Ecological Modelling 311 (2015) 73–87

Contents lists available at ScienceDirect

Ecological Modelling

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

Analysis of the socioecological structure and dynamics of the territory using a hybrid Bayesian network classifier

R.F. Ropero[a,*], P.A. Aguilera[a], R. Rumí[b]

[a] Informatics and Environment Laboratory, Department of Biology and Geology, University of Almería, Carretera de Sacramento s/n, C.P. 04120, La Cañada de San Urbano, Almería, Spain
[b] Department of Mathematics, University of Almería, Carretera de Sacramento s/n, C.P. 04120, La Cañada de San Urbano, Almería, Spain

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form 29 April 2015
Accepted 5 May 2015

Keywords:
Hierarchical classifier
Mixture of Truncated Exponential models
Probabilistic clustering
Socio ecological systems
Global environmental change

ABSTRACT

Territorial planning and management requires that the spatial structure of the socioecological sectors is adequately understood. Several classification techniques exist that have been applied to detect ecological, or socioeconomic sectors, but not simultaneously in the same model; and also, with a limited number of variables. We have developed and applied a new probabilistic methodology – based on hierarchical hybrid Bayesian network classifiers – to identify the different socioecological sectors in Andalusia, a region in southern Spain, and incorporate a scenario of change. Results show that *a priori*, the socioecological structure is highly heterogeneous, with an altitude gradient from the river basin to the mountain peaks. However, under a scenario of global environmental change this heterogeneity is lost, making the territory more vulnerable to any alteration or disturbance. The methodology applied allows dealing with complex problems, containing a large number of variables, by splitting them into several sub-problems that can be easily solved. In the case of territorial planning, each component of the territory is modelled independently before combining them into a general classifier model. Furthermore, it can be applied to any complex unsupervised classification problem with no modification to the methodology.

1. Introduction

The process of territorial planning and management requires that the spatial structure of the territory is adequately understood, particularly given the current context of global environmental change (GEC) (Basurto et al., 2013; Clark and Dickson, 2003; Hufnagl-Eichiner et al., 2011; Kotova et al., 2000; Turner et al., 2003). Spatial analysis allows the territory to be divided into a number of different units or ecological sectors (Schmitz et al., 2005), which can reflect the spatial patterns caused by ecological interactions between the elements of the territory (Jackson et al., 2012; Martín de Agar et al., 1995).

To obtain these sectors, a variety of methodologies have been applied including both subjective methods – based on expert knowledge – and objective ones, based on the data available (Chuman and Romportl, 2010; Schmitz et al., 2005; Trincsi et al., 2014; Vezeanu et al., 2010). One of the most important methodologies is classification, with recent advances promoted by the development of new technologies, such as GIS techniques and software. The most common classification methodologies are based on spatial overlapping of thematic maps and other GIS techniques (Villamagna et al., 2014), the study of satellite images (Rapinel et al., 2014) and various statistical methods, such as hard-clustering or geospatial analysis (Giménez-Casalduero et al., 2011; Liu et al., 2014; Ruiz-Labourdette et al., 2011; Trincsi et al., 2014; Vezeanu et al., 2010) to perform data analysis and ecological mapping (Lahr and Kooistra, 2010). Even though the methodologies mentioned provide robust and appropriate results, they have certain limitations, which basically relate to the amount of information the models can cope with and the rigidity of the boundaries between the different sectors identified (Niederscheider et al., 2014; Smith and Brennan, 2012). Moreover, human's role in nature is being recognized, and new tools are required that can include socioeconomic components in the same way as other components of natural systems, so configuring a socioecological system (SES) (Challies et al., 2014; Dearing et al., 2014). Thus, other methodologies that are capable of overcoming these problems need to be considered (Strand, 2011).

A novel proposal is Bayesian networks (BNs), a multivariate statistical model based on probability theory, whose ability to model environmental problems has been demonstrate over recent

Abbreviations: HBN, hybrid Bayesian network; SES, socio ecological systems; GEC, global environmental change; MTE, mixture of truncated exponential model.

* Corresponding author.
E-mail addresses: rosa.ropero@ual.es (R.F. Ropero), aguilera@ual.es (P.A. Aguilera), rrumi@ual.es (R. Rumí).

