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, Dept. of Biology and Geology, University of Almería. Carretera de Sacramento s/n, C.P. 04120, La Cañada de San Urbano, Almería, Spain.*
[b] *Dept. of Mathematics, University of Almería. Carretera de Sacramento s/n, C.P. 04120, La Cañada de San Urbano, Almería, Spain.*

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

[1] rosa.ropero@ual.es (R.F.Ropero, *corresponding author), aguilera@ual.es (P.A. Aguilera), rrumi@ual.es (R. Rumí

[2] HBN, hybrid Bayesian network; SES, Socio Ecological Systems; GEC, Global Environmental Change; MTE, Mixture of Truncated Exponential model.

problem with no modification to the methodology.

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

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 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 (Figure 1) is the second largest Autonomous Region of Spain – comprising eight provinces – and the most-densely populated. It covers a surface area[3] 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 Morena* and *Sistema*

[3]Data from the Spanish Statistical Institute

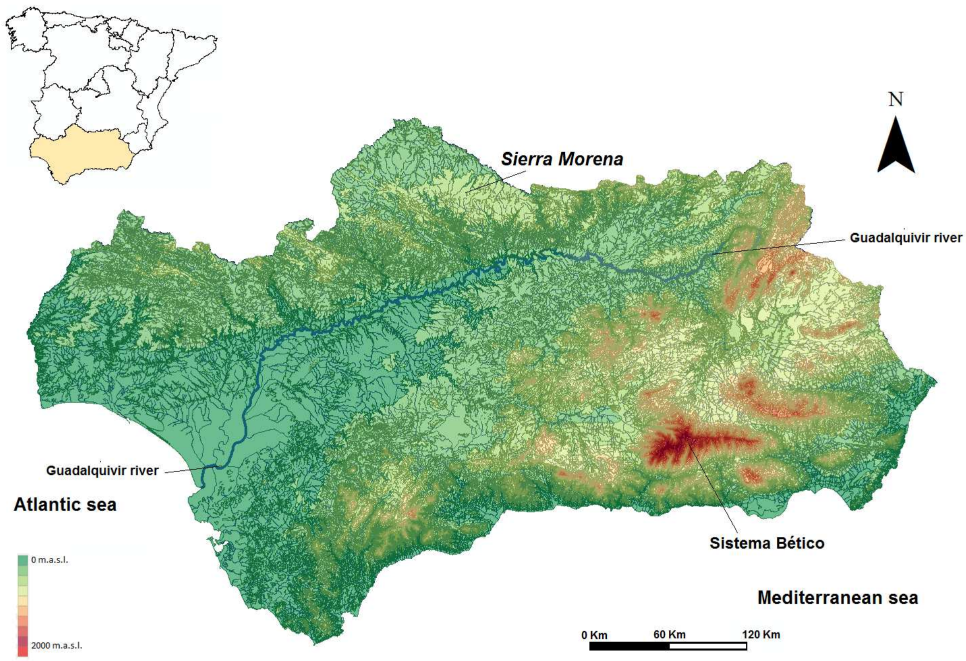


Figure 1: Study area.

Bético, which boast the highest peaks in Spain, lying above 3000 m. a.s.l. The landscape is quite heterogeneous, with huge differences between the densely populated and irrigated rich croplands areas of the river basin and coastlands, to the sparsely populated forested areas of the uplands.

Its climate is similarly heterogeneous. Even though Andalusia is included in the Mediterranean climate zone, there are stark differences between different areas. The climate in the southeast part is semiarid, with less than 200 mm of annual rainfall in several areas, whilst the middle and northern parts are under a continental climate influence, with more than 4000 mm rainfall.

2.2. Data collection

In accordance with the environmental and socioeconomic characteristics of the territory, six groups of variables were selected for the HBN hierarchical classifier model.

Environmental information (Appendix A) was collected from Andalusian

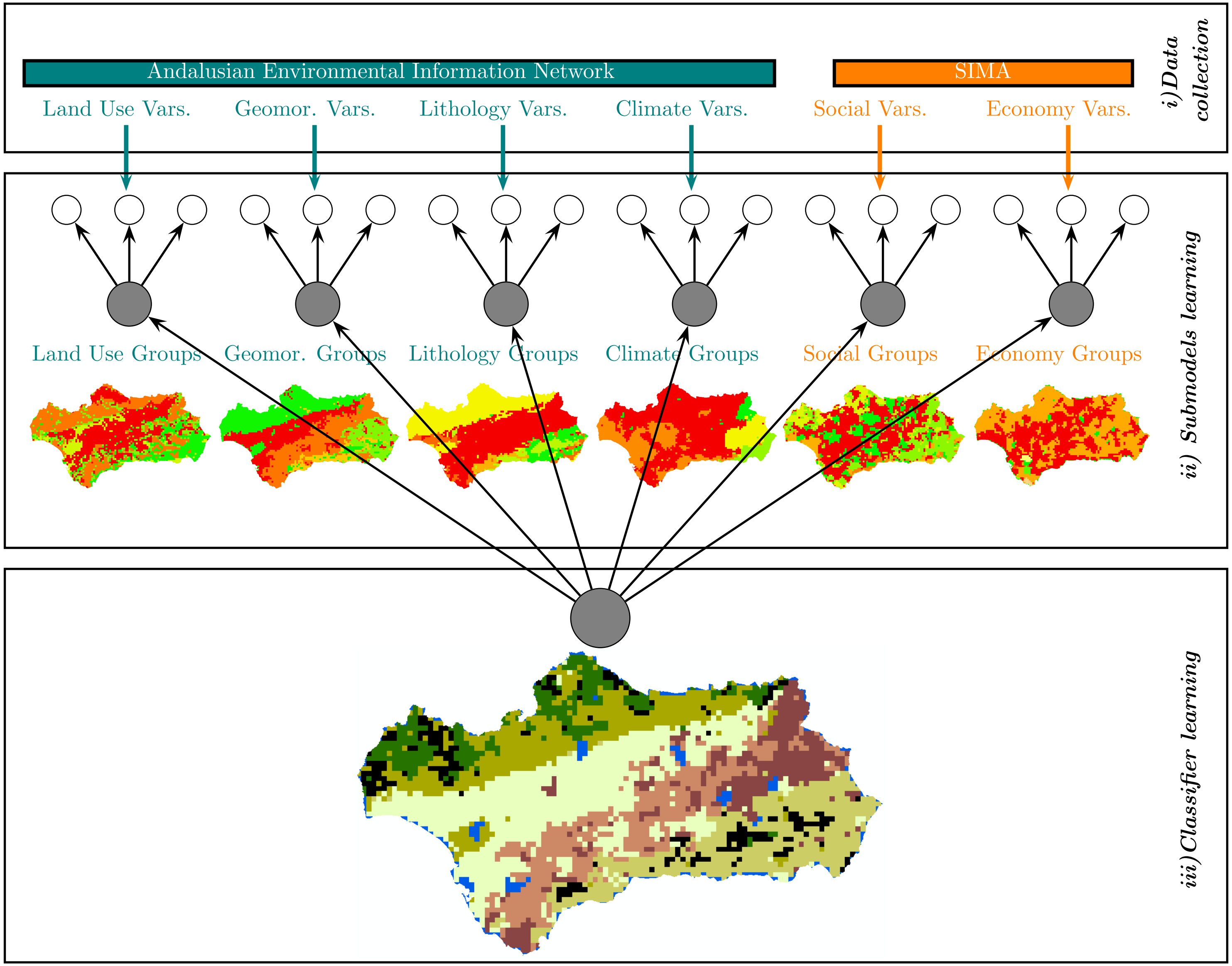


Figure 2: Methodological diagram of the hierarchical classifier model divided into three steps: i) Data collection (Subsection 2.2), ii) Submodels learning (Subsection 2.3) and iii) Meta-classifier learning (Subsection 2.4). White nodes refer to original variables (either discrete or continuous), grey nodes refer to artificial discrete class variables, which represent the membership of each observation to sub-models groups (*i.e.* Land uses groups) and classifier sectors respectively. SIMA, Andalusian Multiterritorial Information System; Vars., Variables; Geomor., Geomorphology.

Environmental Information Network[4] (Figure 2 *i)*) and divided into four different sub-models: land use, geomorphology, lithology and climate. ArcGis v10.0 (ESRI, 2006) was used to retrieve the data, using a grid of 5x5 km. Land use, geomorphology and lithology variables are expressed as the percentage of the surface area of each grid cell, whilst climatic variables are expressed as an absolute value per grid cell (see Appendix A for a detailed explanation).

The Andalusian Multiterritorial Information System [5] was searched to obtain social and economic information for each municipality to feed to the corresponding sub-models (Figure 2 *i)*). In order to obtain information that related to uniform spatial units, ArcGis v10.0 (ESRI, 2006) was used to transform the data into a 5x5 km grid by overlapping it onto the municipal information shape file. In this way two cases were found: *i)* grid cells containing only one municipality, where the information was collected; *ii)* grid cells that overlap two or more municipalities; in these cases variables were obtained as a weighted mean of each municipal values. Variables are expressed in different ways, such as rates, percentage of the municipal population, percentage surface area of the territory (see Appendix A for a detailed explanation).

