Environmental Modelling & Software 26 (2011) 1150–1157

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Uncertainty in classification and delineation of landscapes: A probabilistic approach to landscape modeling

Geir-Harald Strand*

Norwegian Forest and Landscape Institute, PO Box 115, N-1431 Ås, Norway

A R T I C L E I N F O

Article history:
Received 16 February 2010
Received in revised form
20 January 2011
Accepted 18 March 2011
Available online 15 April 2011

Keywords:
Landscape model
Landscape character
Landscape region
Logistic regression
Probability map
Uncertainty

A B S T R A C T

A landscape region can be drawn on a map as a geographic feature with distinct boundaries. Reality is, however, that the change from one landscape to another usually is gradual and that landscapes therefore have uncertain or undetermined boundaries. A thematic map of landscape regions is therefore a too simple model of the landscape. An alternative approach is to consider landscape categories as purely theoretical concepts. With this perspective, a particular geographical location can be more or less affiliated with a number of different landscape categories. Such a conception of landscape does not lead to a traditional thematic map of uniform, non-overlapping regions, but to a landscape model composed of multiple overlapping probability surfaces. This article shows how such a landscape model can be established using binary logistic regression. The method is tested and the result is assessed against an existing landscape map of Norway much used in policy impact analysis in this country. The overall objective is to develop a data driven landscape model that can supplement, elucidate and for some purposes maybe even replace, the qualitative landscape description represented by the traditional landscape map.

1. Introduction

The objective of landscape mapping has conventionally been to delineate internally homogeneous areas with a distinct and uniform landscape character. This approach is common in thematic mapping and presumes that crisp spatial entities can be defined in a meaningful way and require strict adherence to well-defined, formal rules in order to ensure that independent producers of the same map will arrive at fairly identical results (eg Baily, 2005). Still, the definition and delineation of the individual classes often remain vague. The presumptions about categorical classes with crisp spatial boundaries have therefore also been challenged (Ahlqvist and Shortridge, 2010; Fisher, 2000).

Classification and delineation of crisp regions with well-defined boundaries is a customary practice in landscape mapping. This is demonstrated by the 51 national and regional European examples of Landscape Character Assessment examined by Wascher et al. (2005). Still, the term Landscape Character Assessment (Wascher, 2005) does signify a change of methodology away from from the interpretative towards more analytical landscape mapping. Landscape character is defined as "a distinct, recognisable and consistent pattern of elements in the landscape that makes one landscape different from another" (CASNH, 2001). This definition emphasizes explicit recognition of the individual elements that constitute the landscape. These elements can be used as a basis for a formal and systematic approach to landscape mapping (Mücher et al., 2010) subsequently allowing places to be compared in terms of their landscape character (Galatowitsch et al., 2009).

Today, the availability of large and diverse spatial databases provides a new platform for landscape characterization. The individual elements that compose the landscape character (climate, geology, vegetation, hydrology, artificial structures etc) can be represented as descriptive attributes attached to the spatial landscape units. The original spatial units of these databases can also form building blocks in order to construct landscape regions or similar spatial entities using Geographic Information Systems (GIS) and related technology (Metzger et al., 2005).

Progress in image analysis has furthermore brought forward segmentation techniques that expand the toolkit available for construction of landscape regions through clustering of adjacent spatial units with similar characteristics in an automatic but still meaningful way. The recently completed hierarchical European Landscape Classification (LANMAP) by Mücher et al. (2010) is an example. Although limited to using biophysical data, due to the lack of consistent and European-wide socio-cultural and historical data at the appropriate scale, LANMAP clearly demonstrates the advantages

* Tel.: +47 64949699; fax: +47 64948001.
E-mail address: ghs@skogoglandskap.no.

1364-8152/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2011.03.005

and potential of a formal and parameterized (ie defined by a set of defining attributes) approach to landscape characterization.

The parameterization of landscape mapping implied in the concept of Landscape Character Assessment is accompanied by an increased acknowledgement of the vagueness of landscape classification and landscape delineation. This recognition of vagueness was initially linked to the geometry and described as geographical objects with indeterminate or uncertain boundaries (Burrough, 1996). Later development has seen an increasing use of semantic approaches involving fuzzy set theory (Fisher, 1996; Fisher et al., 2006) and probabilistic models (Mladenoff et al., 1999; Osborne et al., 2001) in order to quantify and communicate uncertainty in the thematic classification as well.

Semantic and probabilistic approaches are to some extent competing approaches to uncertainty in geographical analysis (Kosko, 1990). From the viewpoint of semantic uncertainty (Shi, 2010), a parameterized regional partition (eg a landscape map) can be represented as a collection of rough fuzzy membership functions (Ahlqvist, 2005) and the semantic similarity between pairs of categorical data classes (eg landscape classes) can be characterized as a continuous numeric variable (Ahlqvist and Shortridge, 2010). The starting point of the statistical approach is the same parameterized regional partition but is using probability theory and statistical methods to explore similarities and differences between classes, examine the probability of class membership for individual locations and compare locations in terms of their characteristics (Shi, 2010). It is the probabilistic approach that is used in the present paper.

The Norwegian *Reference System for Landscapes* (Puschmann, 1998, 2005; Groom, 2005; Wascher et al., 2005) is a well established hierarchical classification system for the Norwegian landscape. The implementation of the system was built upon existing scientific expertise reflecting a long history of Norwegian landscape science, but there is no formal description of the interpretative method that was employed. The resulting landscape map and the related classification system are therefore to some extent elusive. With respect to recent development in Landscape Character Assessment, the Norwegian *Reference System for Landscapes* therefore could do with both parameterization and quantification of the uncertainty entailed by the system.