http://dx.doi.org/10.1016/j.ecolmodel.2015.05.008
0304-3800/

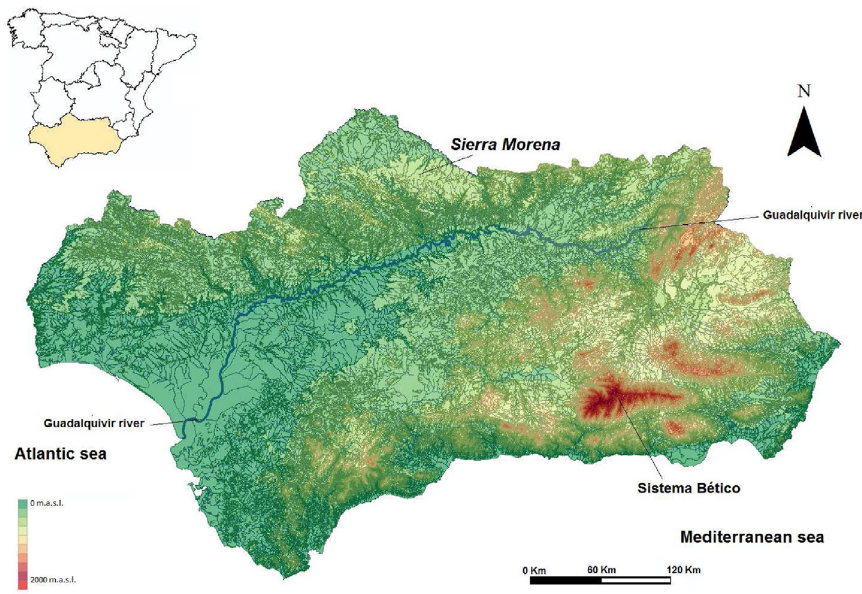

Fig. 1. Study area.

decades (Aguilera et al., 2011; Borsuk et al., 2004, 2006; Kelly et al., 2013; Langmead et al., 2009). BNs consist of a set of nodes (representing the variables of the model) connected by several links, which express relationships of statistical (in)dependence, modelled by means of probability distributions (Jensen et al., 1990; Jensen and Nielsen, 2007; Shenoy and Shafer, 1990). This makes BNs powerful and robust tools, yet their results are also easily interpreted by non-experts and stakeholders, so allowing them to be included in the model learning and validation processes (Hamilton et al., 2015; Tiller et al., 2013; Varis and Kuikka, 1999). Additionally, their probabilistic approach allows risk and uncertainty to be estimating with greater accuracy than using other models (Liu et al., 2012; Marcot, 2012; Uusitalo, 2007).

One of their most important advantages in the environmental field is that BNs can manage both continuous and discrete data in the same hybrid model, even though they were originally proposed only for discrete data (Aguilera et al., 2011; Wilson et al., 2008). In the presence of continuous variables in the data, the most common solution is to discretize them (Keshtkar et al., 2013; Renken and Mumby, 2009), which involves loss of relevant information and of precision (Uusitalo, 2007). To avoid discretization and treat continuous variables, the Conditional Gaussian model has been proposed. However, this imposes certain limitations on the structure; (i) continuous data has to follow a normal distribution, and (ii) a discrete variable cannot have a continuous parent (Lauritzen, 1992). One way to deal with hybrid BN (HBNs) models, without discretizing continuous variables and limitations in the model structure, is to use the Mixture of Truncated Exponential models (MTE) to represent the probability distributions of the variables in the HBNs. This model is able to deal with any distribution function (Moral et al., 2001). In order to avoid computational complexity problems, simpler and fixed structures have been proposed, especially for classification tasks, such as naïve Bayes (Duda et al., 2001; Friedman et al., 1997), which reduce the number of parameters to be estimated but which yield appropriate results (Fernandes et al., 2010).

A classification problem in which no information about the class variable is available (called an unsupervised classification or clustering problem) can be solved by a BN classifier (Aguilera et al., 2013; Anderberg, 1973; Fernández et al., 2014; Gieder et al., 2014). This soft-clustering methodology implies the partition of the data into groups in such a way that the observations belonging to one group are similar to each other but differ from the observations in the other groups. As BNs express the results by means of probability distribution functions, each identified group is composed of a set of different observations with a high probability of belonging to it. BNs also allow the behaviour of the system to be modelled under a scenario of change using probabilistic propagation (Aguilera et al., 2011; Liedloff and Smith, 2010).

Our objective is to develop a new methodological approach based on a HBN hierarchical classifier and apply it to characterize the socioecological structure of a territory, and study its dynamic under different drivers of GEC, in the Spanish region of Andalusia. This mathematical approach is considered hierarchical, since the model is divided into two levels of classification; in the first, both natural and socioeconomic components are modelled using independent HBN sub-models, with the aim of classifying the territory into several groups. In the second, the sub-models are joined into a classifier model that divides the territory into several socioecological sectors. Once the model is learned and the socioecological structure of the territory has been identified, a scenario of change is included. The paper is organized as follows: Section 2 describes the methodological approach used; Section 3 describe the results of both the current situation and under a GEC scenario; Section 4 discusses the results and the methodological approach is shown; finally, Section 5 draw a number of conclusions.

2. Materials and methods

2.1. Study area

Andalusia (Fig. 1) is the second largest Autonomous Region of Spain – comprising eight provinces – and the most-densely populated. It covers a surface area[1] of 87.600 km^2, which represents 17.3% of the national territory. Bounded by the Mediterranean Sea and Atlantic Ocean, Andalusia lies on the frontier between Europe and Africa and contains a mixture of landscapes and cultural heritage from both continents.

Andalusian terrain covers a wide range of altitude, from the Guadalquivir river basin to the mountainous ranges of the *Sierra*

[1] Data from the Spanish Statistical Institute.