On using landscape metrics for landscape similarity search

Jacek Niesterowicz, Tomasz F. Stepinski

Space Informatics Lab, Department of Geography, University of Cincinnati, Cincinnati, USA, OH 45221-0131, USA

Abstract

Landscape similarity search involves finding landscapes from among a large collection that are similar to a query landscape. An example of such collection is a large land cover map subdivided into a grid of smaller local landscapes, a query is a local landscape of interest, and the task is to find other local landscapes within a map which are perceptually similar to the query. Landscape search and the related task of pattern-based regionalization, requires a measure of similarity – a function which quantifies the level of likeness between two landscapes. The standard approach is to use the Euclidean distance between vectors of landscape metrics derived from the two landscapes, but no in-depth analysis of this approach has been conducted. In this paper we investigate the performance of different implementations of the standard similarity measure. Five different implementations are tested against each other and against a control similarity measure based on histograms of class co-occurrence features and the Jensen-Shannon divergence. Testing consists of a series of numerical experiments combined with visual assessments on a set of 400 3km-scale landscapes. Based on the cases where visual assessment provides definitive answer, we have determined that the standard similarity measure is sensitive to the way landscape metrics are normalized and, additionally, to whether weights aimed at controlling the relative contribution of landscape composition vs. configuration are used. The standard measure achieves the best performance when metrics are normalized using their extreme values extracted from all possible landscapes, not just the landscapes in the given collection, and when weights are assigned so the combined influence of composition metrics on the similarity value equals the combined influence of configuration metrics. We have also determined that the control similarity measure outperforms all implementations of the standard measure.

Keywords: landscape pattern, landscape metrics, similarity search, regionalization, similarity measure

1. Introduction

Similarity search is a key technology for data scientists with applications in information retrieval, data mining, decision making, event detection etc. (Zezula, 2012). It can be defined as the retrieval of the “closest” objects to a query in a database. Recently, as the notion of landscape as a mosaic of land cover patches spreads from its original application in ecology to general analysis of land use/land cover (LULC) patterns (Uuemaa et al., 2013), the similarity search is starting to be applied to landscapes. Two types of analyzes utilize similarity search: (a) regionalization of LULC maps into landscape pattern types (Wickham and Norton, 1994), and (b) landscape search, where the goal is to identify from among a large set of different landscapes those showing perceptual similarity to a template landscape. It is assumed that two perceptually similar landscapes express closely related meanings, and fulfill similar functions.

Pattern-based regionalization aims at delineation of a pixel-based LULC map into sub-regions characterized by unique, stationary LULC patterns. This is achieved by dividing the entire map into a regular grid of local landscapes (rectangular blocks of pixels interpreted as mosaics of LULC class patches restricted to the interior of a block) which are clustered in data space (using a similarity measure) to discover and delineate landscape pattern types present in the LULC map. From the computational perspective regionalization (Werlen, 2009) is an unsupervised classification resulting in the generalization of the original LULC map into a useful overview across a large area. Cardille and Lambois (2009) ap-
plied such a method to the 1992 National Land Cover Dataset (Vogelmann et al., 2001) and identified 17 landscape pattern types summarizing typical patterns (on the spatial scale of 6.5km) of LULC across the conterminous U.S. Using the Earth Observation for Sustainable Development of Forests land-cover (EOSD) dataset (Wulder et al., 2008) Partington and Cardille (2013) regionalized the forested part of Quebec, Canada into several landscape pattern types (at a spatial scale of 30km) and concluded that the method fulfilled the basic needs of land management across large areas: produced an overview of a large area, while highlighting useful subsets for closer inspection. Long et al. (2010) regionalized 5.5 million ha of primarily forested land situated within the interior plateau of British Columbia, Canada into six landscape pattern types (at a spatial scale of 1km) using also the EOSD data.

Unlike a regionalization task, a landscape search task relies on the principles of supervised classification. Given a local landscape (or a set of landscapes) of interest it discovers other local landscapes having similar structures. Andrew et al. (2012) utilized landscape search to identify de facto protected areas in boreal Canada. They used structures (patterns of land cover classes) of local landscapes in formally protected areas as a training set for finding local landscapes with similar structures in areas which were not formally protected but presently inaccessible. The identified areas are valuable candidates for protected area expansion. In a similar study Cardille et al. (2012) identified representative local landscapes located in Ontario’s parks and protected areas and searched for similar landscapes in parts of Ontario which are currently unprotected. Their goal was to select some of the identified areas for long-term monitoring to establish benefits of protection. In ecology, Dilts et al. (2010) recognized the need for a landscape search in order to identify control sites for sites that will experience treatment (for example, a road construction or forest harvesting). In epidemiology, Roux et al. (2011) used landscape search to infer the occurrence of Chagas disease. They identified landscapes structurally similar to those where the disease was known to be present to enable closer inspection for the possible presence of the disease.

