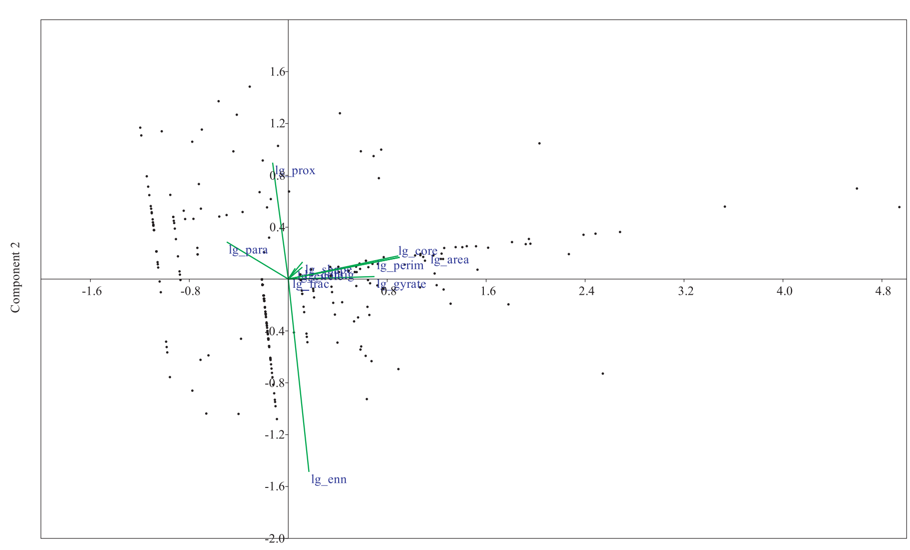

Fig. 6. Biplot of the landscape metrics in case of the dataset containing data of 100 m cell size.

Please cite this article in press as: Szabó, S., et al., Factors biasing the correlation structure of patch level landscape metrics. Ecol. Indicat. (2013), http://dx.doi.org/10.1016/j.ecolind.2013.06.030

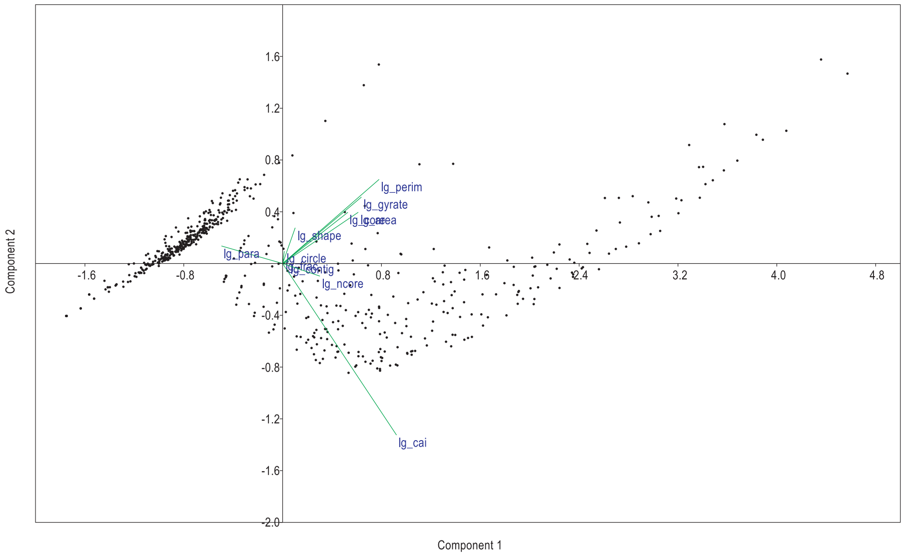

Fig. 7. Biplot of the landscape metrics in case of the dataset containing data of 5 m cell size using different variables (PROX and ENN metrics were changed to CAI and NCORE).

with the others. PARA was in strong negative correlation with the rest of the variables. The directions of the vectors were more or less the same but the variances differed.