Variables were selected by experts and from literature review; they were preprocessed with the aim of avoiding repeated information. The preprocessing step included the elimination of variables providing equivalent information by means of the analysis of a correlation matrix, and the selection of the appropriate level of detail in the shape file information. In addition, environmental variables comprising more than 70% of data equal to zero were discretized using the equal frequency method into three different states (0- no presence; 1- low presence; 2- high presence. Thresholds of each variable are shown in Appendix A). The final data set contained 3630 grid cells and 151 variables, both discrete and continuous.

2.3. Sub-models learning

This section describes the steps for constructing each of the six sub-models (Table 1) included in the first level of the classifier (Figure 2 *ii)*). They are based on the probabilistic clustering methodology using HBNs as

[4] http://www.juntadeandalucia.es/medioambiente/site/rediam

[5] http://www.juntadeandalucia.es/institutodeestadisticaycartografia/sima/index2-en.htm

Table 1: Sub-models characteristics. No., number; Vars., variables.

Sub-model	No. Vars.	Discrete Vars.	Continuous Vars.
Land Use	10	0	10
Geomorphology	50	48	2
Lithology	41	39	2
Climate	7	0	7
Social	18	4	14
Economy	25	15	10
Total	**151**	**106**	**45**

proposed by Fernández et al. (2014), and implemented in the Elvira software (Elvira-Consortium, 2002). Figure 3 shows an outline of this methodology. The relationships between variables cannot be expressed using a *Conditional Gaussian* model for two reasons (see Section 1): the variables in this dataset do not follow a normal distribution, and also, even though in the models developed in this paper no discrete variable has a continuous parent, if a more complex model such as the Tree Augmented Network (Friedman et al., 1997) is selected as the baseline, method then this second constraint is not fulfilled either. So, the MTE model, which avoids these limitations, is used to model the probability distributions involved in the construction of the network (For more information about MTE models see Cobb et al. (2007); Rumí and Salmerón (2007); Rumí et al. (2006)).

The corresponding sub-models have a naïve Bayes structure (Figure 4), in which the links between the feature variables ($X_1, \ldots, X_n$) and the class variable, H, express the conditional probability distribution $p(X_n \mid H)$. If new information is known about the feature variable X_n it is incorporated to the model and the conditional probability distribution of H is updated.

Taking the Land Use sub-model as an example, feature variables are expressed as the presence of different land uses types in Andalusia, collected from the 5x5 km grid, whilst the class variable expresses the membership of each individual grid cell (corresponding to each data sample) to a group with similar land use characteristics. The methodology applied consists of two steps:

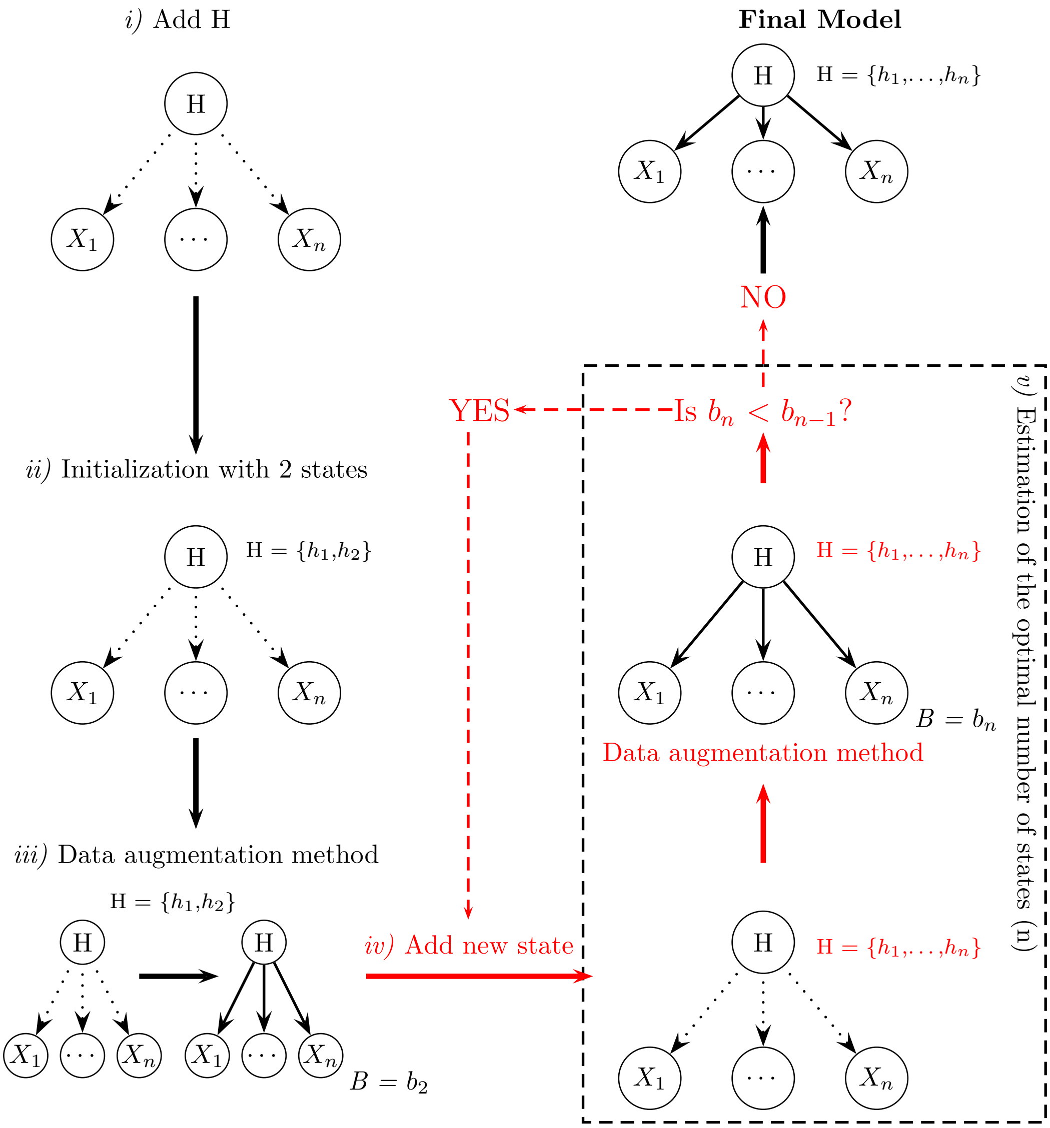


Figure 3: Outline of the HBNs probabilistic clustering methodology to construct both sub-models and the classifier. Dotted lines represent the relationships between the variables when the parameters of the probability distribution functions have not been yet estimated. B, BIC score.

1. *Estimation of the optimal number of states.* Initially, no information about the class variable is given, so we consider it as a hidden variable H, whose values are missing (Figure 3 *i)*). Firstly, we consider only two states for variable H, *i.e.*, two different land use groups that are uniformly distributed (the same probability value for each grid cell of belonging to both groups, *i.e.*, 50%) (Figure 3 *ii)*). Now, the model is estimated based on the *data augmentation* method (Tanner and Wong, 1987), an iterative procedure similar to the Expectation Maximization algorithm (Lauritzen, 1995) as follows: *a)* the values of H are simulated for each data sample according to the probability distribution of H, updated specifically for the corresponding data sample, and *b)* the parameters of the probability distribution are re-estimated according to the new simulated data. In each iteration, the BIC score of the model is computed, and the process is repeated until there is no improvement. In this way, the optimal parameters of the probability distribution function of the model with two states and its likelihood value are obtained (Figure 3 *iii)*). The following step consists of a new iterative process in which a new state (a new land use group) is included in variable H by splitting one of the existing states (Figure 3 *iv)*). The model is again re-estimated (by repeating the *data augmentation* method) and the BIC score is compared with the previous run. The process is repeated until there is no improvement in the BIC score, so achieving the final model containing the optimal number of states (Figure 3 *v)*).
2. *Computation of the probability of each grid cell belonging to each group.* Once we have obtained the final model (with the optimal number of class variable states, *i.e.*, the optimal number of land use groups), the next step consists of probability propagation, also called the *inference* process (For more information see Rumí and Salmerón (2007)). In this step, all the available information (land use variables) for each data sample is input into the model as a new value called *evidence*, and propagated through the network, updating the probability distribution of the class variable. Finally, from this new distribution the most probable land use group (state of the variable H) for each data sample, it means, for each grid cell, is achieved.

2.4. Classifier learning

Once the various sub-models are learned, the next step consists of joining them in the second level of classification in the classifier model (Figure 2

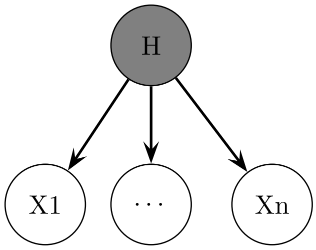


Figure 4: Example of the naïve Bayes structure. $X_1, \ldots, X_n$ are the features variables which can be both discrete or continuous; H, is the hidden discrete class variable that represents the membership of each observation to a group.

iii)). A new virtual data set is created where the feature variables are the results of the previous six sub-models (*i.e.*, the most probable land use, geomorphology, lithology, climate, social and economic group for each grid cell), whilst the hidden class variable expresses the membership of each grid cell to the socioecological sectors.