The denominator of all these investigations is that they all incorporate knowledge discovery and the concept of holistic perceptual measure of similarity between landscapes. A holistic perceptual similarity between two landscapes (hereafter referred to as a similarity) is a function that gives a non-negative number to each pair of landscapes to define a notion of an overall likeness or sameness between them without regard to minute differences and spatial orientation. Thus, it assesses, in a single number, a degree to which two landscapes have similar utility or function. A large number of different similarity measures have been proposed in data science literature Cha (2007). This is because no single similarity measure is appropriate and effective to all problems, instead the most appropriate similarity measure needs to be matched to the data and the analytic task at hand. However, all the aforementioned studies on landscape regionalization and search used the same, standard approach to calculating a similarity measure. This standard approach consists of representing landscapes as feature vectors of landscape metrics (LMs) (Haines-Young and Chopping, 1996; Herzog and Lausch, 2001) and calculating the Euclidean distance between feature vectors to quantify similarity between two landscapes. To the best of our knowledge no study has been conducted which investigates the appropriateness and effectiveness of this approach. As similarity search-based analyzes of landscapes are becoming more frequent, it is important to use an optimal measure of similarity. The aim of this paper is to investigate how different implementations of the standard landscape similarity measure affect the results, and to recommend its best implementation. An additional aim is to compare a standard similarity measure with a measure based on histograms of class co-occurrence features and the Jensen-Shannon divergence.

2. Landscape similarity

Following McGarigal et al. (2002) we use the term “landscape” as ... an interacting mosaic of patches relevant to the phenomenon under consideration. In this paper LULC patterns are referred to as landscapes and the sought after LULC pattern types are referred to as landscape pattern types. Thus, the landscape we consider is the LULC raster which has \( K \) nominal labels \( c_1, \ldots, c_K \) describing \( K \) land cover classes. The entire spatial extent of the LULC raster map is referred to as a region. The region is subdivided (without overlap) into a lattice of local landscapes (LL). Thus, from the data science perspective, a region is a spatial database of LLs. Landscape search can be thought of as a special case of similarity search – the retrieval of “closest” objects (LLs) to a query (a selected LL) in the database. Regionalization can be thought of as a clustering of all objects (LLs) in database. A specific LL (denoted by \( \mathcal{L} \)) is a \( n \times n \) block of pixels. The size \( n \) sets the spatial scale over which the pattern of LULC classes is defined as a local landscape. A landscape similarity measure requires two elements: (a) a mathematical description
of the landscape (called a signature) and (b) a similarity function which takes two signatures as arguments and returns the value of similarity between the landscapes.

2.1. Calculating landscape dissimilarity using landscape metrics

In working with landscapes the standard practice is to use a vector of LMs as a landscape signature and the Euclidean distance as a similarity function. Note that distance, which assesses the degree of “unlikeness” between two patterns, is the opposite of similarity and is better referred to as dissimilarity in the present context. We use the notions of similarity and dissimilarity interchangeably as they are easily convertible.

LMs are algorithms that quantify the specific spatial characteristics of a landscape pattern. A large number of different metrics characterizing individual patches, classes of patches, and entire landscape mosaic have been developed and collected in a single computer program FRAGSTATS (McGarigal et al., 2002). A patch is a contiguous group of same-class pixels. In a standard approach a signature of landscape \( \mathcal{A} \) is a vector \( (a_1, \ldots, a_N) \) consisting of the values of \( N \) different LMs calculated for \( \mathcal{A} \) and the dissimilarity function is the Euclidean distance between two LMs \( \mathcal{A} \) and \( \mathcal{B} \)

\[
d_e(\mathcal{A}, \mathcal{B}) = \sqrt{(\mu_1(a_1-b_1)^2 + \ldots + \mu_N(a_N-b_N)^2)} \tag{1}
\]

where \( (a_1, \ldots, a_N) \) and \( (b_1, \ldots, b_N) \) are vectors of LMs calculated from \( \mathcal{A} \) and \( \mathcal{B} \), respectively, and \( (\mu_1, \ldots, \mu_N) \) are the weights to reflect the relative importance of a given LM to the overall value of dissimilarity. There are a number of issues that arise with the standard approach: (a) Which LMs should be selected for a landscape signature? (b) How should LMs in the signature vector be normalized or standardized if at all? (c) What values of weights should be assigned, if any?

