Comparing Qualitative Raster Maps

WIECZOREK Małgorzata ¹, a and PRZYBYŁ Wojciech ², b

¹ University of Wrocław, Plac Uniwersytecki 1, 50-137 Wrocław, Poland
² Military Institute of Engineer Technology, ul. Obornicka 136, 50-961 Wrocław, Poland

a malgorzata.wieczorek@uwr.edu.pl, b przybyl@witi.wroc.pl

Keywords: Raster Maps, Validation, Filters

Abstract. Spatial diversity of the natural environment can be presented using raster qualitative data. They can be the result of collecting field data or be the result of stochastic modelling of a certain complexity of the environment. One example of such modelling is the determination of similar elements of relief forms on the basis of morphometric variables. The use of unsupervised methods for clustering raster data in modelling can produce different maps. It is necessary to assess the compatibility of the obtained maps in order to assess how significant the differences between them are. The article presents selected stochastic and deterministic methods for assessing the spatial distribution of data. Exemplary methods of measuring variability at the global and local level were discussed.

Introduction

In the era of relatively easy access to numerical data, including raster data, both qualitative and quantitative, the attention of scientists is directed not so much to the ways of obtaining data, but to the ways of using them, analyzing, modelling, and then assessing the results of their validation. Currently, the application of image analysis and classification (i.e. rasters) is used in many fields. The most common methods include the analysis of satellite and meteorological images, morphometric classification, medical image analysis (including tomograms, roentgenograms, ultrasound images, etc.), monitoring of product wear and quality (such as metal products, fabrics, paper, and foodstuffs), as well as analysis of surface images and microscopic images of material structures to monitor tool wear.

Algorithms have already been developed for recognizing objects in satellite images or diagnostic images used in medicine, often using neural networks or fuzzy logic. Comparing two images and assessing whether they are statistically [1] the same when these are modeling results and there is no pattern to relate to is still a challenge, although the topic has already been addressed in the literature [2]. This paper presents two such research situations along with a proposal of measures to assess the similarity of two images. The obtained results and limitations resulting from the measures used were discussed. In the article, the authors focused on exemplary global and local indicators that allow the assessment of differences between two images.

The methodology presented in this article goes beyond its applications solely in the field of topographic maps, as image analysis, particularly for images featuring stochastic fields, can be widely applied in industry and management, significantly impacting product quality enhancement [3-5], as well as modifying management methods [6,7] and related analytics [8,9]. The applications of such analyses encompass various issues, ranging from the assessment of corrosion phenomena [10], both in reference to typical metallic construction materials [11-13] and special alloys [14], to the visual evaluation of separator performance [15], and even to the automatic assessment of surfaces [16], coatings [17-19], and structural welds [20]. Currently, this significantly influences the organization of enterprises that have surpassed the level of Industry 3.0 and are reaching Industry 4.0, impacting the means of automatic control, for instance, in the application of DLC.
and ESD [24-26] coatings, as well as in the automation of thermal insulation evaluation using infrared cameras [27-29]. It is also a significant tool supporting the detection of machine damage [30,31], particularly hydraulic power equipment [32-34].

Automated methods for processing image maps are also becoming a vital tool in process optimization [35-37], including the application of qualitative and/or subjective factors [38-40], although this typically requires prior strong dimensionality reduction [41,42] to avoid unwanted correlations.

**Area of Study**

A fragment of the Owl Mountains (a mountain range in the Sudetes in central-southern Poland, located in the Lower Silesian Voivodeship) was selected for the research area with the intervention introduced so as to obtain a flat area. It is an area of 8,725 m in the E-W direction and 6,725 m in the N-S direction (269 rows x 349 columns). Using a digital elevation model (DEM) with a field resolution of 25 m, a k-median clustering with Manhattan metrics method was developed. This grouping was based on morphometric variables calculated from DEM, such as elevation, slope, aspect, curvature.

![Model 1 (Height, slope, exposure, curvature)](image1)

![Model 2 (Height, slope, curvature)](image2)

![Pattern (Height, Slope, Exposure, Curvature)](image3)

**Fig.1. Models and pattern of the area subject to classification, divided into classes**

Three images (Fig.1) were selected for the calculations, broken down into 4 classes (clusters) of landform elements: Model 1 – developed on the basis of all 4 morphometric variables, Model 2 – obtained on the basis of height, inclination and curvature, and the third considered as a Pattern -model obtained similarly to model 1, but characterized by the greatest intergroup variability (sum of square between clusters). Exposure is not a morphometric variable that is commonly included
in modelling. Having a tool (apart from the expert's assessment) to assess to what extent the inclusion of an exposure gives a significantly different modelling effect, it would be helpful in assessing the legitimacy of its inclusion in the model. An expert assessment of the advantages and disadvantages of modelling elements of sculpture forms with exposure was presented by Wieczorek and Migon [43], but advanced similarity measures were not used in this assessment.

