# Describing multi-layer spatial patterns using an integrated co-occurrence matrix (INCOMA)

Jakub Nowosad[*1] and Tomasz F. Stepinski[†2]

[1]Institute of Geoecology and Geoinformation, Adam Mickiewicz University, Poznan, Poland
[2]Space Informatics Lab, Department of Geography and GIS, University of Cincinnati, Cincinnati, OH, USA

March 29, 2021

**Summary**

Maps of landscape types (LTs) are useful abstractions that facilitate land resources management. However, their creation is difficult as local landscapes arise from a fusion of patterns of many natural themes. This paper introduces the integrated co-occurrence matrix (INCOMA) – a signature for numerical representation of multi-thematic categorical patterns. The region tessellated into local landscapes can be represented by INCOMA, which allows for the identification of LTs using standard clustering techniques. The concept of INCOMA is described, and its application is demonstrated by an unsupervised mapping of LTs in Europe based on combined patterns of land cover, landforms, and soils.

**KEYWORDS:** *spatial patterns, pattern similarity, regionalization, thematic maps, global categorical datasets*

## 1. Introduction

Landscape types (LTs) are groups of areas with recognizable, although often heterogeneous patterns of natural themes (Mücher et al. 2010). Identification and classification of LTs is a foundation for planning and management for sustainable use of land resources, and they also provide first-order information about the geographical distribution of biodiversity and ecological processes.

LTs can be delineated by aggregating areal units with similar spatial patterns, although, identifying and mapping LTs is challenging. First, multiple themes (layers), including topography, land cover, climate, and soil/geology, all contribute to a LT's character, and decisions how to represent their patterns and combine them together need to be made. Second, landscape, as represented by a categorical raster, is a pattern formed by many differently-labeled cells (Omernik and Griffith 2014), rather than an aggregate of cells with the same labels. LTs delineated as aggregates of same-label cells (see, for example Sayre et al. (2014)) represent homogeneous landscape elements rather than heterogeneous landscapes.

Using pattern-based methods, a study area is tessellated into relatively small square blocks of cells - areal units called local landscapes (LLs), and LLs with similar patterns are grouped together into mutually exclusive and exhaustive clusters (Wickham and Norton 1994). Such methods have two core ingredients: (a) pattern signature that provides a numerical embedding of a categorical pattern, and (b) a dissimilarity function that quantifies a degree of dissimilarity between two patterns/signatures.

It is important to note that the existing literature on pattern-based delineation of LTs is restricted to using only a single theme – land cover. The purpose of this paper is to introduce the integrated co-occurrence matrix (INCOMA) – a signature for numerical representation of multi-thematic categorical patterns. INCOMA tabularizes intra-thematic as well as inter-thematic adjacencies, and

[*] nowosad.jakub@gmail.com
[†] stepintz@uc.edu

thus contains information about composition and configuration of patterns representing individual themes as well as information about relative positions of patterns associated with different themes. To demonstrate the ability of INCOMA to identify and map LTs, we apply it to the continent of Europe using three variables – land cover (LC), landforms (LF), and soils (SO).

## 2. Signatures

### 2.1. Co-occurrence matrix

A spatial co-occurrence matrix is a signature used to represent spatial patterns of single layer datasets (Haralick, Shanmugam, and Dinstein 1973). It counts all of the pairs of the adjacent cells in categorical raster data. The result is a  by  square matrix, where  is a number of classes in a raster. This signature can be converted into a normalized co-occurrence histogram, where the sum of all values of the elements equals one. Thus, calculating a dissimilarity between a pair of LLs, needed to cluster them into LT, is tantamount to calculating dissimilarity between two histograms.