2.1.1. Selection of landscape metrics

Compositional metrics (Gustafson, 1998) need to be included in the signature as the composition of a landscape is its primary characteristic. Only landscape-level configuration LMs should also be included in a landscape signature because only landscape-level LMs can be calculated for all LLs regardless of their individual composition. Ideally, a landscape signature should consist of the values of LMs that are independent and together describe adequately the character of the pattern for all possible patterns in a region. Cushman et al. (2008) identified seven landscape structure components – linear combinations of LMs obtained using the principal components analysis (PCA) – that were independent and universal, at least on a set of 531 landscapes (at a ~7km length scale) from across three different regions in the U.S. Conceivably, those components could be good candidates for landscape signatures but because they have been established on the basis of a single scale landscape and over a limited number of all possible U.S. landscapes they may not apply to all landscapes across the entire United States.

Existing work on the regionalization of LLs and on landscape searches use either a small number of hand-picked LMs (Long et al., 2010; Roux et al., 2011; Cardille et al., 2012) or all LMs available in FRAGSTATS (Cardille and Lambois, 2009; Partington and Cardille, 2013). If a large, indiscriminate set of LMs is used a PCA is performed to ensure that calculations of similarity between landscapes were not biased toward certain aspects of landscapes that had been computed redundantly in the initial metric set (Cardille et al., 2012). However, although using PCA eliminates redundancy of metrics in a landscape signature, it does not alleviate issues related to signature normalization and selection of weights.

2.1.2. Normalization and standardization of LMs

Different LMs have different meanings and different ranges of values. In order to obtain a meaningful value for Euclidean distance they need to be scaled to have identical ranges (normalization) or at least similar ranges (z-score standardization), otherwise the Euclidean distance would be dominated by LMs with the largest ranges of values. However, both normalization and standardization have their shortcomings. Some LMs have a priori known ranges, but many do not, thus in practice the ranges are calculated empirically from the data. If the values of a given LM for all LL in a region are a small subset of its possible values the normalization will exaggerate the differences between LLs. Standardization has the same problem, but additionally LMs with different distributions of values will have different ranges after standardization which will further distort the measure of dissimilarity. Standardization was used by Long et al. (2010) and unspecified scaling was used by Cardille and Lambois (2009), Cardille et al. (2012), and Partington and Cardille (2013), as a part of the PCA transformation. Note that the PCA algorithm performs normalization based on the minimum and maximum values of the data from a given region but we will argue that such local normalization is one of the issues when calculating a similarity between landscapes.
2.1.3. Weights

Weights can be assigned to increase or decrease the relative contribution of a given attribute (an LM or a PCA component) to the value of the Euclidean distance. Note that a PCA transformation does not make weights redundant. PCA components are uncorrelated but their relative importance to an overall similarity between landscapes can only be set by an analyst. Weights are set subjectively, but not using the weights is equivalent to setting them all to the same value – a subjective choice that may lead to grossly inaccurate assessment of the level of dissimilarity between two landscapes (see section 4). We suggest that in the absence of any additional information the weights should be set to ensure that total contributions of all composition and all contribution metrics (Boots, 2003) to the similarity value are the same.

2.2. Calculating landscape dissimilarity using histograms of pattern features

Choices other than vectors of LMs and Euclidean distance are possible for defining a landscape signature and for selection of the dissimilarity function. Recently, Jasiewicz and Stepinski (2013) and Stepinski et al. (2014) proposed the calculation of dissimilarity between two landscapes using an approach originally developed for the comparison of images in the context of the Content-Based Image Retrieval (CBIR) (Gevers and Smeulders, 2004; Datta et al., 2008; Lew et al., 2006). This approach uses a histogram of pattern features as a signature and a histogram dissimilarity measure as a distance function. It has the advantage of being extensively studied, for example by Rubner et al. (2001), for the purpose of image retrieval (an analog to landscape search) and proved to be an effective and robust tool for landscape search as well (Stepinski et al., 2014; Niesterowicz and Stepinski, 2013).

Pattern features are simple local elements of a pattern. Many different pattern features are possible, here we discuss only one – a color co-occurrence feature (Barnsley and Barr, 1996; Chang and Krumm, 1999). When applied to landscapes, colors are replaced by LULC classes and a “feature” is a pair of classes assigned to two neighboring pixels. Examples of possible features include forest-urban, forest-forest etc. For a region with \( K \) LULC classes there are \( (K^2 + K)/2 \) different features possible, \( K \) of them correspond to same-class pairs, which measure the composition of the classes in the landscape, and \( (K^2 - K)/2 \) correspond to different-class pairs, which measure the configuration of classes in the landscape. Assuming eight-connectivity between pixels, each pixel generates eight features but their total number is halved, as the same feature is generated twice by the pairs of neighboring pixels. All features are histogrammed and the histogram of co-occurrence features is termed a landscape signature.