Note that, in this level, both feature and class variables are discrete, but the flexibility of the methodology proposed allows this kind of data to be dealt with in exactly the same way as in the previous step. The process is repeated, as explained in Section 2.3 and Figure 3, to obtain the final model with the optimal number of socioecological sectors. Once we know the parameters of the model, the *inference* process is carried out and the probability that a particular grid cell belongs to a particular sector is calculated; then the most probable one is represented.

2.5. Global Environmental Change Scenario

Using the final classifier model obtained, we can predict how the socioecological structure of the territory might change as a consequence of various GEC drivers through the inference or probability propagation process.

Taking the information provided by the Intergovernmental Panel on Climate Change, both national and regional governments have developed climate change scenarios for their particular territory. A number of reports and studies have been written about the impact of these scenarios on the economy, on society, and on land use and land cover (Gasca, 2014; Méndez-

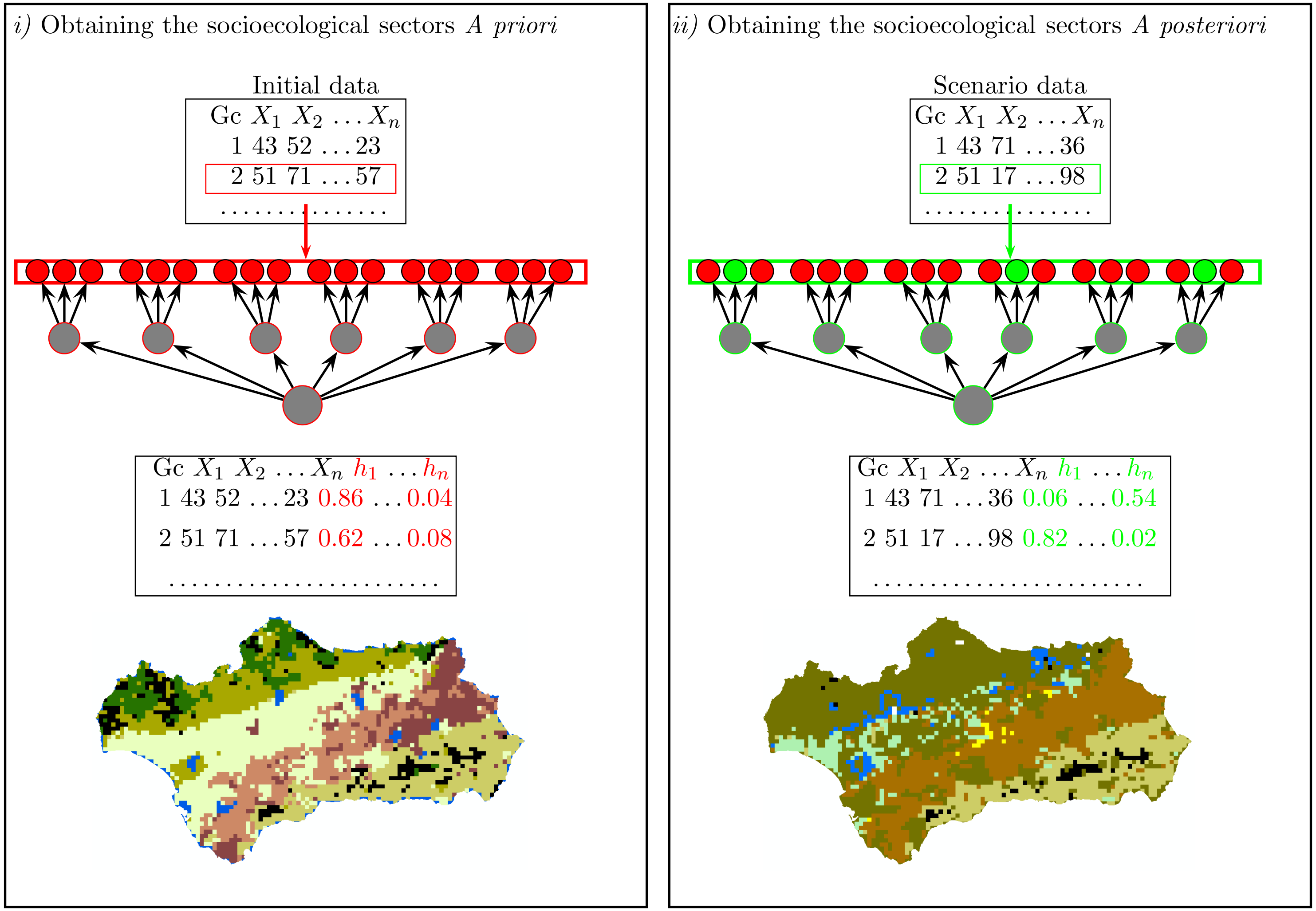


Figure 5: Methodological diagram of the Inference process. *A priori* the information about the current situation is introduced into the model and propagated to obtain the probability of each grid cell (Gc) belonging to socioecological sectors. *A posteriori*, information about drivers of GEC is collected and included - as new values or evidences - into several variables of the classifier model, and the probability values are updated.

Jiménez, 2012; Nieto and Linares, 2011). In Andalusia, two scenarios are considered: A2 and B2 (Méndez-Jiménez, 2012). The A2 scenario describes a heterogeneous world, where self reliance and preservation of local identity are key. Population increases continuously and economic development is based on national decisions (regionally oriented), whilst per capita economic growth and technological change are fragmented and slow (Gasca, 2014; Solomon et al., 2007). By contrast, the B2 scenario describes a situation in which economic development is not important and the environmental and socioeconomic problems are solved at local level. This scenario implies a slow population increase (Gasca, 2014; Solomon et al., 2007). In our study we focused on the A2 scenario - the 2040 horizon scenario for Andalusia, since we consider it closer to the current trend of socioecological change.

The information for the evidences was collected from the Assessment of the International Panel on Climate Change (Stocker et al., 2013), from national and regional reports (Gasca, 2014; Méndez-Jiménez, 2012; Nieto and Linares, 2011), and from the Andalusian Environmental Information Network.

One advantage of BNs is that it is not necessary to include information for all feature variables in order to be able to make the prediction (Ropero et al., 2014b). Rather, only new information is included as evidences in those variables in which we have knowledge about their change. In our case, evidences are included for the variables of climate, land use and economic sub-models (Table 2). Lithology and Geomorphology are consider stable. Whilst no reliable information about social changes is available, no evidences have been introduced into these variables (For a detailed explanation of the scenario of change, see Appendix B). Once the evidences are introduced, they are propagated using an inference algorithm from the sub-models to the classifier, updating the distribution of the socioecological sectors in Andalusia (Figure 5 *ii)*).

3. Results

3.1. A priori results

Figure 6 shows the socioecological structure of Andalusia in the current situation, which identifies eight different sectors. Several non-parametric hypothesis test (Chi-square for discrete variables and Kruskall-Wallis for continuous variables) were carried out to check if significant differences exist

Table 2: Variables in which new evidences are introduced under the scenario of GEC.

Sub-model	Variables	Appendix
Climate	Annual average rainfall; Annual average temperature	Appendix B.1
Land Use	Dense woodland; Irrigated cropland; Rain-fed cropland	Appendix B.2
Economy	Business Activities Tax in primary sector; Business Activities Tax in secondary sector; Business Activities Tax in tertiary sector; Tertiary sector employment; Number of rural hotels; Winter water consumption; Summer water consumption; Farming units cattles; Farming units pigs	Appendix B.3

between these sectors. Using a significance level of 0.05, the tests showed that the differences between sectors are significant.

The sectors are aligned geographically with a southwest to northeast orientation, following a gradient of increasing altitude from the *Guadalquivir* river basin to the peaks of *Sierra Morena* and *Sistema Bético* mountain ranges *Mountain peaks* sector. Figure 7 shows the box plot of certain variables, as an example of how this gradient is revealed (*i.e.* rainfed crops surface increase from the mountain peak to the *Guadalquivir* river.)

The first sector, called *Guadalquivir river* covers the river basin area, with its gentle geomorphology of rich sedimentary plains, whose climate enables an important rainfed agriculture to be practiced. This sector is the one most-affected by human activities, containing few natural areas and supporting a wealthy population with a high level of education.

In the foothills of the mountains to the north and south, there are two transitional bands of mixed cropland with forestland, subject to cooler, wetter weather. From the socioeconomic point of view, both areas have significant agricultural activity, but their wealth and structure are different: there are fewer urban areas, lower level of education, lower income per capita, and a change from agricultural areas to one with a high proportion of natural areas (Figure 7).

The northern transitional band can be differentiated into two sectors:

- *Northern transition, medium socioeconomic sector*. Located along the edge of the river basin plain, it is dedicated to agricultural activity with a slightly less wealthy population who are educated to a lower level than the *Guadalquivir* sector. This area still contains some areas of significant agricultural investment.

- *Northern transition, low socioeconomic sector.* Located on the hillslopes of the *Sierra Morena*, its landscape is woodland with some patches of rainfed crops. The main difference with the other northern transitional sector is its socioeconomic structure, which corresponds to a sparse population of poorer ageing people.

The differences between these two sectors and the river basin area are slight and gradual. By contrast, to the south, the transition band - also represented by two sectors- shows greater contrast and clearer differences to the river plain:

- *Southern transition, contrast sector.* This is characterized by a steep, eroded relief, containing contrasting areas and an important livestock activity. Close to the river *Guadalquivir*, its socioeconomy comprises a wealthier population with a high agricultural investment. At higher elevations in this sector, the population is characterized by higher migration rates and the economic variables are more depressed than in the previous one.

