Computers & Geosciences 50 (2013) 128–135

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

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

Uncertainty in ecosystem mapping by remote sensing

Duccio Rocchini [a,*], Giles M. Foody [b], Harini Nagendra [c,d], Carlo Ricotta [e], Madhur Anand [f], Kate S. He [g], Valerio Amici [h], Birgit Kleinschmit [i], Michael Förster [i], Sebastian Schmidtlein [j,k], Hannes Feilhauer [l], Anne Ghisla [a], Markus Metz [a], Markus Neteler [a]

[a] Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, GIS and Remote Sensing Unit, Via E. Mach 1, 38010 S. Michele all'Adige (TN), Italy
[b] School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[c] Ashoka Trust for Research in Ecology and the Environment, Royal Enclave, Srirampura, Jakkur PO, Bangalore 560064, India
[d] Center for the Study of Institutions, Population, and Environmental Change, Indiana University, 408 N. Indiana Avenue, Bloomington, IN 47408, USA
[e] Department of Environmental Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy
[f] School of Environmental Sciences, University of Guelph, Guelph, Ontario, Canada N1G 2W1
[g] Department of Biological Sciences, Murray State University, Murray, KY 42071, USA
[h] BIOCONNET, Biodiversity and Conservation Network, Department of Environmental Science "G. Sarfatti", University of Siena, Via P.A. Mattioli 4, 53100 Siena, Italy
[i] Department of Geoinformation for Environmental Planning, Technical University of Berlin, Straße des 17. Juni 145, 10623 Berlin, Germany
[j] Center for Remote Sensing of Land Surfaces, University of Bonn, Walter-Flex-Str. 3, 53113 Bonn, Germany
[k] Vegetation Geography, University of Bonn, Meckenheimer Allee 166, D-53115 Bonn, Germany
[l] Department of Geography, University of Erlangen-Nuremberg, Kochstr. 4/4, 91054 Erlangen, Germany

ARTICLE INFO

Article history:
Received 24 March 2012
Received in revised form
17 May 2012
Accepted 18 May 2012
Available online 29 May 2012

Keywords:
Ecosystem complexity
Ecosystem mapping
Fuzzy sets
Geosciences
Remote sensing
Uncertainty

ABSTRACT

The classification of remotely sensed images such as aerial photographs or satellite sensor images for deriving ecosystem-related maps (e.g., land cover, land use, vegetation, soil) is generally based on clustering of spatial entities within a spectral space. In most cases, Boolean logic is applied in order to map landscape patterns. One major concern is that this implies an ability to divide the gradual variability of the Earth's surface into a finite number of discrete non-overlapping classes, which are considered to be exhaustively defined and mutually exclusive. This type of approach is often inappropriate given the continuous nature of many ecosystem properties. Moreover, the standard data processing and image classification methods used will involve the loss of information as the continuous quantitative spectral information is degraded into a set of discrete classes. This leads to uncertainty in the products resulting from the use of remote sensing tools.

It follows that any estimated ecosystem property has an associated error and/or uncertainty of unknown magnitude, and that the statistical quantification of uncertainty should be a core part of scientific research using remote sensing. In this paper we will review recent attempts to take explicitly into account uncertainty when mapping ecosystems.

1. Introduction

Mapping and modeling the complexity of ecosystems and their changes over time is a key issue in spatial ecology and biogeography. Remote sensing has been acknowledged as one of the most powerful methods to map abiotic and biotic components of ecosystems (including land cover, land use, vegetation, soils) and estimate their changes over time.

The mechanism used to create maps based on remote sensing data is to derive classification algorithms to label pixels. Formally speaking, let $S=\{1,2,3,\ldots,n\}$ be a set of pixels; clustering algorithms seek to find out the most probable/possible partition of S. Quintana (2006) provides a detailed mathematical review of clustering algorithms. In most cases classes are derived relying on Boolean rules where classes are sharply defined and pixels are generally associated to a class on the basis of relative spectral similarity.

Regardless the method being used (raster-based or object-oriented classification), the assumptions for carrying out classification are associated with one major drawback: classes are mutually exclusive with discrete boundaries separating each other. Hence, processing and classifying images can result in a substantial loss of information, due to the degradation of continuous quantitative information into discrete classes (Foody, 2000; Palmer et al., 2002).

Many authors have attempted to produce better representations of the true complexity using improved methods for discrete boundaries, e.g., using multi-scale segmentation based

* Corresponding author.
E-mail address: ducciorocchini@gmail.com (D. Rocchini).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.05.022

on hierarchical patch dynamics (e.g., Blaschke, 2010). Nonetheless, it has been widely demonstrated in ecological studies that habitats vary in space in a continuum manner (Rocchini, 2010, for a review) and, in some cases, a standard discrete boundary provides an unrealistic representation (Foody, 1999). As an example, Fig. 1 represents an aerial photo (1 m spatial resolution) of a mountainous landscape (Monte di Mezzocorona and Valle dell'Adige, Trentino, Northern Italy, centre scene coordinates, λ 11°07′02″E, ϕ 46°13′36″N, WGS84 datum, acquisition date May 2006, Fig. 1A). For each unit (e.g., a pixel or a polygon) assigning a membership (μ) approaching 1 to each single class is unrealistic. As an example distinguishing grassland from shrubland in ecotones would be practically impossible (Fig. 1B). On the other hand, anthropic-dominated landscapes may show discrete borders among objects (e.g., fields versus roads). In this cases, depending on the scale, discrete mutually exclusive classification can be reliable (Fig. 1C). Similar examples considering different habitat types can be found Wood and Foody (1989, based on lowland heaths) and Rocchini (2010, based on Mediterranean forests). Ahlqvist et al. (2003) and Comber et al. (2005) provide robust critical reviews on the matter.