Many different measures of dissimilarity between histograms have been discussed in data science literature (Cha, 2007), here we discuss only one – the Jensen-Shannon divergence (Lin, 1991). For two co-occurrence histograms \( A \) and \( B \), representing landscapes \( \mathcal{A} \) and \( \mathcal{B} \), the Jensen-Shannon divergence (JSD) is given by

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

(2)

where \( H(X) \) indicates the value of the Shannon entropy (Shannon, 1948) for the histogram \( X \). The JSD is always defined, symmetric, bounded by 0 and 1, and equal to 0 only if \( A = B \). We use \( d_{JS}(\mathcal{A}, \mathcal{B}) = \sqrt{\text{JSD}(\mathcal{A}, \mathcal{B})} \) as a measure of dissimilarity between the landscapes because \( \sqrt{\text{JSD}} \) has been proved (Endres and Schindelin, 2003) to be a metric.

3. Methods and Data

There are no established methods to compare the performance of different dissimilarity measures for application to landscapes. This differs from the CBIR domain where performances of different image retrieval

---

**Table 1: Six different landscape dissimilarity measures**

<table>
<thead>
<tr>
<th>ID</th>
<th>Abbrev.</th>
<th>Signature</th>
<th>Distance</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CJS</td>
<td>co-occurrence features</td>
<td>Jensen-Shannon</td>
<td>does not use landscape metrics</td>
</tr>
<tr>
<td>2</td>
<td>MET</td>
<td>landscape metrics</td>
<td>Euclidean</td>
<td>metrics normalized using region extrema, no weights</td>
</tr>
<tr>
<td>3</td>
<td>METW</td>
<td>landscape metrics</td>
<td>Euclidean</td>
<td>metrics normalized using region extrema, weights</td>
</tr>
<tr>
<td>4</td>
<td>PC</td>
<td>landscape metrics</td>
<td>Euclidean</td>
<td>uses principal components</td>
</tr>
<tr>
<td>5</td>
<td>METG</td>
<td>landscape metrics</td>
<td>Euclidean</td>
<td>metrics normalized using global extrema, no weights</td>
</tr>
<tr>
<td>6</td>
<td>METWG</td>
<td>landscape metrics</td>
<td>Euclidean</td>
<td>metrics normalized using global extrema, weights</td>
</tr>
</tbody>
</table>
algorithms are measured objectively using a ground truth dataset of images with content pre-labeled by an analyst. This is possible because the objective of CBIR is to retrieve images containing an object that is simple to describe, for example, a dog or a house. On the other hand landscapes are so complex that, in most non-trivial cases, they cannot be assigned a simple semantic label. Thus, using a ground truth dataset for assessing the performance of a landscape dissimilarity measure is not feasible. Instead, we use a series of numerical experiments (Stepinski and Cohen, 2014) utilizing dissimilarity matrices to determine differences between the measures and to judge their relative performance. Given a region with \(N\) LLs, a dissimilarity matrix is the \(N \times N\) matrix containing values of dissimilarity between every pair of LLs. Each dissimilarity method generates a different dissimilarity matrix.

As a test site we use a \(2000 \times 2000\) pixel NLCD 2011 (Homer et al., 2015) region located in the state of Georgia (see Fig. 1A) and centered on 81.87W and 33.16N. There are \(K=15\) different land cover classes within this region (the sixteenth NLCD class, “perennial ice/snow,” is not present at the test site). The region is divided into \(N = 20 \times 20 = 400\) non-overlapping LLs. Each LL has a linear scale of \(n = 100\) pixels, or 3km. For further reference LLs are numbered row-wise starting from the upper-left (LL #1) to the lower-right (LL #400).

Table 1 lists the six different landscape dissimilarity measures selected for comparison. Five of these methods are based on landscape metric signatures and the Euclidean distance, but they differ in the way the LMs data is handled (see sections 2.1.1 to 2.1.3). The sixth method is based on a histogram of co-occurrence features and the Jensen-Shanon divergence (see section 2.2). For the five methods which utilize landscape metrics we calculated 69 metrics. The first 15 metrics are class-level metrics describing the composition of LL pattern. The remaining 54 metrics are landscape-level metrics describing the configuration of LL pattern. The list of the metrics is given in the supplementary material, their names and definitions are the same as in FRAGSTATS (McGarigal et al., 2002). All composition metrics are normalized to a range between 0 (no given LULC class in the LL) to 1 (the entire LL consists of this LULC class). All configuration metrics are normalized to a \((0, 1)\) range by either using their minimum and maximum values on the test set of 400 LLs (local normalization), or their minimum and maximum values on the set of 893,840 LLs of the same scale covering the entire conterminous United States (global normalization). In MET and METG methods no weights are set, and \(\mu_i = 1\) for \(i = 1, \ldots, 69\). For METW and
METWG methods weights are set in the following fashion, \( \mu_i = 69/30 \) for \( i = 1, \ldots, 15 \) and \( \mu_i = 69/108 \) for \( i = 16, \ldots, 69 \). This ensures that the 15 composition metrics have cumulatively the same weight as the 54 configuration metrics. The PC method uses the first 13 principal components (calculated from all 69 metrics) which cumulatively account for 90% of variability. The principal components do not depend on normalization and no weights are assigned to them.