- *Southern transition, heterogeneous sector.* Located in the highlands of the *Sistema Bético*, this sector presents a heterogeneous landscape with significant forest cover, as well as areas with degraded natural vegetation. Croplands are fewer common than in the lower foothills and the population is characterized by ageing and abandonment areas.

Dotted around within these four zones of the northern and southern transition bands are seven patches, which belong to the *Irrigated cropland* sector. These patches have similar characteristics to the sector within they lie, but they are principally dedicated to irrigated croplands and reveal industrial, rather than agricultural, investment. They also contain a significant proportion of urban landscape. Despite this, these patches have the lowest income per capita and the lowest level of education.

At the top of the mountains are several local patches, which make up the *Mountain peaks sector*. In the *Sierra Morena* this sector appears over 400 m.a.s.l. whilst in the *Sistema Bético*, it lies above 500 m.a.s.l., so the weather is colder and rained in the last one. However, both zones contain more natural landscape (forest and scrubland) with some olive groves in the northern part. The geography of these areas comprises an elevated, steep relief, whilst its sparse and ageing population is mainly dedicated to subsistence agriculture.

Finally, the *Mediterranean coast sector* lies on the South face of the *Sistema Bético* foothills, over a mixture of sedimentary, metamorphic, volcanic and even karst materials. Its eroded relief is composed of hills, mountains and coastal plains. It is a warm sector, the driest one of Andalusia, and its heterogeneous landscape includes a high proportion of scrubland and sparse vegetation. From the socioeconomic point of view, this sector is mainly dedicated to the primary sector, though contrasts exist between medium income per capita and medium educational level to poorly developed areas. It also has an important tourism sector.

3.2. A posteriori results

Figure 8 shows the socioecological structure of Andalusia under the GEC scenario. The number of sectors have decreased to seven. As in the *a priori* situation, Chi-square and Kruskall-Wallis tests were carried out. There are significant differences between the sectors *a posteriori*.

Under this scenario of change, the socioecological structure of the territory indicates three main sectors, oriented southwest - northeast. These three sectors contain patches of the four sectors dotted within them (Figure 8). The gradient corresponding to altitude from the river to the mountain peaks is no longer observed.

The sector called *Woodland in the Sierra Morena foothills* now covers the *Sierra Morena* and part of the *Guadalquivir* river basin, as well as several patches in southern Andalusia. It is characterized by woodland and rainfed landscape on the eroded slopes of dry areas. From the socioeconomic point of view, it is a varied sector with an ageing population and a low level of education.

The next sector is called *Woodland in the Sistema Bético foothills*. It is a continuous area that runs from southwest to northeast through Andalusia, comprising woodland with patches of rainfed crops. It corresponds to areas that are depressed socioeconomically, similar to the previous sector.

Among them, some agricultural relic areas are found. They support an agricultural society with a high level of education, a positive natural increase and tourist activity. There is now the *Rainfed cropland* sector, comprising

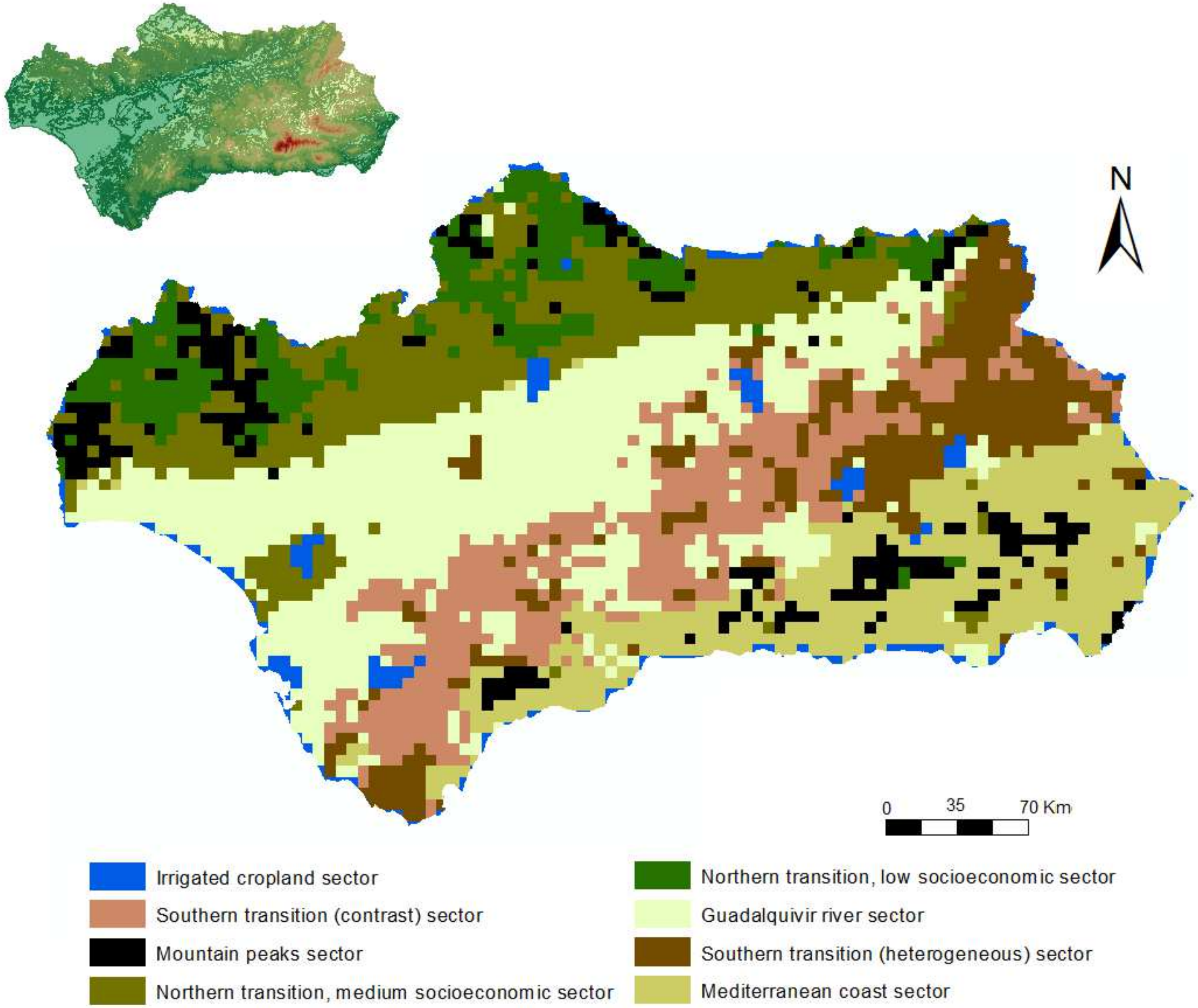


Figure 6: Socioecological sectors of Andalusia, *a priori* results.

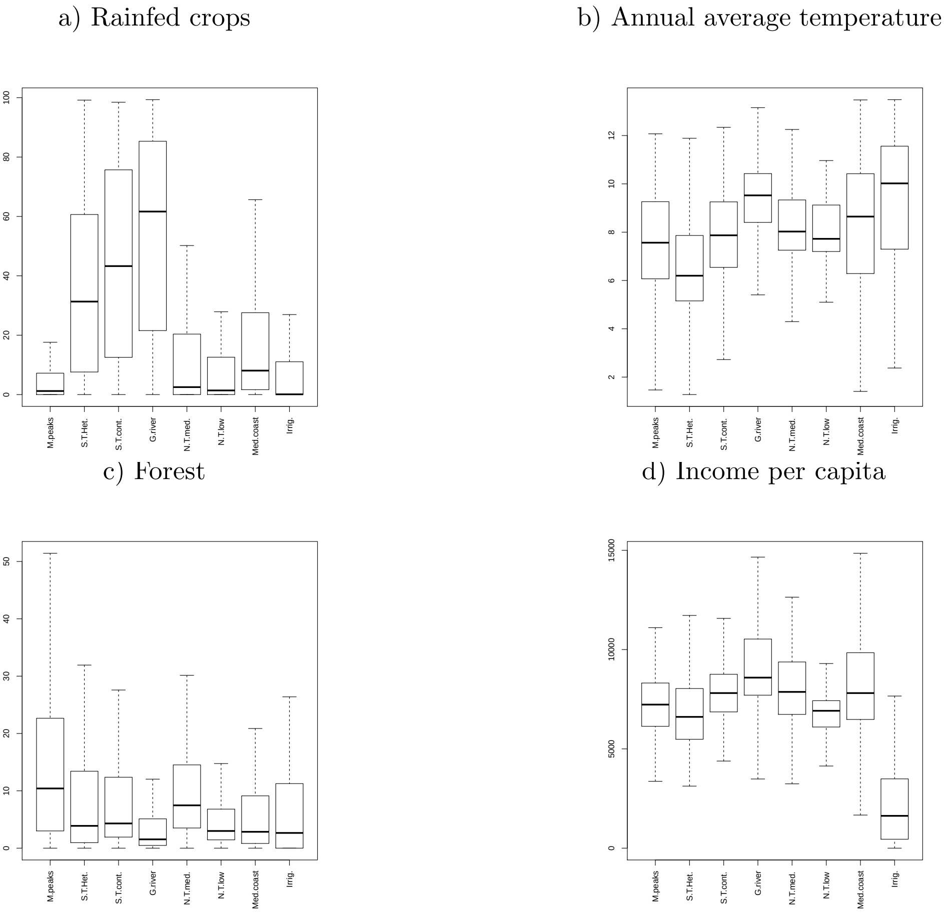


Figure 7: Extension of some land use (Rainfed crops and Forest expressed in percentage of the grid cell), climate (Annual average temperature express in Celsius) and economic (Income per capita express as a rate) variables in *a priori* sectors. M.peaks, Mountain peaks; S.T.Het, Southern transition, heterogeneous; S.T.cont., Southern transition, contrast; G.river, *Guadalquivir* river basin; N.T.med., Northern transition, medium; N.T.low, Northern transition, low; Med.coast, Mediterranean coast; Irrig., Irrigated cropland.

several patches within the river basin and the *Sistema Bético* foothills of rainfed agriculture that contains no natural landscapes. In a similar way, *Woodland-croplands* sector is composed of a number of small patches, mostly located in the river basin area, containing both natural and crop landscapes. The *Irrigated croplands* sector is composed of several patches dedicated to irrigated crops.