We examined the similarities of component matrices inside the landscape types (i.e. subregions, Table 5). The results reflected the importance of the details: the correlation structure of subregions differed more intensively than between the landscape types. Sand dunes, especially, had dissimilar correlation structure. Sand dunes #1 had a larger area and fewer habitat patches, and consequently patch sizes were larger. These characteristics caused the changes in the correlation structure. However, the loess terrain #1 did not differ as much from the others as r_c values indicate in the component matrices.

3.3. Effects of land cover units on the correlation structure

When we applied different sets of LULC classes, there was a relevant decrease in similarities (Table 6). Apart from some “excellently” rated pairs there were only “good” or worse parities. The same landscape types with a different number of classes (e.g. sand dunes$_{14\text{class}}$ – sand dunes$_{7\text{class}}$) had greater differences than those cases when pairs consisted of different landscape types (e.g. sand dunes$_{14\text{class}}$ – loess-based terrain$_{7\text{class}}$). Congruence (r_c) values were mainly rated only as “good” or worse than “good”, and PC1s were somewhat smaller than PC2s, but PC3 similarities were remarkably smaller.

Rank orders were the same as in the case of the 5 m PCA solution (see Table 4).

Table 5
Coefficient of congruence between PCs inside landscape type groups (“excellent” similarities are highlighted in bold).

Subregions	PC1	PC2	PC3
Floodplain #1–floodplain #2	**0.991**	**0.99**	**0.982**
Floodplain #1–floodplain #3	**0.989**	**0.984**	**0.986**
Floodplain #2–floodplain #3	**0.989**	**0.999**	**0.992**
Sand dunes #1–sand dunes #2	**0.986**	0.351	0.339
Sand dunes #1–sand dunes #3	**0.991**	0.383	0.367
Sand dunes #2–sand dunes #3	**0.996**	**0.998**	**0.995**
Loess based terrain #1–loess based terrain #2	0.808	0.976	0.66
Loess based terrain #1–loess based terrain #3	0.794	0.976	0.706
Loess based terrain #2–loess based terrain #3	**0.987**	**0.994**	**0.989**

The biplot diagram of 7 LULC classes showed a similar structure for the variables as in previous PCA solutions (e.g. Fig. 5).

3.4. The effect of different sets of variables

We tested what would happen when the applied spatial metrics differed slightly: we omitted PROX and ENN (in previous PCA solutions PC3) and used NCORE and CAI. This option was run on landscape types. We found the largest effect on component matrices, taking into consideration all the previous tests. There was a relevant difference in the correlation structure of floodplains compared to sand dunes and loess terrain areas. Congruence (r_c) values were only “good” at PC1s, while in case of PC2s r_c they were “poor”, and “terrible” at PC3s. PCA solutions of sand dunes and loess terrain areas were similar at the “excellent” level.

Congruence values indicated differences, but only the ranks revealed the structural changes. PCs contained distinct metrics contrary to what was experienced in previous investigations. For each landscape type factor loadings had different values and ranks had different orders (Table 7). Biplots also showed a new structure (Fig. 7).

3.5. Possibilities of extrapolation

In order to obtain information about the universality of our results we conducted correlation analyses in the test areas (NE-Hungary, Hungary, Portugal). Cross-correlations showed a varied picture of the connections among the variables (Figs. 8 and 9): some metrics correlated strongly with some others in each case: pairs of AREA–CORE and PARA–CONTIG were completely redundant, while AREA and PERIM, SHAPE and FRAC, and, SHAPE and GYRATE had strong correlations with small changes. The magnitudes of the relationships differed in the case of core area metrics (CORE, NCORE, CAI); differences varied on a wide scale (changes ranged from 0 to 0.5 in the Pearson r value). Furthermore, only in the case of CAI and SHAPE did we identify the turn of the direction in the connection, i.e. we may be able to observe negative and positive correlation between these metrics, while on the contrary, correlations of other metrics never changed their signs. PROX and ENN did not correlate with the other metrics; consequently, they can be regarded as the ones providing unique information. Exploring the differences, we identified that the largest ones belonged to those variable pairs whose range was close to zero

Table 6
Coefficient of congruence between PCs inside different land cover classes ("excellent" similarities are highlighted in bold).