In order to compare the relative performance of the six landscape dissimilarity measures we perform four different assessments as described below.

3.1. Correlation of dissimilarity matrices

A 400 \times 400 dissimilarity matrix encapsulates all available information about the relative dissimilarities between LLs in the test site. We calculated six dissimilarity matrices, one for every dissimilarity measure listed in Table 1. The structure of a dissimilarity matrix dictates the results of a potential landscape search over the site or the outcome of site’s regionalization into landscape types. If the two methods yield highly correlated dissimilarity matrices they will yield similar results for landscape search and regionalization, however, weakly correlated distance matrices point to significant differences in search and regionalization results. We calculate a correlation value between every possible pair (15 pairs in all) of dissimilarity matrices using the Mantel test (Mantel, 1967).

3.2. Comparing dissimilarity methods using percentiles of distances

For any pair of LLs at a site we have six different values (stemming from six different similarity measures) quantifying their mutual dissimilarity. However, these values are incomparable due to differing ranges and the shapes of distance distributions stemming from each dissimilarity method (see Fig. 1B). In order to make the distances arising from different dissimilarity measures comparable to each other we transform them to their percentiles using an empirical cumulative distribution function constructed from the values of respective dissimilarity matrices. Thus, for any pair of LLs we have six different values of percentiles. A dissimilarity measure yielding the smallest percentile assesses the two LLs as the most similar, whereas a dissimilarity measure yielding the largest percentile assesses them as the least similar. We conduct a round-robin “tournament” between all dissimilarity measures resulting in 15 different comparisons. In each comparison we identify pairs of LLs for which a difference in percentile distance is large (one measure indicates large similarity between the LLs while the other indicates that LLs are dissimilar). We visually assess these pairs and decide which measure yields a value more in agreement with our assessment. A dissimilarity measure with more “wins” performs better than a dissimilarity measure with less wins.

3.3. Comparing results of search queries

This is another round-robin tournament between the six different dissimilarity measures. Each of the 15 comparisons is aimed at assessing whether the two participating similarity measures return similar search results, and, in the case that they don’t, assesses which measure returns better results. For each of 400 LLs in the test site we run two similarity searches (one for each participating measure) aimed at finding the closest match to a query from among the remaining 399 LLs. We save the five best matches from each search and calculate an average dissimilarity between the two sets of retrieved LLs (using the CJS measure). We then construct a grid of the same dimensions as the lattice of LLs to visualize the results of the similarity search comparison. In this grid the cells are colored according to the value of average dissimilarity between the results of the two searches. Lighter colors indicate LLs for which the two searches returned similar results and darker colors indicate LLs for which the two searches returned different results. We visually assess the cases where search results differ significantly and decide which dissimilarity method performs better.

3.4. Regionalization

We regionalize the test site by clustering the 400 LLs in a distance space. There are several possible methods of performing such clustering including \( k \)-medoids (used by Long et al. (2010), affinity propagation (Frey and Dueck, 2007) (used by Cardille and Lambois (2009) and Cardille et al. (2012)), and finding communities in a graph stemming from translating a dissimilarity matrix into an adjacency matrix (used by Stepinski and Cohen (2014)). Here we use a hierarchical clustering with Ward linkage (Ward, 1963) to perform clustering of LLs. All these methods use a dissimilarity matrix as input and the results are much more sensitive to the structure of the values in a matrix than they are to the choice of clustering algorithm. We examine each regionalization for homogeneity of LULC patterns within each landscape type and assess its overall agreement with how an analyst would delineate the site.
4. Results

Table 2 shows the correlations between dissimilarity matrices calculated using the Mantel test. The above-diagonal values in Table 2 pertain to correlations between the original dissimilarity matrices and the below-diagonal values pertain to correlations between matrices transformed to percentiles; there are no significant differences between the corresponding values of correlation. In general, correlation values between the five matrices stemming from landscape metrics/Euclidean distance measures have relatively high correlation values indicating that the overall structure of dissimilarities between the LLs are similar when calculated using these measures. However, it needs to be pointed out that even small differences in the structure of a distance matrix could lead to significant differences in regionalization outcomes. The correlations between the CJS and the rest of the measures are smaller; the smallest being with the MET measure and the largest with the METWG measure. It follows that the results of the landscape search or regionalization using the CJS measure will be the most similar to the results using the METWG measure and the least similar to results using the MET measure.