**Problem to Solve**

In this case, we are interested in the answer to the question: How to assess which model fits the pattern better? Although they are not identical, is the scale of similarity large enough to consider them identical models? Where to put this limit of assumed similarity? Obtaining answers to such questions would allow us to decide whether it is worth dividing into classes using 3 or 4 morphometric variables as input.

Another situation in which a measure of statistical significance would be needed in determining the differences between two rasters with qualitative data is the processing of maps by means of filtering. The key question is: After how many filters do we get an image that is too smooth? Fig.2 shows the pattern and its modification by carrying out the filtering procedure 5 times. The agreement between the two images is 92.6% (explained later in the article). The parameters set during filtering (the size of the filtering window and the filtering method), the number of filters and the homogeneity of the image are also important. The smaller information noise, the greater the differences between the original and the final image – however, we leave these considerations for another research study.

![Fig.2. Pattern and its modification after 5-fold majority filtration (majority in 3x3 window).](image)

**Research Procedure**

Map comparison procedures are derived from four spatial data analysis traditions:

1. An accuracy rating that characterizes a match (or mismatch) between a reference map considered to be accurate and one or more of its approximations.
2. Detection of changes that are interpreted as a function of time.
3. Model comparison, where the (predicted or simulated) model results are compared with observed data and/or with other model results (pixel-to-pixel consistency is not expected here).
4. The last tradition is to compare the landscape. The key feature of this approach is that the comparison is undertaken against one or more spatial metrics computed from the maps.

In practice, there are a number of methods used to compare and measure differences between raster images. They can be divided into: objective methods - using image descriptors based on mathematical models, and subjective methods – based on observations made by expert observers.
Three methods were selected in the article: differential, using Cohen's kappa coefficient, and based on the Haralick event matrix (Co-occurrence Matrix). These methods were applied globally – to the entire area and locally – to selected fragments of areas. Different areas and adjacent areas were deliberately selected to test the effectiveness of the measures for different types of terrain.

**Global Measures**

The simplest approach to finding differences between two raster maps is to subtract one raster from the other and mark the differences. The results of such subtraction are shown in Fig.3, where the differences between the pattern and Model1 are marked in green, and the differences between the pattern and Model2 in yellow. The overall agreement was 62% for Model1 and 76% for Model2, respectively (Table 1). Setting a threshold for the allowable number of differing pixels is a half-way solution, because there is still the question of distributing these differences in a spatial context.

### Table 1. Value of differences between the models and the standard based on Cohen's kappa coefficient

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen's kappa agreement coefficient</td>
<td>agreement</td>
<td>Cohen's kappa agreement coefficient</td>
<td>agreement</td>
<td></td>
</tr>
<tr>
<td>0.71</td>
<td>62%</td>
<td>0.82</td>
<td>76%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Values of textural properties for Pattern, Model 1 and Model 2

<table>
<thead>
<tr>
<th>Property</th>
<th>Model 1</th>
<th>Standard</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular Second Moment</td>
<td>0.1913</td>
<td>0.1895</td>
<td>0.1864</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.4885</td>
<td>0.3595</td>
<td>0.3797</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.8067</td>
<td>0.8563</td>
<td>0.8496</td>
</tr>
<tr>
<td>Sum of Squares</td>
<td>1.2638</td>
<td>1.2513</td>
<td>1.262</td>
</tr>
<tr>
<td>Inverse Difference Moment</td>
<td>0.9036</td>
<td>0.9104</td>
<td>0.9073</td>
</tr>
<tr>
<td>Sum Average</td>
<td>2.9844</td>
<td>2.9951</td>
<td>3.1949</td>
</tr>
<tr>
<td>Sum Variance</td>
<td>4.5665</td>
<td>4.6457</td>
<td>4.6683</td>
</tr>
<tr>
<td>Sum Entropy</td>
<td>2.3805</td>
<td>2.4063</td>
<td>2.421</td>
</tr>
<tr>
<td>Entropy</td>
<td>2.7962</td>
<td>2.7874</td>
<td>2.8102</td>
</tr>
<tr>
<td>Difference Variance</td>
<td>0.1228</td>
<td>0.1228</td>
<td>0.121</td>
</tr>
<tr>
<td>Difference Entropy</td>
<td>0.8035</td>
<td>0.7724</td>
<td>0.7928</td>
</tr>
<tr>
<td>Information Measures of Correlation I</td>
<td>-0.5898</td>
<td>-0.5966</td>
<td>-0.5861</td>
</tr>
<tr>
<td>Information Measures of Correlation II</td>
<td>0.9498</td>
<td>0.9513</td>
<td>0.9492</td>
</tr>
</tbody>
</table>

The second proposed global measure is Cohen's Kappa Coefficient [44] – determined according to the equation [1]. This coefficient takes values from -1 to 1. The closer the value is to 1, the greater the agreement between the model and the standard. For the study area, Cohen's kappa coefficient was 0.71 for Model1 and 0.82 for Model2 (Table 1). Based on these two global measures, Model2 is closer to the benchmark, but is this close enough to conclude that the maps are consistent at a certain level of statistical significance?

$$K = \frac{\sum_{i=1}^{n} \frac{c_{ji}}{e_{ji}} - \sum_{i=1}^{n} \frac{c_{ji}e_{i}}{e_{ji}}}{\sum_{i=1}^{n} \frac{e_{ji}}{e_{ji}} - \sum_{i=1}^{n} \frac{c_{ji}e_{i}}{e_{ji}}}$$

[1]
where $c_{ij}$ – values in the $i$-row and $j$-column, dots mean summing up e.g. $c_i$ – summing up over all columns in row $i$, $c$ – total sum, $n$ – number of classes.