Lastly, two sectors are found with similar characteristics (and also the same name) as *a priori*, namely the *Mediterranean coast* and the *Mountain peaks* sectors. The *Mediterranean coast* sector covers the same area as before and supports a quite similar socioecological structure. In the same way, the landscapes belonging to the*Mountain peaks* sector are still located at the top of the mountain ranges, but they occur only in the *Sistema Bético* whilst this sector has almost disappeared in the case of *Sierra Morena* (Figure 8).

In order to study the dynamics of the structure of the territory, a confusion matrix was drawn up to highlight the differences between the *a priori* and *a posteriori* situation (Table 3). This matrix represents the percentage of each sector in the *a priori* situation that is included in each of the *a posteriori* sectors. From studying this table, it becomes clear that parts of both the northern and southern transitional areas have been incorporated into the *Woodland in the Sierra Morena foothills* and *Woodland in the Sistema Bético mountain foothills* sectors (Table 3), with corresponding change in landscape to scrubland and degraded vegetation. From the socioeconomic point of view, the diversity and heterogeneity of the transition band between the river basin and the mountain peaks has been minimized and the variables have become more homogeneous.

Whilst, in the *a priori* situation, agricultural activity extended over the river basin and both mountain foothill areas, under this scenario agricultural activity has been reduced to a number of small patches. Both *Rainfed cropland* and *Woodland-croplands* sectors replace part of the previous *Guadalquivir river* sector. However, the *Irrigated crops* sector is no longer located in the same areas as *a priori*; now these occur at higher altitude - within the *Northern transition, medium socioeconomy* (Table 3).

The *Mediterranean coast* sector, is a heterogeneous area quite similar to the *a priori* one. From the socioeconomic point of view, they have similar characteristics, but the climate under this A2 scenario is warmer and drier.

Lastly, the *Mountain peaks sector* covers the same geographical area as *a priori*, but the extent of these areas has decreased. Under the A2 scenario of change, the mountain peaks show greater presence of forest and scrublands.

Table 3: Confusion matrix showing the percentage of grid cells in common between each *a priori* and *a posteriori* sectors.

		A posteriori						
		Woodland in *Sierra Morena*	Rainfed crops	Mountainous peaks	Irrigated crops	Woodland & crops	Woodland in *Sistema Bético*	Mediterranean coast
A priori	Irrigated crops	14.6	0	1.5	0.9	0	0.2	2.05
	Southern transition (contrast)	3.4	42.8	0	0	0.5	43.4	2.4
	Mountain peaks	8.57	0	87.7	0	0	0	4.8
	Northern transition, medium	26.1	0	1.5	88.2	0	0.2	0.7
	Northern transition, low	16.6	0	0	0	0	0	0
	Guadalquivir river	25.6	57.1	1.5	10.7	99.5	21	4.3
	Southern transition (heterogeneity)	1.3	0	0	0	0	34.9	0
	Mediterranean coast	3.6	0	7.7	0	0	0.1	85.6
	Total	100	100	100	100	100	100	100

The fall in both temperature and rainfall occurs because this sector now occurs at higher altitude (in both areas, this sector is found above 600 m.a.s.l. in the *a posteriori*, whilst in *a priori* corresponded to land above 400-500 m.a.s.l.).

4. Discussion

4.1. HBNs classifier

Ecological modelling requires new methodological approaches that are capable of dealing with the heterogeneity inherent in natural systems, especially under the current framework of GEC (Challies et al., 2014). Traditional clustering techniques have been extensively applied to solve environmental problems (Giménez-Casalduero et al., 2011; Jackson et al., 2012) but in the case of detecting socioecological sectors, they would obtain poorer results (Ropero et al., 2014a). Firstly, they usually have a limit on the number of variables that can be included. The methodology proposed in this paper highlights the ability of BNs to manage datasets containing a high number of variables and observations providing robust and easy-to-interpret results due to the proposed structure. Since it is based on a hierarchical classifier - in which the problem is split into sub-problems - the model is able to deal

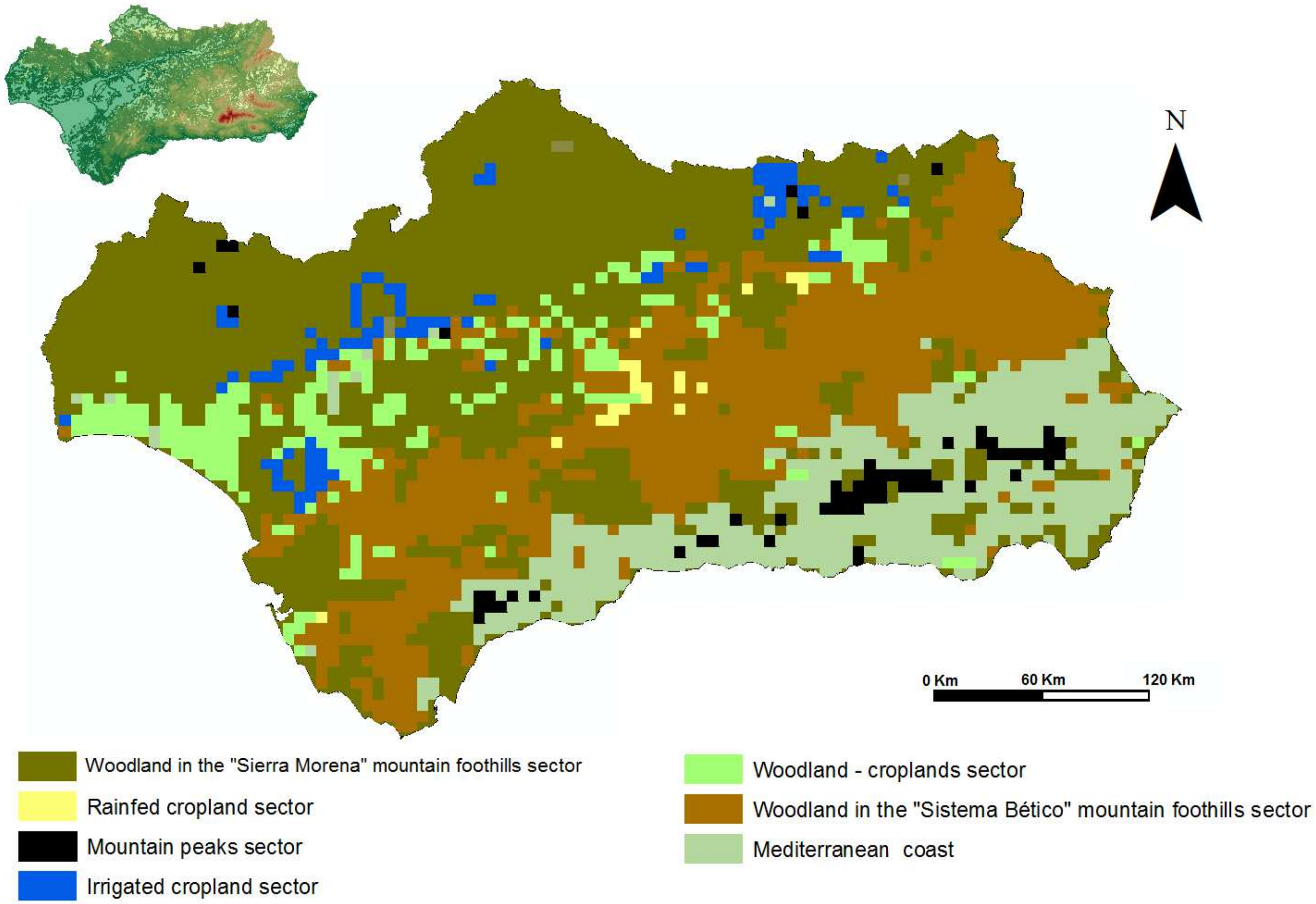


Figure 8: Socioecological sectors of Andalusia, results *a posteriori*.

with this really complex task, simplifying the problem in the manner of a divide and conquer. In addition, it allows the inclusion of new groups of variables into the final classifier, if necessary (*i.e.* include species distribution information as a new group of variables).