### 2.2. Integrated co-occurrence matrix

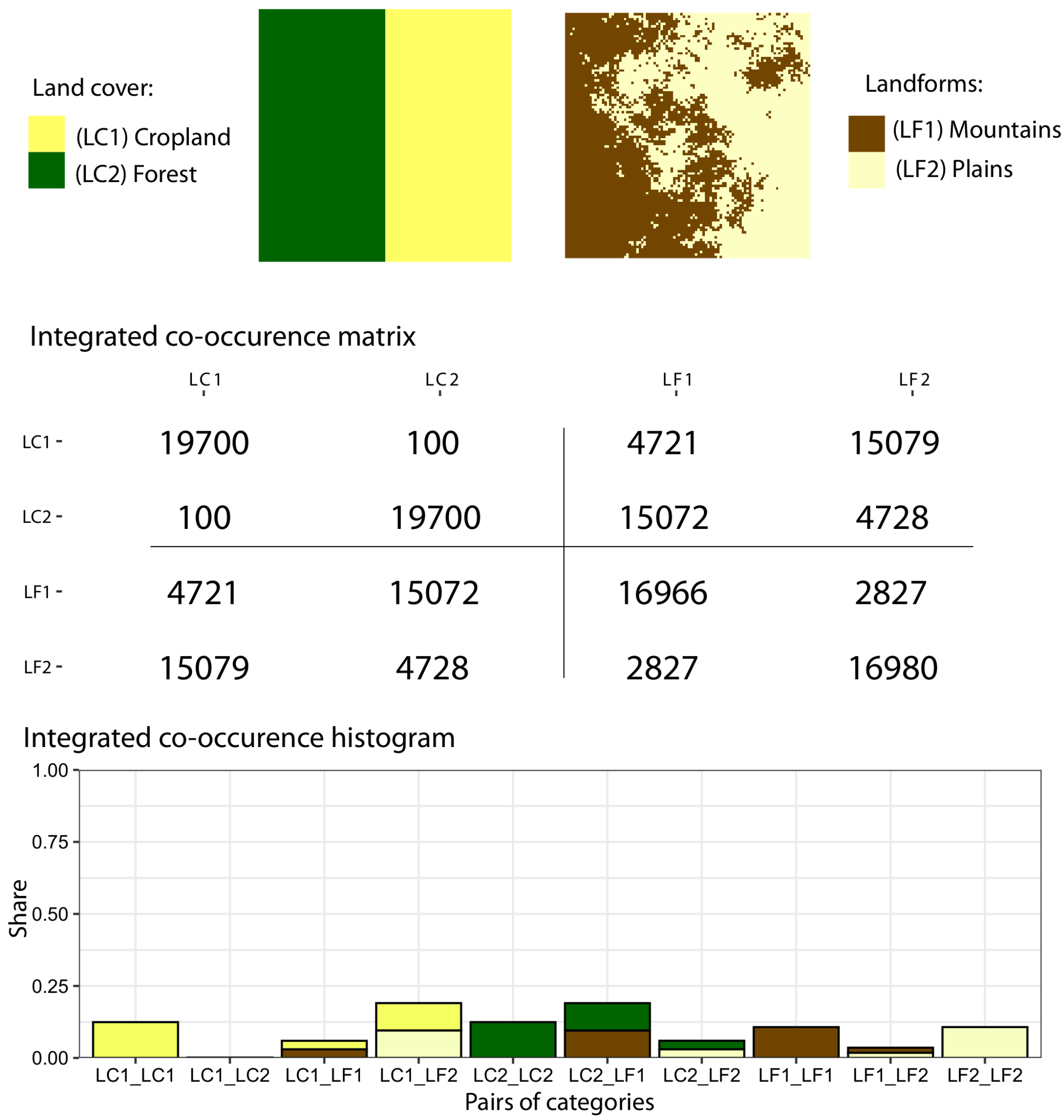


***Figure 1***

*Example of a set of categorical rasters and their spatial signatures, integrated co-occurrence matrix and normalized integrated co-occurrence histogram.*

INCOMA, based on an earlier approach for representing the color and intensity of pixel neighborhoods for images (Vadivel, Sural, and Majumdar 2007), is an extension of the co-occurrence matrix to two or more themes/layers. INCOMA representing two themes has four parts (Figure 1). The upper left part is a co-occurrence matrix for the first (land cover) raster. The bottom right part is a co-occurrence matrix for the second (landforms) raster. The remaining two parts count co-occurrences between classes of land cover and classes of landforms (inter-thematic co-occurrence). INCOMA contains not only information about patterns of all themes but also information about relative positions of different themes' patterns.

INCOMA can also be converted into a one-dimensional form - a normalized integrated co-occurrence histogram (Figure 1).

## 3. Dissimilarity function

Comparison of two LLs represented by signatures is straightforward using distance measures, such as the Jenson-Shannon divergence:

$$JSD(A,B)=H\left(\frac{A+B}{2}\right)-\frac{1}{2}[H(A)+H(B)], \quad (1)$$

where $A$ is the normalized integrated co-occurrence histogram for one LL, $B$ is the integrated co-occurrence histogram for the second LL, and $H(A)$ and $H(B)$ are the values of the Shannon entropy for each histogram.