Fig. 2 shows the distributions of variable \( d^p_X(A, B) - d^p_Y(A, B) \), where \( d^p_X(A, B) \) is a distance between two LLs \( A \) and \( B \) calculated using a measure \( X \) and expressed as a percentile. These distributions show the spread of disparity between different dissimilarity measures. They are peaked around the value of 0 indicating that for the majority of LLs pairs different measures yield similar assessments. When two measures give a similar assessment of the dissimilarities between landscapes, we cannot assess, via visual inspection of LLs, which of these two values better reflects the difference between two patterns. This is because human perception can only assess alikeness between patterns qualitatively but not quantitatively as humans have difficulties with putting a number on the degree of alikeness. We have determined experimentally that the minimum difference between the two dissimilarity values (expressed in percentiles) needs to be 0.5 for us to judge which of these two measures reflects an actual state of alikeness between the two LLs. The numbers on the left and right of each panel in Fig. 2 give the percentage of the LL pairs for which the difference in dissimilarity assessment is > 0.5. Only for these LL pairs are we able to visually determine which measure gives a better assess-
Fig. 3 shows eight examples of LL pairs for which we can visually determine which measures gives a dissimilarity value in a better agreement with our visual assessment. For simplicity only two measures are considered per each LL pair. The examples labeled A to D pertain to comparisons between measures based on the Euclidean distance, whereas the examples labeled E to H pertain to comparisons between the Euclidean-based measures and the CJS. Each example shows two LLs and the values of dissimilarity between them as calculated by the two measures under consideration. Recall that these dissimilarities are given in terms of percentiles. Thus, in the example A the MET indicates that 92% of LL pairs in the test site are more similar than LLs represented by tiles 355 and 389. On the other hand, the METG indicates that only 12% of LL pairs in the test site are more similar than tiles 355 and 389. Given the variety of LLs in the test site it is our judgment that a dissimilarity value yielded by METG corresponds better to a human perception of landscape similarity than the value yielded by MET. We indicate our judgment by putting an asterisk next to METG. Notice that by restricting our visual comparison only to the
Figure 4: Visualization of the components of the Euclidean distance between two local landscapes. The upper panel visualizes the components of the distance between tiles 153 and 195 whereas the bottom panel visualizes components of the distance between tiles 8 and 11. The components are labeled by their metric ID (from 1 to 69) and their values are given on the vertical axis. Different colors correspond to different normalizations of metrics and/or presence or absence of weights. The values of distances (expressed as percentiles) are given in a legend.

As can be seen in Fig. 3 the Euclidean-based measures can sometimes give an assessment of dissimilarity between LLs which is at odds with what an analyst would assess. It is important to pinpoint specific reasons for this disagreement. The Euclidean distance (eq.1) consists of the sum of squared differences between values of individual metrics in the two LLs. To visualize the Euclidean distance between two LLs we plot its 69 components \( |a_i - b_i| \), \( i = 1, \ldots, 69 \). Fig. 4 shows such visualization for two pairs of LLs.

The upper part of Fig. 4 visualizes the Euclidean distance between tiles 153 and 195. The LLs represented by these tiles are perceptually similar but MET assesses them as highly dissimilar (99%). The METW and METG measures also assess them as dissimilar but to a smaller degree (96% and 71%, respectively), while the METWG measure assesses them as moderately similar (22%). The panels A to C on Fig. 4 visualize the four different Euclidean distances between LLs 153 and 195 corresponding to MET, METW, METG, and METWG. To provide better clarity only two distances are compared on each panel. For the Euclidean distance to assess a small value of dissimilarity all its components need to be small. As we can observe in panel A this is not the case for neither MET nor METW. Many elements corresponding to configuration metrics are large meaning that the values of these metrics in the two LLs are significantly different. Because the LLs are visually similar the values of these metrics are not really sig-
significantly different but they become different after being subjected to local normalization (MET) that exaggerate the differences. Adding weights (METW) does not eliminate the problem. The problem is alleviated by performing global normalization (panel C) which better preserves actual differences between the values of metrics. Overall, this example illustrates an importance of using global normalization when assessing landscape similarity by means of the metrics-based Euclidean distance.