Secondly, the majority of the distances used in traditional *unsupervised* classification methodologies can not deal with both continuous and discrete variables in the same hybrid model (Ropero et al., 2014a). It has been demonstrated how BNs are able to deal with both discrete and continuous data, without the need to discretize the continuous variables (Ropero et al., 2014b). In this paper, the same methodology is applied, whether variables are discrete or continuous, without the need to modify the data or the methodology (Sections 2.3 and 2.4).

Finally, when data are of different magnitudes, (for example, land use variables are expressed as percentage, whilst some social variables such as age are expressed as a rate or number) some variables could have more impact on the model than the rest, and need to be standardized. Since BNs are based on probability distribution functions, they can cope with those differences without data transformation beforehand.

4.2. Socioecological structure and dynamics of the territory

Andalusia is a heterogeneous Mediterranean region, where extensive beaches lie only a short distance from high and wild mountain peaks, and where large extensions of homogeneous monocrops lie a short distance from heterogeneous subsistence crops. However, there is a clear difference between the Mediterranean coast and inland Andalusia (which are separated by the *Sistema Bético* mountain range).

Under the current situation, in inland Andalusia there is a clear separation between socioecological sectors. There is a transition from the lowland river basin to the mountain peaks, which is reflected by a gradual change from an agriculturally rich society to forestland and rural structure, with high emigration rates, illiteracy and abandonment areas. This heterogeneity implies a wide variety of ecosystems which, in turn, supports great biodiversity - Andalusia, being a Mediterranean region, is a global biodiversity hotspot (Myers et al., 2000). Inland Andalusia supports a strong economic sector, with opportunities for a huge range of economic activities (tourism, agriculture, and industry between others). However, its socioeconomy is mainly based on extensive (homogeneous) single crop farms, on which a large percentage of the population depend for their livelihood. Under the scenario of GEC, this structure is lost and the diversity and richness of the socioeconomic structure will tend to decrease.

In comparison to the *a priori* situation, changes in the environmental conditions will cause a shift in the optimal growing areas for several crop species (including olive, wheat and barley) (Méndez-Jiménez, 2012). For that reason, the agricultural diversity would be reduced to a number of relict areas and provokes the irrigated crops to shift to a higher altitudes in the *Guadalquivir* river basin area. In turn, this would provoke changes in the socioecological structure of the territory. The loss of socioecological heterogeneity would provoke a decrease in the resilience of Andalusian ecosystems (Virah-Sawmy et al., 2009), making them vulnerable to any disturbance from either natural disaster or socioeconomic and political decisions.

In contrast, in the case of the *Mediterranean coast* sector, even though the GEC scenario implies a decrease in the extent of agricultural activities, the socioeconomic characteristics would be hardly affected. This area supports an important tourist industry, apart from agriculture. Due to both increases in temperature and a longer warm season, tourism might benefit under GEC. Coastal areas would see an increase in the tertiary sector (Méndez-Jiménez, 2012). Under the A2 scenario of change, the socioeconomic heterogeneity would help to mitigate the impact on the socioecological structure of the territory and the effects of GEC would be less profound than in inland Andalusia.

As far as the *Mountain peaks* sector is concerned, our results show an increase in the surface area of forest, but further work is needed to study these areas, since climate change could provoke the extinction of the species unable to climb in altitude in the search for colder conditions (Méndez-Jiménez, 2012). On the other hand, the warmer conditions would allow an increase in population, including tourism, which might provide an opportunity in these areas to develop a sustainable touristic activity (Méndez-Jiménez, 2012).

5. Conclussions

This work presents a new methodological proposal based on HBNs hierarchical classifier and applied to identify the socioecological structure of a territory. The dynamics of the territory under a scenario of GEC was studied. The methodology proposed was able to model the heterogeneity of the territory under a probabilistic framework. The hierarchical classifier structure splits the problem into several sub-problems, in such a way that they can each be studied in detail; it is also feasible to include a new group of variables if necessary. In future work, not only would the most probable sub-model group be included in the second level of this hierarchical structure but also its probability.

Under an A2 scenario of GEC, it is demonstrated how Andalusia would tend to suffer a loss in its inherent territorial heterogeneity. This might involve important losses in environmental and social diversity, as well as a decrease in resilience that would leave the territory more vulnerable to impacts arising from political and economic decisions or natural disasters.

Even though, in this paper, this methodology has been applied to a specific case, it can be applied to any complex *unsupervised* classification problem.

Acknowledgements

This work has been supported by the Spanish Ministry of Economy and Competitiveness through project TIN2013-46638-C3-1-P, by Junta de Andalucía through project P12-TIC-2541, and from ERDF funds. R. F. Ropero is supported by the FPU research grant, AP2012-2117, funded by the Spanish Ministry of Education, Culture and Sport. We are grateful to the anonymous reviewers and the editor for their constructive and useful comments on the manuscript.

Table A.4: Variables included the **Social** Sub-Model. P/A, Presence / Absence

Variable	Type of Variable	Units	Thresholds
Rate of school attendance between 14 and17 years old	Continuous	Rate	-
Population average age	Discrete	Year	37.9; 40.9
Number of libraries	Discrete	Number per population in each municipality	P/A
Number of Cinemas	Discrete	Number per population in each municipality	P/A
Number of private schools	Continuous	Number per population in each municipality	-
Number of public schools	Continuous	Number per population in each municipality	-
Health care centres	Continuous	Number per population in each municipality	-
Number of pharmacies	Continuous	Number per population in each municipality	-
Rate of iliteracy	Continuous	Percentage of the municipal population	-
Primary studies	Continuous	Percentage of the municipal population	-
Secondary studies	Continuous	Percentage of the municipal population	-
Tertiary studies	Continuous	Percentage of the municipal population	-
National Emigration	Continuous	Percentage of the municipal population	-
Foreign Emigration	Continuous	Percentage of the municipal population	-
National Immigration	Continuous	Percentage of the municipal population	-
Foreign Immigration	Continuous	Percentage of the municipal population	-
Natural increase	Continuous	Rate	-
Total population	Discrete	Population per 25 Km^2	474.1; 1320.4

Appendix A. Variables included in the model

In this appendix variables including in each Sub-Model are shown.

Table A.5: Variables included the **Economic** Sub-Model.

Variable	Type of Variable	Units	Thresholds
Employed population	Discrete	Rate	39.9; 44.3
Internet facilities	Discrete	Number per head of population in each municipality	9.2; 12.8
Number of bank branches	Discrete	Number per head of population in each municipality	0.07; 0.09
Unemployment rate	Continuous	Percentage of the municipal population	-
Business Activities Tax in primary sector	Continuous	Rate	-
Business Activities Tax in secondary sector	Discrete	Rate	20.0; 24.4
Business Activities Tax in tertiary sector	Discrete	Rate	71.8; 78.4
Primary sector employment	Discrete	Percentage of the employed population	16.9; 27.3
Secondary sector employment	Continuous	Percentage of the employed population	-
Tertiary sector employment	Continuous	Percentage of the employed population	-
Number of agricultural cooperatives	Continuous	Percentage per municipal territory	-
Number of home owners	Discrete	Percentage of the total flats in the municipality	80.6; 86.5
Number of rented homes	Continuous	Percentage of the total flats in the municipality	-
Agricultural investment	Discrete	Percentage per municipal territory	0.44; 22.9
Industrial investment	Discrete	Percentage per municipal territory	1.6; 38.5
Investment in tertiary sector activities	Discrete	Percentage per municipal territory	0.01; 8.4
Income per capita	Continuous	Rate	-
Number of hotels	Discrete	Percentage per municipal territory	0.6; 2.1
Number of campsites	Discrete	Percentage per municipal territory	0.001; 0.08
Number of rural hotels	Discrete	Percentage per municipal territory	0.027; 0.23
Winter water consumption	Continuous	Percentage per municipal territory	-
Summer water consumption	Continuous	Percentage per municipal territory	-
Farming units bovines	Continuous	Percentage per municipal territory	-
Farming units ovines	Continuous	Percentage per municipal territory	-
Farming units goats	Continuous	Percentage per municipal territory	-
Farming units equines	Discrete	Percentage per municipal territory	6.9; 19.47
Farming units pigs	Discrete	Percentage per municipal territory	23.7; 320.5

Table A.6: Variables included the **Climate** Sub-Model.

Variable	Type of Variable	Unit
Evapotranspiration rate	Continuous	mm per year
Annual average temperature	Continuous	Celsius
Annual average rainfall	Continuous	mm
Spring number of rainfall days	Continuous	days
Winter number of rainfall days	Continuous	days
Summer average rainfall	Continuous	mm
Winter average rainfall	Continuous	mm

Table A.7: Variables included the **Land Use** Sub-Model, expressed as the percentage of the cell surface area.