Landscape types by number of LULC classes	PC1	PC2	PC3
floodplain$_{14class}$ – floodplain$_{7class}$	0.899	0.82	0.366
sand dunes$_{14class}$ – sand dunes$_{7class}$	0.834	**0.993**	−0.045
loess based terrain $_{14class}$ – loess based terrain$_{7class}$	0.896	**0.988**	−0.207
floodplain$_{14class}$ – sand dunes$_{7class}$	0.889	0.812	0.201
floodplain$_{14class}$ – loess based terrain$_{7class}$	0.915	0.777	0.415
sand dunes$_{14class}$ – loess based terrain$_{7class}$	0.869	0.988	−0.15

Table 7
Rank orders of some selected component matrix of PCA solutions (landscape metrics are highlighted in italics where factor loadings had similar values in the component matrix).

PCs	Floodplain	Sand dunes	Loess area
PC1	CORE > AREA > CAI > GYRATE > PERIM > NCORE	CORE > AREA CAI > NCORE	CORE > AREA > CAI > PARA > GYRATE > PERIM > NCORE
PC2	FRAC > CIRCLE > SHAPE	FRAC > CIRCLE > SHAPE	FRAC > SHAPE > CIRCLE
PC3	CONTIG > PARA	CONTIG > PARA > GYRATE > PERIM	CONTIG

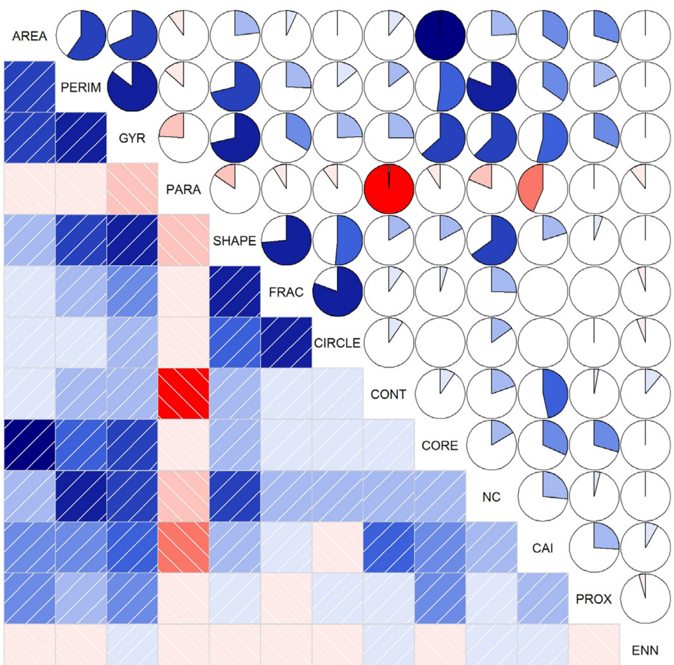

Fig. 8. Correlogram of the landscape metrics of the Tiszazug study area (14 categories, 5 m resolution).

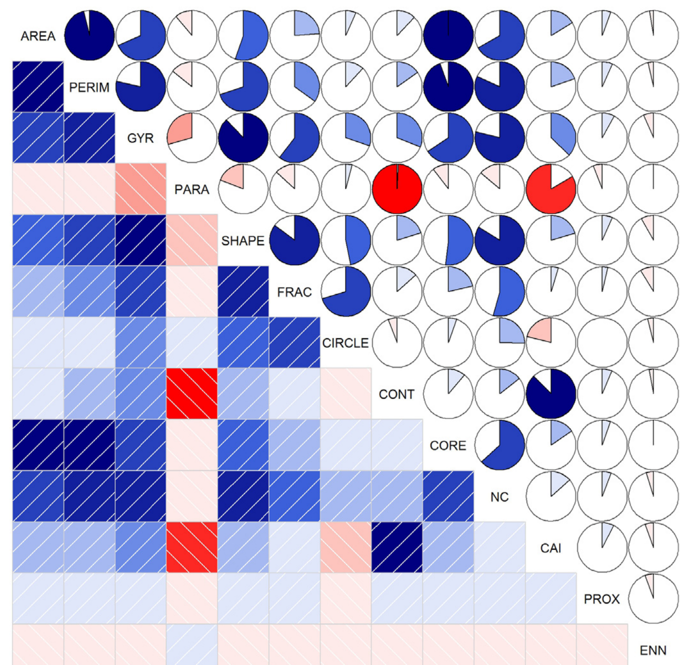

Fig. 9. Correlogram of the landscape metrics of Portugal (14 categories, 100 m resolution).