## 4. Software

We developed two open-source R packages: comat (https://github.com/Nowosad/comat) to calculate INCOMA and motif (https://github.com/Nowosad/motif) to perform pattern-based spatial analysis.

## 5. Identifying and mapping landscape types in Europe

To delineate LTs in Europe, we used three themes/datasets: 2018 C3S land cover (ECMWF 2019), the Hammond's landform regions (Karagulle et al. 2017), and USDA soil taxonomy dataset (Hengl et al. 2017). The datasets were reprojected and resampled to the same grid for Europe with 300 meters resolution. Additionally, they were simplified, resulting in nine land cover categories, four landform classes, and 12 soil categories.

Next, the study area was divided into ~40,000 15 km x 15 km (50 x 50 cells) square blocks (LLs). Spatial patterns of the three themes were used to calculate INCOMA signature for each LLs. Identification of LTs was achieved via clustering LLs using the $K$-means clustering with $K=20$ and the $JSD$ as a distance between signatures (Figure 2).

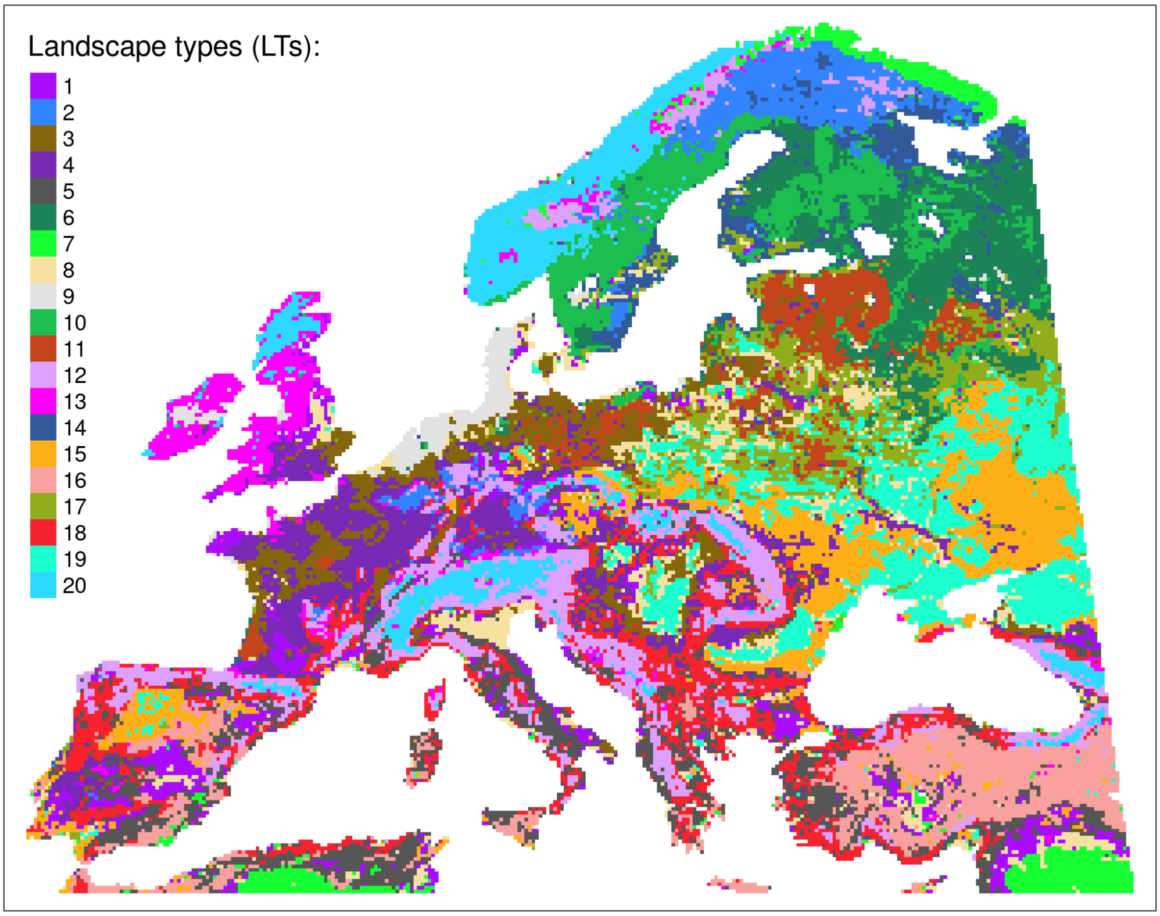


***Figure 2*** *Map of 20 landscape types in Europe obtained by clustering local landscapes of INCOMA embeddings of patterns of land cover, landforms, and soils.*