Variable	Type of Variable
Heterogeneous cropland	Continuous
Landscape with scarce vegetation	Continuous
Dense Woodland	Continuous
Scrubland	Continuous
Woodland with scrub	Continuous
Woodland with herbaceous vegetation	Continuous
Human infrastructure	Continuous
Irrigated cropland	Continuous
Rainfed cropland	Continuous
Water surface	Continuous

Table A.8: Variables included in the **Lithology** Sub-Model, expressed as the percentage of the cell surface area

Variable	Type of Variable	Thresholds	Variable	Type of Variable	Thresholds
Amphibolite	Discrete	0.001; 0.078	Basic volcanic complex	Discrete	0.001; 0.069
Clay with red sand	Discrete	0.001; 0.23	Clay with marl	Discrete	0.001; 0.25
Clay with limestone	Discrete	0.001; 0.09	Clay with dolomite	Discrete	0.002; 0.17
Sand	Discrete	0.001; 0.42	Sand and marl	Discrete	0.001; 0.16
Sand and silt	Continuous	-	Silicaceous sandstone	Discrete	0.001; 0.41
Sandstone with marl	Discrete	0.001; 0.16	Calcarenite	Continuous	-
Metamorphosized limestone	Discrete	0.001; 0.14	Limestone with dolomite	Discrete	0.001; 0.22
Greywacke	Discrete	0.001; 0.07	Volcanic complex	Discrete	0.001; 0.30
Conglomerates in sand	Discrete	0.001; 0.22	Conglomerate in lutite	Discrete	0.001; 0.10
Quartzite	Discrete	0.001; 0.12	Schist and quartzite	Discrete	0.001; 0.12
Schists with gneiss	Discrete	0.001; 0.24	Phyllite	Discrete	0.001; 0.21
Grabo	Discrete	0.001; 0.07	Gneiss	Discrete	0.001; 0.13
Granite	Discrete	0.001; 0.18	Granodiorite	Discrete	0.001; 0.37
Silt with clay	Discrete	0.001; 0.48	Breccia in marl	Discrete	0.001; 0.13
Marl with limestone	Discrete	0.001; 0.20	Marl and gypsum	Discrete	0.001; 0.19
Marl with sandstone	Discrete	0.001; 0.16	Marly limestone	Discrete	0.001; 0.10
Metabasite	Discrete	0.011; 0.023	Mica schist	Discrete	0.001; 0.28
Marble	Discrete	0.001; 0.12	Peridotite	Discrete	0.001; 0.18
Calcoschist slate	Discrete	0.001; 0.19	Quartzite slate	Discrete	0.001; 0.37
Schisty slate	Discrete	0.001; 0.36	Greywacke slate	Discrete	0.001; 0.49
Volcanic complex of Cabo de Gata	Discrete	0.001; 0.69			

Table A.9: Variables included in the **Geomorphology** Sub-Model, expressed as the percentage of the cell surface area.

Variable	Type of Variable	Thresholds	Variable	Type of Variable	Thresholds
Badland	Discrete	0.001; 0.17	Gully	Discrete	0.001; 0.09
Scree	Discrete	0.001; 0.022	Structural outlier	Discrete	0.001; 0.061
Marl outlier	Discrete	0.001; 0.087	Metamorphosized outlier	Discrete	0.001; 0.077
Gypsum outlier	Discrete	0.001; 0.12	Crested hill	Discrete	0.001; 0.19
Eroded hills	Discrete	0.001; 0.14	Peripheral depression	Discrete	0.0012; 0.23
Piedmont hills	Discrete	0.001; 0.096	Structural hill	Discrete	0.001; 0.15
Conglomerate hills	Discrete	0.001; 0.067	Volcanic hill	Discrete	0.001; 0.083
Hill of intrusive rock	Discrete	0.001; 0.15	Gypsum hill	Discrete	0.001; 0.12
Dissected knoll (outlier)	Continuous	-	Alluvial fan	Discrete	0.001; 0.036
Crest	Discrete	0.001; 0.044	Cuvette	Discrete	0.001; 0.035
Conserved glacis	Discrete	0.001; 0.061	Dissected glacis	Discrete	0.001; 0.085
River bed	Discrete	0.001; 0.045	Colluvia	Discrete	0.001; 0.037
Floodplain	Discrete	0.001; 0.11	Floodplain	Discrete	0.001; 0.10
Former mudflat	Discrete	0.001; 0.33	Glacis	Discrete	0.001; 0.13
Peneplain	Discrete	0.0011; 0.37	Piedmont	Discrete	0.001; 0.045
Karstified shelf	Discrete	0.001; 0.16	Granite pluton	Discrete	0.001; 0.51
Shallow erosion surface	Discrete	0.001; 0.11	Seasonal watercourse	Discrete	0.001; 0.037
Laminated relief	Discrete	0.001; 0.38	Tabletop relief	Discrete	0.001; 0.059
Appalachian mountain chain	Discrete	0.001; 0.48	Intrusive mountain chain	DIscrete	0.001; 0.068
Metamorphic mountain chain	Discrete	0.001; 0.077	Conglomerate mountain chain	Discrete	0.001; 0.14
Marly mountain chain	Discrete	0.001; 0.10	Slate mountain chain	Continuous	-
Volcanic mountain chain	Discrete	0.001; 0.13	Scarcely dissected erosion relief	Discrete	0.001; 0.18
Moderately dissected erosión surface	Discrete	0.001; 0.21	Highly dissected erosion relief	Discrete	0.001; 0.20
Peneplanization	Discrete	0.002; 0.28	Low terrace	Discrete	0.001; 0.072
Terrace	Discrete	0.001; 0.029	Medium terrace	Discrete	0.001; 0.091

Appendix B. Information used to define the Scenario of Global Environmental Change

Information to describe the impact of several GEC drivers on different sectors of the natural and social-economic environments in Andalusia were collected from various sources: the Assessment of the International Panel on Climate Change (Stocker et al., 2013), national and regional reports (Gasca, 2014; Méndez-Jiménez, 2012; Nieto and Linares, 2011), and from the Andalusian Environmental Information Network. Due to the high heterogeneity of the Andalusian relief, the impact of the GEC scenario varied between different areas. This appendix explains these changes in detail.

Appendix B.1. Climate change

Climate change is one of the most important and commonly studied natural drivers modelled under different perspectives and methodologies (Keenan et al., 2011; Rubidge et al., 2011; Quisthoudt et al., 2013). Its interactions with land use provoke changes in the structure of both natural and socioeconomic components through different agents (Anderson-Teixeira et al., 2013; Claesson and Nycander, 2013). In Andalusia, the A2 scenario implies an increase in temperature (of up to 4 degrees in some locations), and changes in rainfall distribution (Figure B.9). Data about the predicted value of both temperature and rainfall variables for each grid cell can be obtained from the Andalusian Environmental Information Network. These data were included as evidences in the Climate sub-model variables; Annual average rainfall and Annual average temperature.

Appendix B.2. Land use changes

The pattern of land uses supports ecosystems and societies due to the fact that any alteration of land use leads to changes in biodiversity, primary production, alterations in soil productivity and the capacity to provide ecosystem services to societies (Lambin et al., 2001). In Spain, several reports based on information from the International Panel on Climate Change have been written to describe the expected change in land uses. Our study used information from the 2040 scenario of land use change (Nieto and Linares, 2011; Méndez-Jiménez, 2012). The expected changes include several that relate to the distribution of vegetation, both crops and forest species. Figure B.10 shows the percentage presence of certain species under the current situation

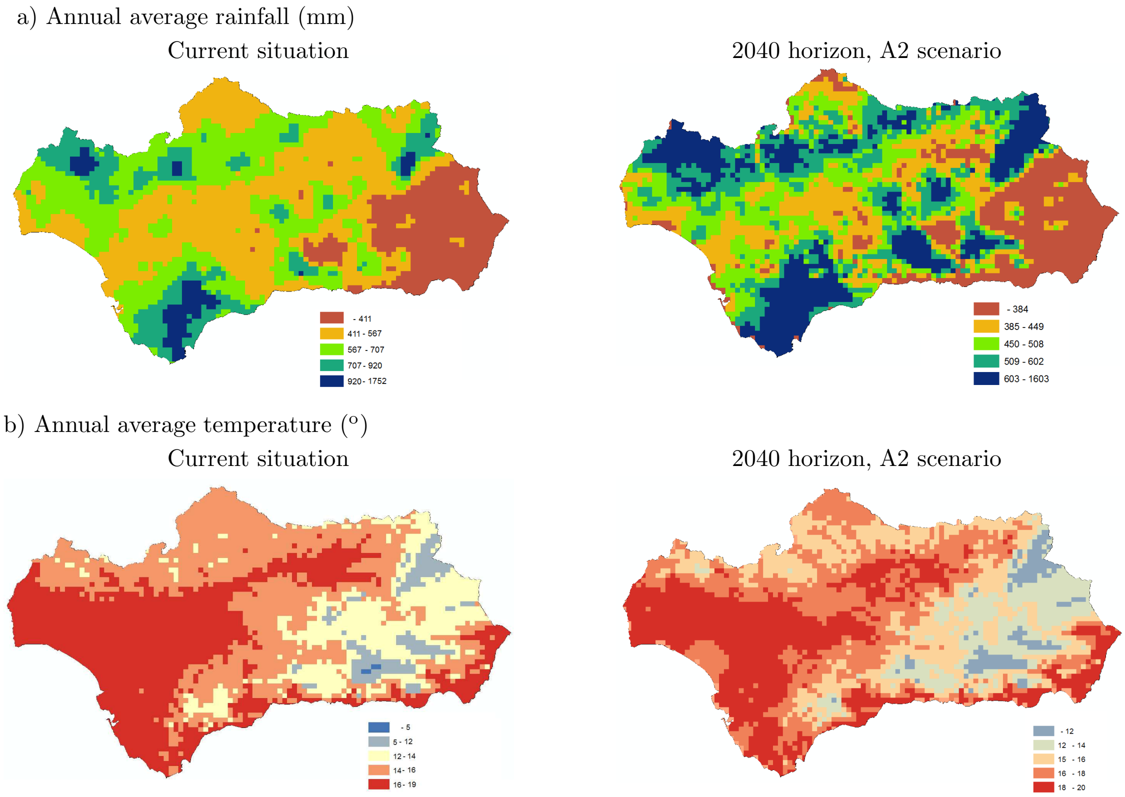


Figure B.9: Comparison between annual average rainfall (a) and temperature (b) in the current situation, and under the 2040 horizon A2 scenario of change.