in the case of the test group; consequently, here, correlations were almost the same (Table 8). These structures were similar to those we calculated in the analysis of the Tiszazug test area, and provided further information about the variability of the metrics.

The Wilcoxon paired test (between test area and study area group, see Fig. 3) revealed that there was no significant difference between the ranges of the correlations ($W = 1785$, $z = 1.665$, $p = 0.096$); therefore, our calculations in that small study area can be regarded as general outcomes.

Table 8
Ranges of Pearson correlation coefficients extracted from 4 correlation matrices (calculated from control dataset).

	AREA	PERIM	GYR	PARA	SHAPE	FRAC	CIRCLE	CONT	CORE	NC	CAI	PROX
PERIM	0.35											
GYR	0.02	0.08										
PARA	0.07	0.09	0.11									
SHAPE	0.12	0.19	0.05	0.05								
FRAC	0.03	0.07	0.03	0.02	0.07							
CIRCLE	0.00	0.08	0.03	0.02	0.01	0.03						
CONT	0.08	0.09	0.12	0.00	0.06	0.03	0.02					
CORE	0.00	0.41	0.02	0.06	0.16	0.05	0.00	0.07				
NC	0.46	0.04	0.01	0.10	0.09	0.02	0.01	0.10	0.51			
CAI	0.21	0.20	0.27	0.02	0.24	0.28	0.34	0.03	0.20	0.18		
PROX	0.27	0.15	0.28	0.08	0.05	0.01	0.04	0.09	0.27	0.03	0.25	
ENN	0.00	0.02	0.08	0.03	0.08	0.07	0.06	0.03	0.00	0.04	0.07	0.00

Please cite this article in press as: Szabó, S., et al., Factors biasing the correlation structure of patch level landscape metrics. Ecol. Indicat. (2013), http://dx.doi.org/10.1016/j.ecolind.2013.06.030

4. Discussion

4.1. Issues of geometric resolution

Pixel size relevantly influences pattern metrics (Saura and Martinez-Millán, 2001; Szabó et al., 2012; Wickham and Riitters, 1995). However, resolution did not have as great an effect on the structure as might be expected (see Wu et al., 2002). We found that the Tiszazug study site, for example (with a 243 km^2 area and a 5 m cell size) had a very similar correlation structure to the site in Portugal (with an area of almost 90,000 km^2 and a 100 m cell size, see Figs. 8 and 9). A 5 m cell size was ideal for analysing all the examined landscape metrics; however, 50 and 100 m cell sizes were only the "skeletonised" variants of the original ones. Small patches were eliminated or merged into larger ones and the whole pattern changed (Saura, 2004); nevertheless, the correlation structure showed only small alterations. Besides, we have to consider the computational limits deriving from the scale and cell size. Analysis of large areas can be carried out only with small scale, i.e. coarse pixels size and, conversely, small areas (large scale) can be investigated with high resolution (O'Neill et al., 1996; Wu et al., 2000). However, we experienced that our upper limit of computation was in high accordance with the number of patches (it was about 40,000 patches).