The lower part of Fig. 4 visualizes the Euclidean distance between tiles 8 and 11. The LLs represented by these tiles are perceptually dissimilar but MET assesses them as highly similar (2%). For the Euclidean distance to assess LLs as dissimilar some of its components need to be large. As can be seen on panel D few composition-related components of the Euclidean distance (MET) are large but all configuration-related components are small resulting in a small value of the Euclidean distance. Assigning weights (METW) boosts influence of composition resulting in more realistic assessment of dissimilarity (45%). Panel F shows that using global normalization that a combination of global normalization and weights (METWG) results in a more acceptable (67%) assessment of dissimilarity between these two LLs. Overall, this example illustrates the importance of using weights when assessing landscape similarity by means of the metrics-based Euclidean distance.

Fig. 5 shows the results of a round-robin tournament of search queries as described in section 3.3. A white color indicates LLs for which the two dissimilarity measures yield similar matches while a red-to-brown color
indicates LLs for which the two measures yield different matches. In places where matches are different we review them and assess which measure resulted in matches closer to analyst’s perception; this measure is indicated by an asterisk. The results are the same as in the previous comparison illustrated on Fig. 3. In most areas of the test site all queries yield similar results, but in the areas where queries results are different the CJS measure yields matches which are closest to analyst’s perception, followed by the METWG. The MET and PC measures yield matches which are farthest from analyst’s perception. Fig. 5 also reveals that similarity searches performed on LLs located at the boundary of urban area and wetland are, in general, the most sensitive to the measure used, whereas LLs located at the bottom part of the region are least sensitive.

Our last assessment is the actual regionalization of the site as described in section 3.4. The aim of this test is to find whether different dissimilarity measures lead to different delineations of landscape types (LTs) and if so, which dissimilarity measure yields the best regionalization as judged by visual inspection. We look for stationarity of LULC pattern within each LT and diversity of patterns between different LTs. We use clustering of LLs to perform regionalization, the number of landscape types (LTs) is a free parameter. Here we show only results for regionalization into six LTs, a number experimentally determined to yield distinct yet stationary patterns. Only three dissimilarity measures, CJS (the overall preferred measure based on the previous tests), METWG (the preferred measure among the LM-based similarity measures) and MET (the least preferred measure) are tested for regionalization.

Fig. 6 shows the regionalization of the site using the CJS measure. The map of LTs is located in the middle of the figure and is surrounded by rectangles grouping most of LLs belonging to individual LTs. Combining all LLs within a given LT into a single contiguous rectangle helps in visual assessment of LT’s pattern stationarity and in judging uniqueness of each LT. The CJS measure did a good job delineating major types of landscape in the test site: urban-dominated (type 1), forest-dominated (type 3), wetland-dominated (type 6), cropland-dominated (type 5), and the two mixed cover mosaics (type 2 and type 4). Note that the names given to these LTs are composition-based even so the LTs are patterns characterized by composition and configuration of LULC patches. More in-depth investigation of delineated LTs is beyond the scope of this paper but could result in giving them names that could reflect both composition as well as configuration. The pattern in each LT is relatively uniform, although urban-dominated LT includes some wetland spots and wetland-dominated areas include some forest spots. This is a consequence of our arbitrary division of the site into square LLs and our selection of a 3km scale.

Fig. 7 shows the regionalization of the site using the METWG measure. We judge the METWG regionalization as worse than the CJS regionalization because the LT #1 has a non-stationary pattern (containing urban and non-urban mosaics) and because the LT #4 is not distinct enough from the LT #6. The METWG measure isolates urban-dominated mosaics if seven instead of six LTs are assumed, but it always stays one step behind the CJS measure which, at seven LTs, divides mosaic type 2 into two more uniform mosaics.

Finally, Fig. 8 shows the regionalization of the site using the MET measure. The MET yields unacceptable regionalization which lacks pattern stationarity across most of LTs. The urban-dominated mosaic is not isolated, the forest and wetland mosaics are in a single LT, and the meaning of the types 6 and 4 is unclear. The poor performance of the MET measure on regionalization task is not surprising in light of our earlier tests.

5. Conclusions and discussion

Similarity search is a technology that is beginning to be recognized as a valuable tool for landscape analysis, especially with application to very large LULC datasets like NLCD and EOSD mentioned in section 1, as well as GLC30 (Chen et al., 2014), CORINE, and CropScape (Han et al., 2012). Several papers utilizing the concept of similarity search for land cover patterns have already been published (see section 1) and more work is certain to follow. Our aim in this paper was to assess the performance of the standard implementation of landscape similarity search to inform future practitioners of the method about potential pitfalls and to suggest best practice approaches. Our overall conclusion is that calculating landscape similarity using LMs and the Euclidean distance is susceptible to undesirable artifacts and needs to be performed with attention to details. The core of the issue is how to combine information from many LMs having different meanings, wide ranges of values, and varying degrees of relevance into a single number that can effectively quantify the overall similarity between two landscapes in a meaningful way. This problem, not unique to landscape similarity, is addressed in machine learning by normalization and weighting of contributing features (LMs). However, with application to landscapes it has not been clear how exactly to perform these two pre-processing steps.
Figure 6: Regionalization of the test site using the CJS dissimilarity measure (center). Local landscapes in each landscape type are grouped into corresponding rectangles so the character and uniformity of their land cover pattern can be assessed visually. See Fig. 3 for NLCD legend.