Regarding the aforementioned loss of information, classification can implicitly degrade information which in turn results in uncertainty in the data and related outcomes. The uncertainty related to the classification process often remains hidden in the output maps, thus it cannot be readily accounted for during further analysis. For instance, maximum-likelihood classification of remotely sensed data simply leads to pixel membership to each class, while additional information generated during the classification process, such as posterior probabilities, could be output (Foody et al., 1992). Although some attempts exist that aim to map and preserve this uncertainty for further analyses (e.g., Ohmann and Gregory, 2002; Ohmann et al., 2011), this issue requires further attention.

According to Rocchini and Ricotta (2007) we will generally refer to uncertainty as both (i) vagueness, namely the lack of sharpness of relevant distinctions, and (ii) ambiguity, arising from conflicting distinctions (discordance, Klir and Wierman, 1999). In this paper, we will review the progress made in geosciences and ecology for taking explicitly into account uncertainty when mapping ecosystems and related environmental phenomena.

2. Uncertainty related to input data for ecosystem mapping

An accurate supervised classification of remotely sensed images requires appropriate ground reference data which are often derived from field training sites. There are many sources of uncertainty in the training stage of a supervised classification, such as class definitions, subjectivity of field data collection and the mixed pixel problem.

Since plant species represent the bulk of habitat structure (Chiarucci, 2007), training sites are often derived from plant sampling-based field surveys, for which one of the main problems lies in the definition of plant communities, an issue raised as early as 1926 by Gleason (1926). A formal definition has not been and will not be accepted globally (Chiarucci, 2007). Moreover, there is an intrinsic difficulty in judging survey completeness (Palmer et al., 2002). This is generally true for all observational sciences; geosciences are not free from such uncertainty as a result of a partial input (Henley, 2006).

There are a number of provoking papers dealing with problems in the discrimination of species in the field, including operator bias (Bacaro et al., 2009), taxonomic inflation (Isaac et al., 2004; Knapp et al., 2005) and more generally taxonomic uncertainty (Guilhaumon et al., 2008; Cayuela et al., 2011), i.e., the subjectivity of field biologists in acquiring species lists which is expected to increase error variance instead of obtaining accurate information on field data.

The effect of imperfect species reference data have been discussed mainly in relation to species distribution modeling (Foody, 2011; Rocchini et al., 2011), in which labeling accuracy together with sample size and pseudo-absence data may lead to biased models of species distribution over space. The same reasoning applies to input field data for generating ecosystem/habitats maps.

Evidence exists about the possibility that abrupt classification of vegetation types, especially at the species hierarchical level, can present misleading or even erroneous results (Schmidtlein and Sassin, 2004). This is due to the often continuous transition of the vegetation assemblages due to changes in environmental gradients (e.g., moisture) and self-organization in vegetation. Alternative approaches like ordination methods aim to extract major floristic gradients describing the variation of the assemblages as metric variables, thus still retaining the continuous

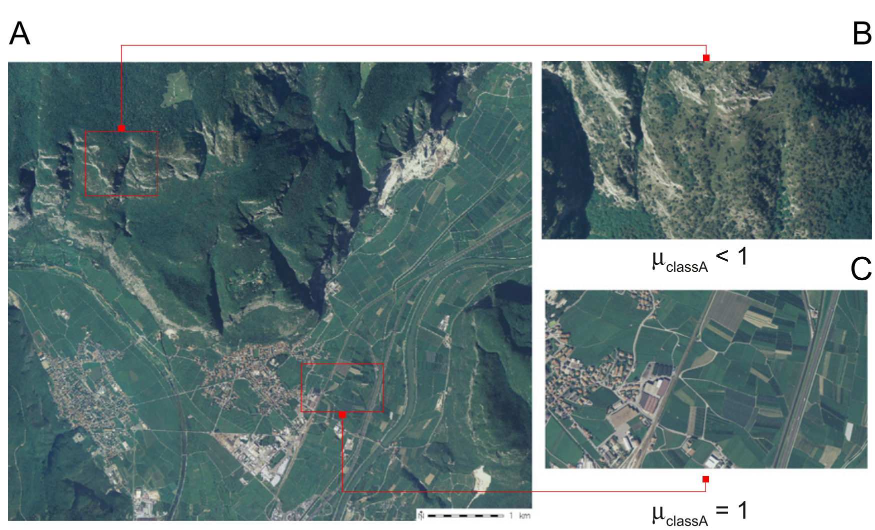

Fig. 1. An aerial photo (spatial resolution = 1 m, acquisition date: May 2006) of a mountainous landscape of Trentino (Monte Mezzocorona), Northern Italy (A). In case of high heterogeneity (B) the membership to a defined class (e.g., classA = woodland) is not exhaustive (high uncertainty). In case of a homogeneous landscape (C) the membership to a defined class (e.g., classA = crops) may be complete equaling 1. Similar examples can be found in Wood and Foody (1989) and Rocchini (2010).