4.2. Issues of thematic resolution

There were several previous studies on the thematic resolution (Baldwin et al., 2004; Buyantuyev and Wu, 2007; Szabó et al., 2012; Turner et al., 1989b) and it was found that many indices were influenced by the number of land cover types. These studies dealt with class and landscape level metrics; however, we explored significant effect on patch level, too. Different land cover classes caused relevant differences in the factor structure. Application of fewer classes involves the merging of given patches, but it is not identical to the changes caused by increasing cell sizes. Due to the merging classes it is not only small parts that are incorporated into larger ones; even large patches can be plotted as one. Landscape patterns can form in completely different ways with a different number of land cover classes, or it may be the case that the changes are not relevant, depending on the composition. In our study, changes were significant, as was reflected in low r_c values (varying according to PCs). When one works with a certain type of data, its thematic resolution is given and possibly all LULC classes are preserved. Consequently, all investigations use a different number of classes, thus according to our results, findings cannot be compared.

4.3. Map extent: influence of area on the cross-correlations

Map extent also can bias the results. This means that both the area and the borders of the examined units are influencing factors. On the one hand, area determines the possible number of patches (but this also depends on scale, cell size and minimum mapping unit), edge length, and core area, it is thus probable that their value will increase in the case of larger areas (Baldwin et al., 2004). However, it was proved that their standardized formulae were sensitive, too (Baldwin et al., 2004; Saura and Martinez-Millán, 2001). We were dealing with patch level metrics so the extent only biased the number of observed patches and their characteristics, the above mentioned effects are true when we summarize them (e.g. count, calculate simple or area weighted average) on class or landscape level. On the other hand, borders can relevantly skew the calculation of shape metrics by cutting away the outer parts. Turner et al. (1989a) remarked that if the system borders are correct, the experimental model can predict dynamic processes. In our case the question is whether we can be sure that this line runs on the right place. It calls into question the problem of multiscale input data (i.e. we have a large scale land cover map and the coverage of official borders is only small scaled).

Our study areas covered different landscapes from 22 km^2 to 100,000 km^2. Between landscape types there were smaller differences; component structure, however, was more distinct among the subregions. This result indicates that the common origin of subregions was not an overriding factor in determining their correlation structure. Regarding landscape types, subregions of sand dunes differed from each other more than they did from a floodplain or loess terrain area. Changes in correlation structure were reflected in a multivariate way. Remarkably, that we did not find relevant differences among the correlation structure even in the case of countrywide investigations (Hungary, Portugal).

4.4. Correlation structure and the problems of comparisons

We saw that PCA was able to reveal similarities in the correlation structure; however, these were only occasional. All outcomes depend on the specific characteristics of the variables. Identical variables can facilitate the sphericity of the n dimension space or in other cases, cause its deformation and lower the KMO values. The component matrix consists of the factors (principal components) and the variables. If one changes the variables, results in a new solution, causing changes in the component matrix. The final ranks of factor loadings depend on the number of factors, the number of variables, the communality of the variables and their correlations (Jolliffe, 2002). According to the outcomes, factor loadings had minor differences within the given PCs among the different PCA solutions, while there were large differences between the PCs. Thus, patch level metrics showed stable membership in the PCs. PC1 and PC2 were comprised of area-edge and shape metrics; furthermore, PC3 consisted of aggregation metrics. Perimeter is an element of the formula of PARA, FRAC and SHAPE since it is an input parameter in their formulae. Area is an input parameter of PARA, CORE and CIRCLE. Consequently, their common appearance in the first two PCs was not surprising.

If we use the factor scores as artificial variables (e.g. Schindler et al., 2008; Tinker et al., 1998), we can use r_c values to estimate similarities. However, if we use the component matrix to choose the most relevant variables from a set of metrics, considering that PCs provide uncorrelated groups of variables and the ranking of the variables is based on the factor loadings, selected variables can be misleading. If a given rank of metrics was derived from the factor (component) loadings, and differences are small, we can easily find that a metric is not the most relevant one. Therefore, it is advisable to choose the metrics which can be justified in the given analysis. This is in accordance with the findings of Uueemaa et al. (2011) and Leitao and Ahern (2002).