The quality of delineation depends on how similar patterns are within a single LT and how dissimilar are patterns between different LTs. The intra-cluster dissimilarity of an LT, $\delta$, is calculated based on an average dissimilarity between all LLs within a zone. Smaller values of $\delta$ indicate better-defined zones/LTs. The inter-cluster dissimilarity between a given LT and other LTs is calculated based on a distance between the given cluster and the rest of the clusters using the average linkage. The metric $\beta$ for a given LT is an average of values of average linkage between this LT and all other LTs. Larger values of $\beta$ indicate more distinct zones/LTs.

Values of metric $\beta$ are mostly high (between 0.63-0.91), which means that LTs are distinct from each other. Values of metric $\delta$ have a large range, from as small as 0.12 for LT 19 to as large as 0.73 for LT 7.

Inhomogeneity can also be assessed visually with a pattern mosaic. Pattern mosaic is an artificial rearrangement of a subset of randomly selected areas belonging to a given zone. Figure 3 shows examples of two pattern mosaics, for the most homogeneous LT 19 and the least homogeneous LT 7. LT 19 has land cover consisting mostly of agricultural areas with smaller, dispersed patches of settlements and forests located on plains. This LT's soils are mostly mollisols, with substantial areas of alfisols, both naturally fertile soils. On the other hand, all three layers in LT 7 contains LLs characterized by various, not similar patterns, suggesting that this cluster should be divided further into two or three LTs. This LT was delineated by $K$-means algorithm as a single entity based on its dissimilarity to all other nineteen LTs ($\beta$ of 0.91).