Figure 7: Regionalization of the test site using the METWG dissimilarity measure (center). Local landscapes in each landscape type are grouped into corresponding rectangles so the character and uniformity of their land cover pattern can be assessed visually. See Fig. 3 for NLCD legend.
The focus in previous work (Cardille and Lambois, 2009; Partington and Cardille, 2013) was on eliminating correlated LMs by means of PCA to avoid biasing the similarity assessment toward latent landscape features described by multiple LMs. However, the correlation-based PCA algorithm always normalizes or standardizes LMs on the basis of data to be transformed into principal components (local normalization). As we have demonstrated this can lead to spurious results nullifying any benefit of eliminating correlated LMs. What is needed is a normalization over a theoretical range of LMs. However, many LMs do not have well-defined ranges, instead their ranges are case-dependent. The only viable solution is to establish the ranges of LMs empirically from the biggest and broadest possible set of landscapes. We calculated LMs for ~900,000 3km scale LLs across the entire U.S. and used resultant ranges of LMs to perform global normalization of LMs calculated from 400 LLs in our test site. To achieve this task we used our custom-made, high-performance GRASS-GIS (Neteler and Mitasova, 2007) extension module (available upon request from the authors). Our module reads in the entire NLCD, partitions it into 3km scale LLs and calculates the 69 metrics (see supplementary material for the list of the metrics) for each of ~900,000 LLs. This calculation takes only 30 minutes on a workstation equipped with a single Core i7 processor running at 2.8GHz and having 16 GB memory. However, a similar calculation is hardly practical for the majority of potential authors using FRAGSTATS to calculate LMs. Another problem is that landscape pattern has two broad aspects, composition and configuration, but there are typically many more LMs characterizing landscape configuration than there are LMs characterizing its composition. This imbalance is not taken care of by the PCA; the only way to control relative contributions of landscape composition and landscape configuration to the overall measure of similarity is to introduce weights. In the absence of any additional information we propose to assign individual weights so that the combined weight of the composition metrics is equal to the combined weight of the configuration metrics.

Our numerical experiments revealed that the Euclidean distance-based measure with global normalization and weights outperforms (yields dissimilarity values closer to an analyst assessment) Euclidean distance-based measures utilizing other data pre-processing methods including the PCA. Although our conclusion was reached on the basis of one particular test site and for a single scale of landscape, our subsequent analysis revealed the theoretical underpinnings of our findings. Therefore, we are confident that our conclusions hold for all landscapes and all scales. Inadequacies of local normalization and a lack of weights are expected to have the most pronounced negative effect on similarity measures with smaller scale landscapes where it is easier to
find LLs dominated by a single LULC class (like the LL #153 or #195 in our study) or having similar configurations but different compositions (like the LL #8 or #11 in our study) – types of LLs particularly sensitive to proper normalization and weighting. Thus, in studies like that of Partington and Cardille (2013), where large ~30km scale LLs have been used, the resultant regionalization may be of sufficient quality despite being based on the PCA pre-processing. This however cannot be assessed from their paper which does not provide a measure-independent (for example, visual) assessment of uniformity of pattern within each landscape type.

Finally, we have found that the control method based on LULC class co-occurrence features and the Jensen-Shannon divergence yields values of dissimilarity between LLs which are in the closest agreement with an analyst’s assessment. This may seem surprising as this method uses only a single co-occurrence feature (similar to join count metric). However, instead of calculating an average count it describes a multi-categorical landscape using a histogram of join counts between all 15 LULC classes. Such histogram constitutes a rich description (120 values) which implicitly takes into account properties of pattern that could be explicitly described by a set of landscape metrics. By using a histogram rather than feature vector to represent the landscape the CJS method avoids problems with normalization and weights as all bins have the same meaning (counts) and are normalized to the total number of possible features (joins). This makes the CJS method robust in application to assessment of an overall perceptual similarity between landscapes. The performance of this method can be further assessed using an interactive GeoWeb application LandEx (Stepinski et al., 2014) that implements landscape overall similarity search using the NLCD over the entire U.S. LandEx is available at http://sil.uc.edu. On the other hand CJS method cannot be applied to assess similarity targeted at a specific property of landscape pattern.

Acknowledgments. This work was supported by the University of Cincinnati Space Exploration Institute and by the National Science Centre (NCN) under Grant DEC-2012/07/B/ST6/01206.

References