Regarding the comparisons, in spite of the fact that statistical tests provide differing results with different input data, our findings show that the structures, at least at the level of PCs (i.e. groups) were identical in every combination. Different cell size, landscape and LULC numbers did not bias the outputs more than the difference in variables. Factor structure was significantly transformed when we changed 2 spatial metrics in the set. The coefficient of congruency was sensitive to the changes in factor loadings, while biplots and correlograms were able to show the variables in the multivariate space and were not biased by the applied parameters. All diagrams showed a similar picture; groups of metrics were in high accordance with the factor loadings when we used the same variables. However, one has to keep in mind that the similarity of correlation structure does not mean the similarity of the compared landscapes (see the example of the Tiszazug and Portugal). This only means that the investigated variables are not influenced by the input parameters.

Our experience in the testing phase of the generalization possibilities showed the efficiency of correlation matrices. Correlations are calculated pairwise and are not influenced by the number of variables, i.e. correlation between two variables does not change when we investigate more or fewer pairs at the same time. Therefore, we can apply different sets of landscape metrics. Both biplots and correlograms visualize the structures, and the coefficients can be evaluated statistically. Ranges, i.e. the difference between the maximum and minimum correlations coefficients of the variable pairs, showed clearly those pairs where the influencing factors were ineffective. Table 8 reflected that it was metrics with absolute values which experienced especially larger changes (AREA, PERIM, NP), although some standardized ones also had high variance in accordance with Baldwin et al. (2004).

5. Conclusions

Multivariate techniques are useful tools in environmental sciences; they can make it easier to interpret large datasets with many variables. The application of PCA as a popular multivariate method is not new, but this study attempted to reveal the biasing factors of the correlation structure of landscape metrics. It is important to ask what the limits of the researchers' findings are: are they limited to the given investigation or can they be extrapolated? Our results showed that some factors (cell size, landscape type,) do not influence the correlation structure on a significant scale (according to the r_c values), but if we use different number of LULC classes or sets of metrics, the outcomes show large differences.

As a part of data mining and interpretation, comparisons can be carried out with the evaluation of r_c (coefficient of congruence), using the ranks of the component matrix, or graphically with biplots or correlograms. Generally, r_c can hide the real differences, and it may mislead us in our judgement of the distinction between PCA solutions. Factor loadings provide ranks which can be compared with other ranks. Biplots show the variables with their directions and variance and are insensitive to the factors biasing the variables' relationships. Besides this, our suggestion is to apply the evaluation of the correlation matrices by extracting the ranges of correlation coefficients by variable pairs (see Fig. 3).

Q3 Uncited references

Chust et al. (2004) and Heegaard et al. (2007).

Acknowledgements

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and TÁMOP-4.2.2/B-10/1-2010-0024. The project is co-financed by the European Union and European Social Fund.

References

Aluja-Fabregat, A., Colom, R., Abad, F., Juan-Espinoza, M., 2000. Sex differences in general intelligence defined as g among young adolescents. Pers. Indiv. Differ. 28, 813–820.

Baldwin, D.J.B., Weaver, K., Schnekenburger, F., Perera, A.H., 2004. Sensitivity of landscape pattern indices to input data characteristics on real landscapes: implications for their use in natural disturbance emulation. Landsc. Ecol. 19, 255–271.

Baranyai, G., Saura, S., Podani, J., Jordán, F., 2011. Contribution of habitat patches to network connectivity. Redundancy and uniqueness of topological indices. Ecol. Indic. 11, 1301–1310.

Buyantuyev, A., Wu, J., 2007. Effects of thematic resolution on landscape pattern analysis. Landsc. Ecol. 22, 7–13.

Cale, P.G., Hobbs, R.J., 1994. Landscape heterogeneity indices: problems of scale and applicability, with particular reference to animal habitat description. Pacific Conserv. Biol. 1, 183–193.

Chust, G., Ducrot, D., Pretus, J.L., 2004. Land cover mapping with patch-derived landscape indices. Landsc. Urban Plan. 69, 437–449.

Cushman, S.A., McGarigal, K., Neel, M.C., 2008. Parsimony in landscape metrics: strength, universality and consistency. Ecol. Indic. 8, 691–703.

Darmstad, W.E., 2009. Spatial metrics – useful indicators for society or mainly fun tools for landscape ecologists. Nor. Geogr. Tidsskr 63, 246–254.