The present paper explores the spatial aspect of uncertain classification and delineation in the Norwegian *Reference System for Landscapes* by employing binary logistic regression. This statistical method represents a probabilistic approach to spatial uncertainty. The method requires a parameterization of the classes found in the nomenclature of the reference system. This is discussed in Section 2. The method itself is described in Section 3. Results are presented in Section 4 and discussed in Section 5. A summary with conclusions is found in the final Section 6.

2. Material

The reference data set used in this study was the set of ten "*Farming Landscape Regions*" (Hereafter: FLRs) defined as the uppermost hierarchical level in the Norwegian *Reference System for Landscapes* (Puschmann et al., 1999). The FLRs are listed in Table 1. The FLRs are technically a set of ten uniform, non-overlapping regions that constitute a complete national coverage where every location in Norway is assigned to one and only one FLR. The classification system is based on the interpretation of six criteria. These criteria are landform, geology, drainage structure, vegetation, agriculture and technical constructions. The analysts carried out manual interpretation following a qualitative description of the interpretative method. Existing thematic maps, photographs and frequent field visits were also employed.

The standard Norwegian statistical grid SSB5KM (Strand and Bloch, 2009) with quadratic grid cells of 25 km^2 (5 × 5 km) was used as the spatial framework for the study. SSB5KM covers Norway and adjacent sea areas and consists of 19,455 grid cells. The grid is a suitable starting point for modeling because the spatial units are uniform with respect to size and shape. SSB5KM was also used for cartographic presentation of the results, as shown in Fig. 2 and a number of the following figures.

Table 1
The ten Farming Landscape Regions (FLRs) found in Norway (Puschmann et al., 1999).

1	Southern coastal landscapes
2	Lowland landscapes
3	Southern forest landscapes
4	Valley landscapes
5	Southern mountain landscapes
6	Southern fiord landscapes
7	Northern forest landscapes
8	Northern fiord landscapes
9	Northern coastal landscapes
10	Northern mountain landscapes

In order to represent the FLRs in grid format, a vector y_i {i|1…10} was attached to each grid cell. Element y_{ij} for grid cell j was set to 1 if FLR i was present in the grid cell, otherwise to 0. Since SSB5KM and FLRs are independent spatial datasets, their boundaries rarely coincide and a grid cell in SSB5KM may therefore intersect more than one FLR. A single FLR was present in 11,685 grid cells while two FLRs were present in 5490 grid cells. 260 grid cells intersected three FLRs and three grid cells intersected four FLRs. In addition 2017 grid cells did not intersect any FLR. These were grid cells in the ocean, outside the limits of the landscape data set. The binary matrix y_{ij} {i|1…10, j|1…19,455} represents the dependent variables in the statistical model of the landscape regions.

Sixteen independent or explanatory variables were also used in the model. These were purposively selected in order to represent as closely as possible the factors forming the basis for the delineation of the FLRs. The variables were compiled from a number of sources and represented as a vector x_i {i|1…16} attached to each grid cell in SSB5KM. The matrix x_{ij} {i|1…16, j|1…19,455} represents the independent variables in the statistical model of the landscape regions.

Location was represented as x_1 (northing). Northing was measured as the coordinate of the center of the grid cell in UTM-33/WGS84(EUREF89) rounded to the nearest kilometer. Easting was not used in the model due to Norway's placement diagonally across several longitudes, but replaced by a measurement of the distance to the coast (described below).

Topography was captured with two variables. Maximum elevation (meters) above sea level (x_2) and relief (x_3) measured as difference (in meters) between highest and lowest elevation above sea level. Both datasets were obtained from the national digital elevation model interpolated to a 100 × 100 m grid from 20 m contour lines. Maximum elevation and relief for each grid cell were computed from the set of points from the elevation model falling inside each SSB5KM grid cell.

Distance to the coast x_4 was measured in kilometer from the center of each grid cell to the nearest point on the Norwegian coastline (including islands). For grid cells containing sea area, the distance to the coastline was set to 0 irrespective of whether the center point fell on land or not. The reference data set was the digital version of the national topographic map (scale 1:50,000).

Different aspects of land cover and land use were captured by variables x_5–x_9. These variables were compiled from the digital land resource map AR5. AR5 is a national land resource map for cartography at scale 1:5000 (Bjørdal and Bjørkelo, 2006) included in the national spatial data infrastructure. The data set contains information about land cover, land use and land capability and is maintained and updated locally by the municipal administrations. (A viewer is available at http://kilden.skogoglandskap.no).

The variables obtained from AR5 were (x_5) agricultural land (x_6) infield pasture (x_7) forest land (x_8) inland water bodies (below the tree line) and (x_9) area above the tree line. For grid cells located on the border between Norway and a neighboring country, the fraction was computed as the fraction of the part of the grid cell falling on Norwegian territory.

Human impact was captured by (x_{10}) total ground area of buildings, measured in square meters and (x_{11}) total population of each grid cell. These data were obtained from the national register of buildings (GAB) and from national census data provided by Statistics Norway.

The six remaining variables described the farming systems. These were percentage of the agricultural area used for production of cereal (x_{12}) and grass (x_{13}). The livestock was represented by number of cattle (x_{14}) and sheep (x_{15}). Finally, the number of summer farms (x_{16}) in each grid cell was recorded. Data was obtained from the national register of agricultural subsidies and from the national register of buildings (GAB).

3. Method

The approach used in this study distinguishes between the "landscape regions" actually mapped in the reference system, and the abstract "landscape categories" that these landscape regions attempt to represent cartographically. The landscape category can be understood as a specific landscape character that may be present in variable degrees throughout the landscape. The landscape category can therefore be