The third method used to determine the differences was the Cartesian distance from the textural properties according to Haralick [45], determined on the basis of the Co-occurrence Matrix. Table 2 contains the values of individual properties for the pattern and models.

**Table 3.** The value of differences (taking into account the classes) between the models and the standard based on the Cartesian distance from the textural properties according to Haralick after averaging for angles

<table>
<thead>
<tr>
<th>Model</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model1</td>
<td>0.173</td>
<td>0.057</td>
<td>0.189</td>
<td>0.071</td>
</tr>
<tr>
<td>Model2</td>
<td>0.115</td>
<td>0.003</td>
<td>0.040</td>
<td>0.167</td>
</tr>
<tr>
<td>The best fitting</td>
<td>model 2</td>
<td>model 2</td>
<td>model 2</td>
<td>model 1</td>
</tr>
</tbody>
</table>

Fig. 3. The concept of locality in similarity measures can be understood as a similarity within subclusters or a measure of global similarity calculated for a smaller area (subregion - $R_1, R_2$).

Fig. 3 shows that greater differences occur in the western part of the study area ($R_1$ rectangle). Difference pixels are larger clusters and there are more of them. In the eastern part of the area, the differences are minor and do not constitute compact patches ($R_2$ rectangle). This experiment shows that differences (Table 3) can be distributed very differently in space, so one global measure of this nature is not sufficient. In addition to the spatial distribution of differences, there is also the issue of their share in a given class, which was already indicated in their research by Pontius et al. [46] and Boots and Csillag [42]. Hence, a local approach seems to be more promising.

**Local Measures**

In order to determine local measures, 2 sets of sub-areas were designated. One resulted from the simple division of the entire area into equal adjacent sub-areas (Fig.4). The second method of division was to designate two sub-areas that differed according to the field expert (Fig.5).
Therefore, the determination of local measures consisted in determining global measures for 9 modules (tiles being sub-areas of the entire study area) and for two selected frames, for which the largest and smallest differences between the Pattern and Models 1 and 2 were recorded (Table 4). It is worth noting that the Haralick distance does not indicate Model 2 as closer to the standard for all sub-areas, which is the case for Cohen's kappa coefficient. For Cohen's kappa coefficient, a value of 0.7 was adopted as the limit for a satisfactory fit. Although model 2 has a better match to the pattern everywhere, the critical value of 0.7 was exceeded only for 4 out of 9 sub-regions.
Table 4. Summary of differences based on Haralick distance and Cohen's Kappa coefficient for Region 1 (R1) and Region 2 (R2)

<table>
<thead>
<tr>
<th>Module</th>
<th>Distances according to Haralick</th>
<th>Cohen's Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>1</td>
<td>0.54</td>
<td>0.70</td>
</tr>
<tr>
<td>2</td>
<td>1.48</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>1.66</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>1.17</td>
</tr>
<tr>
<td>5</td>
<td>0.46</td>
<td>0.29</td>
</tr>
<tr>
<td>6</td>
<td>1.47</td>
<td>0.30</td>
</tr>
<tr>
<td>7</td>
<td>0.32</td>
<td>0.87</td>
</tr>
<tr>
<td>8</td>
<td>1.40</td>
<td>0.42</td>
</tr>
<tr>
<td>9</td>
<td>0.64</td>
<td>0.13</td>
</tr>
<tr>
<td>R1</td>
<td>1.01</td>
<td>0.84</td>
</tr>
<tr>
<td>R2</td>
<td>0.92</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Conclusions
The number of pixels in which two raster maps differ is very specific information, but not sufficient to consider two maps similar or not, because there is no information about the spatial distribution of these differences. The presented measures and indicators are an attempt to develop a tool to assess the significance of differences between qualitative raster maps in the quantitative and spatial context. Even with expert knowledge, it is difficult to indicate the boundary between similar and dissimilar depiction of a given phenomenon. None of the presented measures is sufficient to unequivocally indicate the critical value allowing two maps to be considered similar. The experiment indicates that you should look for a combination of different indicators. The second approach is to determine similarity measures for subregions and then combine them. However, at this stage it is difficult to indicate the optimal number of subregions. The question remains how to combine the obtained measures for subregions into one summary indicator, which requires further research.

References