ESRI, 2008. ArcGIS – A Complete Integrated System. Environmental Systems Research Institute, Inc., Redlands, CA www.esri.com/software/arcgis (accessed September 2012).

Forman, R.T.T., Godron, M., 1986. Landscape Ecology. John Wiley and Sons.

Gábris, Gy., Túri, Z., 2008. Sand movement in historical times in the area of the Tiszazug. Hung. Geogr. Bull. 132, 241–250 (in Hungarian).

Girvetz, E.H., Thorne, J.H., Berry, A.M., Jaeger, J., 2008. Integration of landscape fragmentation analysis into regional planning: a statewide multi-scale case study from California, USA. Landsc. Urban Plan. 86, 205–218.

Griffith, J.A., Martinko, E.A., Price, K.P., 2000. Landscape structure analysis of Kansas at three scales. Landsc. Urban Plan. 52, 45–61.

Haines-Young, R., Chopping, M., 1996. Quantifying landscape structure: a review of landscape indices and their application to forested landscapes. Prog. Phys. Geogr. 20, 418–445.

Hammer, Ř., Harper, D.A.T., Ryan, P.D., 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4, 1–9, http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Hargis, C.D., Bissonette, J.A., David, J.L., 1998. The behavior of landscape metrics commonly used in the study of habitat fragmentation. Landsc. Ecol. 13, 167–186.

Heegaard, E., Økland, R.H., Brati, H., Darmstad, W.E., Engan, G., Pedersen, O., Solstad, H., 2007. Regularity of species richness relationships to patch size and shape. Ecography 30, 589–597.

Jaeger, J., Bertiller, R., Schwick, C., Müller, K., Steinmeier, C., Ewald, C.K., Ghazoul, J., 2008. Implementing landscape fragmentation as an indicator in the Swiss monitoring system of sustainable development (MONET). J. Environ. Manage. 88, 737–751.

Jolliffe, I.T., 2002. Principal Component Analysis, 2nd edition. New York, Berlin, Heidelberg, Springer.

Kabacoff, R.I., 2011. R in Action. Manning, Shelter Island.

Kareiva, P., Wennergren, U., 1995. Connecting landscape pattern to ecosystem and population processes. Nature 373, 299–302.

Kupfer, A.J., 2012. Landscape ecology and biogeography: rethinking landscape metrics in a post-FRAGSTATS landscape. Prog. Phys. Geogr. 36, 400–420.

Leitao, A.B., Ahern, J., 2002. Applying landscape ecological concepts and metrics in sustainable landscape planning. Landsc. Urban Plan. 59, 65–93.

Li, H., Wu, J., 2004. Use and misuse of landscape indices. Landsc. Ecol. 19, 389–399.

MacCallum, R.C., Widaman, K.F., Zhang, S., Hong, S., 1999. Sample size in factor analysis. Psychol. Methods 4, 84–99.

Magura, T., Ködöböcz, V., Tóthmérész, B., 2001. Effects of habitat fragmentation on carabids in forest patches. J. Biogeogr. 28, 129–138.

Marosi, S., Szilárd, J. (Eds.), 1969. The Great Hungarian Plain of the Tisza. Landscape Geography of Hungary, vol. 2. Academic Press, Budapest, pp. 94–97 (in Hungarian).

McGarigal, K., Marks, B.J., 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR, 122 p.

McGarigal, K., McCombs, W.C., 1995. Relationships between landscape structure and breeding birds in the Oregon coast range. Ecol. Monogr. 65, 235–260.

O'Neill, R.V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Ritters, K.H., Wickham, J.D., 1996. Scale problems in reporting landscape pattern at regional scale. Landsc. Ecol. 11, 169–180.

Penn-Bressel, G., 2005. Begrenzung der Landschaftszerschneidung bei der Planung von Verkehrswegen. GAIA 14, 130–134.

Riitters, K.H., O'Neill, R.V., Hunsaker, C.V., Wickham, J.D., Yankee, D.H., Timmins, S.P., Jones, K.B., Jackson, B.L., 1995. A factor analysis of landscape pattern and structure matrics. Landsc. Ecol. 10, 23–40.