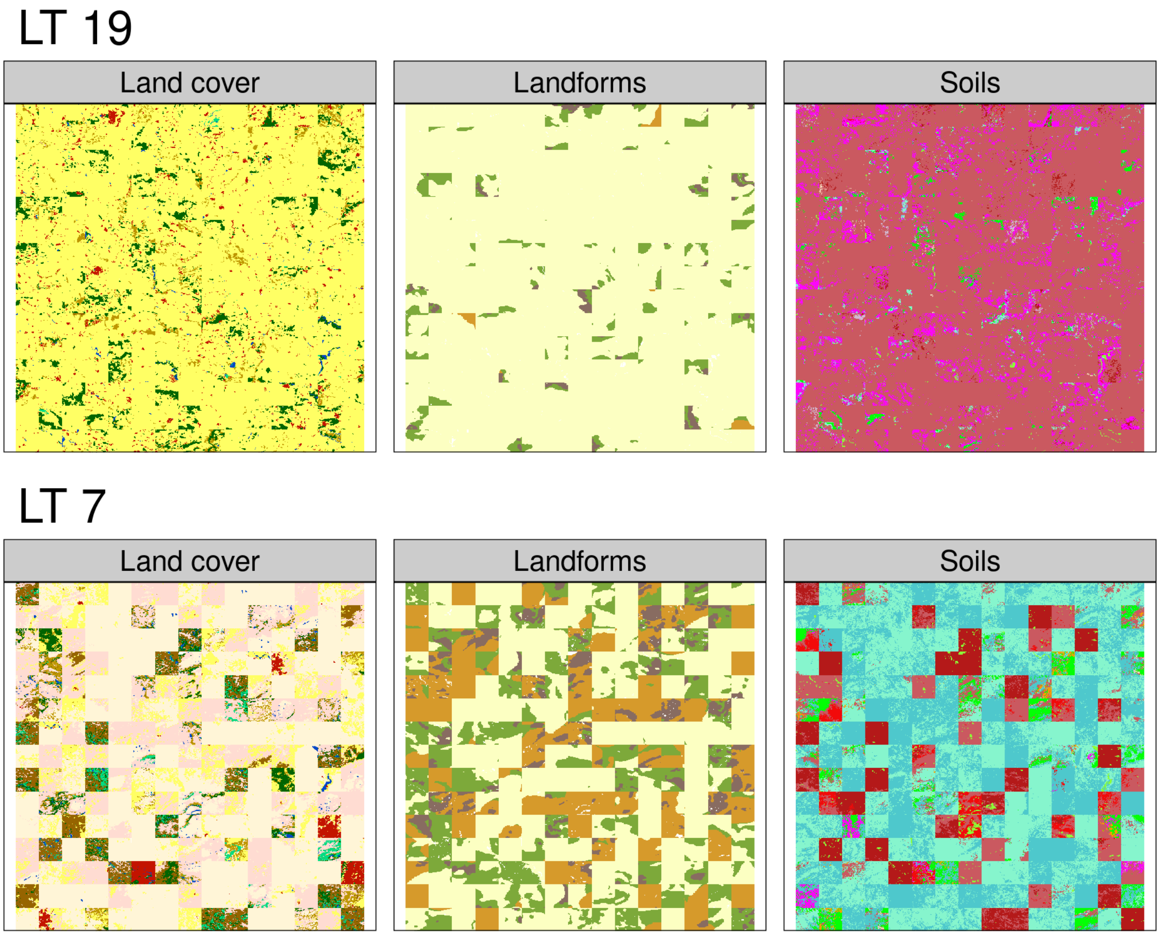


***Figure 3*** *Examples of two landscape types: the most homogeneous LT19 and the least homogeneous LT7.*

## 6. Conclusions

INCOMA brings together an interest in landscape classification with unsupervised machine learning methods like clustering and segmentation and provides a principled way to make maps of landscape types. INCOMA can also find application in tasks other than mapping LTs, which also require calculating similarities between multi-thematic patterns, such as content-based search and retrieval in spatial databases. The purpose of search and retrieval is to find LLs most similar to the query local landscape ($LL_0$), which can be achieved by calculating INCOMA-based dissimilarities between $LL_0$ and each LL in the area of interest and select LLs characterized by the smallest dissimilarities values as the answer to the query.

## References


ECMWF (2019). "ICDR Land Cover Product User Guide and Specification."

Haralick, Robert M., K. Shanmugam, and Its'Hak Dinstein (1973). "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–21. https://doi.org/10.1109/TSMC.1973.4309314

Hengl, Tomislav, Jorge Mendes de Jesus, Gerard BM Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, et al. (2017). "SoilGrids250m: Global Gridded

Soil Information Based on Machine Learning.” *PLoS One* 12 (2): e0169748.

Karagulle, Deniz, Charlie Frye, Roger Sayre, Sean Breyer, Peter Aniello, Randy Vaughan, and Dawn Wright (2017). “Modeling Global Hammond Landform Regions from 250-m Elevation Data.” *Transactions in GIS* 21 (5): 1040–60. https://doi.org/10.1111/tgis.12265

Mücher, C. A., J. A. Klijn, D. M. Wascher, and J. H. Schaminée (2010). “A new European Landscape Classification (LANMAP): A transparent, flexible and user-oriented methodology to distinguish landscapes.” *Ecological Indicators* 10(1): 87–103.

Omernik, J. M., and G. E. Griffith (2014). “Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework.” *Environmental Management* 54 (6): 1249–66.

Sayre, R., J. Dangermond, C. Frye, R. Vaughan, P. Aniello, S. Breyer, D. Cribbs, et al. (2014). “A new map of global ecological land units – an ecophysiographic stratification approach.” Tech. rep., Washington, DC: Association of American Geographers.

Vadivel, A., Shamik Sural, and A. K. Majumdar (2007). “An Integrated Color and Intensity Co-Occurrence Matrix.” *Pattern Recognition Letters* 28 (8): 974–83. https://doi.org/10.1016/j.patrec.2007.01.004

Wickham, J. D., and D. J. Norton (1994). “Mapping and analyzing landscape patterns.” *Landscape Ecology* 9 (1): 7–23.

**Biographies**

Jakub Nowosad is an assistant professor in the Institute of Geoecology and Geoinformation at Adam Mickiewicz University, Poznan, Poland. His main research is focused on developing and applying spatial methods to broaden our understanding of processes and patterns in the environment. It includes creating and improving geocomputational methods and software.

Tomasz Stepinski is a professor at the University of Cincinnati and a Director of Space Informatics Lab. His recent area of research is a development of automated tools for intelligent and intuitive exploration of very large Earth and planetary datasets. He is also interested in computational approaches to geodemographics.