Saura, S., 2004. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. Landsc. Ecol. 19, 197–209.

Schindler, S., Poirazidis, K., Wrbka, T., 2008. Towards a core set of landscape metrics for biodiversity assessments: a case study from Dadia National Park, Greece. Ecol. Indic. 8, 502–514.

Schindler, S., Wehrden, H., Poirazidis, K., Wrbka, T., Kati, V., 2013. Multiscale performance of landscape metrics as indicators of species richness of plants, insects and vertebrates. Ecol. Indic. 31, 41–48.

Skånes, H.M., Bunce, R.G.H., 1997. Directions of landscape change (1741–1993) in Virestad, Sweden – characterised by multivariate analysis. Landsc. Urban Plan. 38, 61–75.

SPSS Inc., 2007. SPSS Statistics Base 17.0 for Windows User's Guide. SPSS Inc., Chicago, IL.

Stone, D., 2007. Big plans and little plans: delivering land use change designed by landscape ecology. Acta Geogr. Debrecina Landsc. Environ. 1, 68–74.

Szabó, Sz., Csorba, P., Szilassi, P., 2012. Tools for landscape ecological planning – scale, and aggregation sensitivity of the contagion type landscape metric indices. Carpath. J. Earth Environ. Sci. 7, 127–136.

Szilassi, P., Jordán, G., Kovács, F., Van Rompaey, A., Van Dessel, W., 2010. Investigating the link between soil quality and agricultural land use change. A case study in the Lake Balaton catchment, Hungary. Carpath. J. Earth Environ. Sci. 5, 61–70.

Tinker, D.B., Resor, C.A.C., Beauvais, G.P., Kipfmueller, K.F., Fernandes, C.I., Baker, W.L., 1998. Watershed analysis of forest fragmentation by clearcuts and roads in a Wyoming forest. Landsc. Ecol. 13, 149–165.

Please cite this article in press as: Szabó, S., et al., Factors biasing the correlation structure of patch level landscape metrics. Ecol. Indicat. (2013), http://dx.doi.org/10.1016/j.ecolind.2013.06.030

Tischendorf, L., 2001. Can landscape indices predict ecological processes consistently? Landsc. Ecol. 16, 235–254.

Turner, M.G., 2005. Landscape ecology: what is the state of the science? Annu. Rev. Ecol. Evol. Syst. 36, 319–344.

Turner, M.G., Dale, V.H., Gardner, R.H., 1989a. Predicting across scales: theory development and testing. Landsc. Ecol. 3, 245–252.

Turner, M.G., O'Neill, R.V., Gardner, R.H., Milne, B.T., 1989b. Effects of changing spatial scale on the analysis of landscape pattern. Landsc. Ecol. 3, 153–162.

Ueemaa, E., Roosaare, J., Oja, T., Mander, Ü., 2011. Analysing the spatial structure of the Estonian landscapes: which landscape metrics are the most suitable for comparing different landscapes? Est. J. Ecol. 60, 70–80.

Waltz, U., 2011. Landscape structure, landscape metrics and biodiversity. Living Rev. Landsc. Res. 5, 3 http://www.livingreviews.org/lrlr-2011-3

Watkins, M.W., 2005. Invariance (Computer Software). Ed and Psych Associates, State College, PA.

Wickham, J.D., Riitters, K.H., 1995. Sensitivity of landscape metrics to pixel size. Int. J. Remote Sens. 16, 3585–3594.

Wright, K., 2012. Corrgram package. CRAN Repository, http://cran.r-project.org/web/packages/corrgram/corrgram.pdf www.esri.com/software/arcgis (accessed June 2013).

Wu, J., Jelinski, D.E., Luck, M., Tueller, P.T., 2000. Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. Geogr. Inf. Sci. 6, 6–19.

Wu, J., Shen, W., Sun, W., Tueller, P.T., 2002. Empirical patterns of the effects of changing scale on landscape metrics. Landsc. Ecol. 17, 761–782.