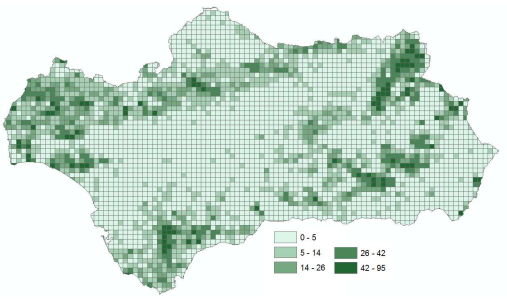
and under the GEC scenario. Information was collected from regional reports (Méndez-Jiménez, 2012) and processed by ArcGIS to transform it into 5x5 grid information. These new values were included into the model as evidences in the following Land use sub-model variables: Dense woodland, Irrigated cropland, and Rainfed cropland.

Appendix B.3. Economic change

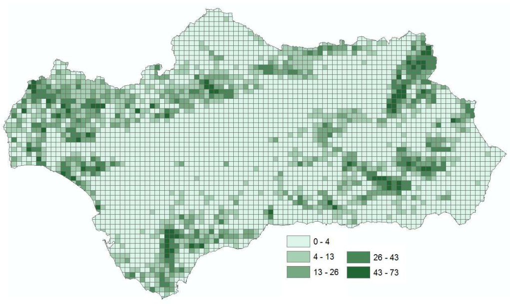
SES are dynamic systems including several socioeconomic drivers that affect ecosystems; at the same time, they contain natural drivers affecting societies in an iterative process (Cadenasso et al., 2006; Haberl et al., 2006). Due to the alteration of natural conditions, several changes are expected in the economic and social component of the SES. No reliable information was found about changes in social variables, but economic changes were identified. Two economic sectors are important in Andalusia. The first is the primary sector (livestock and agriculture). Modifications in this sector are reflected in the Land use sub-model (as changes to the extent of Rainfed crops and Irrigated crops variables). The second is the Tourism sector, which could be affected in the future if climate and weather conditions change. Information was collected from regional reports (Méndez-Jiménez, 2012) and introduced as evidences in the following variables: Business activities tax in primary, secondary and tertiary sectors, tertiary sector employment, number of rural hotels, winter and summer water consumption, and farming units cattle and pigs. Figure B.11 shows modifications of some of these variables as an example.

a) Dense woodland (percentage of the grid cell)

Current situation

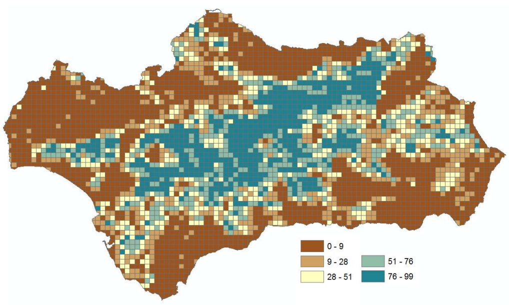


2040 horizon, A2 scenario



b) Rainfed cropland (percentage of the grid cell)

Current situation



2040 horizon, A2 scenario

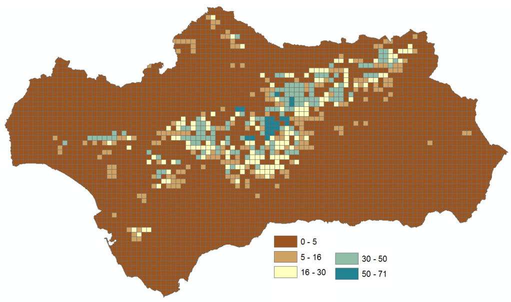
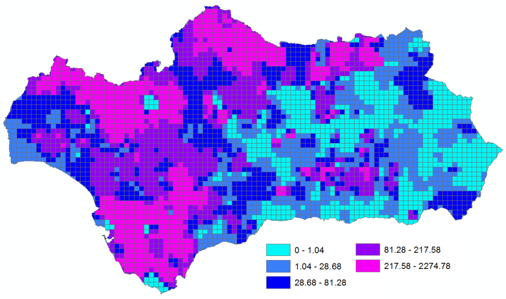


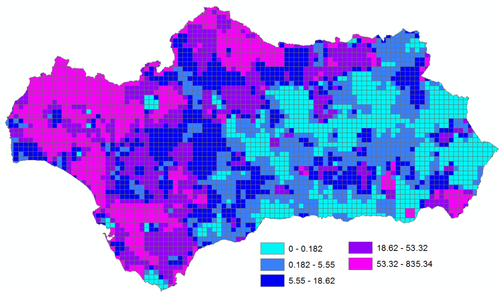
Figure B.10: Comparison between dense woodland (a), and rainfed cropland (b) in the current situation, and under the 2040 horizon A2 scenario of change.

a) Farming units cattle (Livestock units)

Current situation

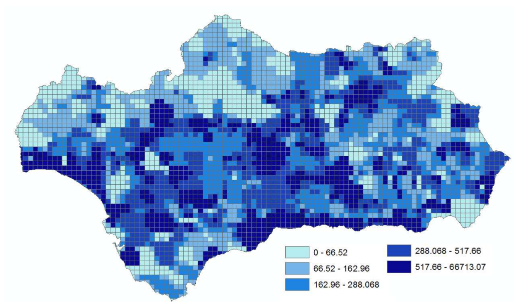


2040 horizon, A2 scenario



b) Summer water consumption (mm)

Current situation



2040 horizon, A2 scenario

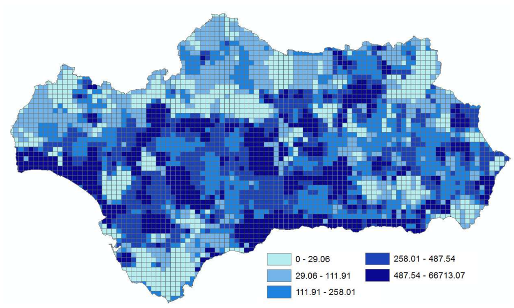


Figure B.11: Comparison between farming units cattle (a), and summer water consumption (b) in the current situation, and under the 2040 horizon A2 scenario of change.

References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian networks in environmental modelling. Environmental Modelling & Software 26, 1376–1388.

Aguilera, P. A., Fernández, A., Ropero, R. F., Molina, L., 2013. Groundwater quality assessment using data clustering based on hybrid Bayesian networks. Stochastic Environmental Research & Risk Assessment 27 (2), 435–447.

Anderberg, M. R., 1973. Cluster Analysis for Applications. Academic Press.

Anderson-Teixeira, K. J., Miller, A., Mohan, J., Hudiburg, T., Duval, B. D., Delucia, E., 2013. Altered dynamics of forest recovery under a changing climate. Global Change Biology 19, 2001–2021.

Basurto, X., Gelcich, S., Ostrom, E., 2013. The social-ecological system framework as a knowledge classificatory system for benthic small-scale fisheries. Global Environmental Change 23, 1366–1380.

Borsuk, M. E., Reichert, P., Peter, A., Schager, E., Burkhardt-Holm, P., 2006. Assessing the decline of brown trout (*Salmo trutta*) in Swiss rivers using Bayesian probability networks. Ecological Modelling 192, 224–244.

Borsuk, M. E., Stow, C. A., Reckhow, K. H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173, 219–239.

Cadenasso, M., Pickett, S., Grove, J., 2006. Dimensions of ecosystem complexity: Heterogeneity, connectivity, and history. Ecological Complexity 3, 1–12.

Challies, E., Newig, J., Lenschow, A., 2014. What role for social-ecological systems research in governing global teleconnections?. Global Environmental Change 27, 32–40.

Chuman, T., Romportl, D., 2010. Multivariate classification analysis of cultural landscapes: An example from the Czech Republic. Landscape and Urban Planning 98, 200–209.

Claesson, J., Nycander, J., 2013. Combined effect of global warming and increased CO2-concentration on vegetation growth in water-limited conditions. Ecological Modelling 256, 23–30.

Clark, W. C., Dickson, N. M., 2003. Sustainability science: The emerging research program. PNAS 100, 8059–8061.

Cobb, B., Rumí, R., Salmerón, A., 2007. Bayesian networks models with discrete and continuous variables. Advances in probabilistic graphical models, Ch. Studies in Fuzziness and Soft Computing, pp. 81–102.

Dearing, J., Wang, R., Zhang, K., Dyke, J. G., Haberl, H., Hossain, M. S. Langdon, P. G., Lenton, T., Raworth, K., Brown, S., Carstensen, J., Cole, M. J., Cornell, S. E., Dawson, T. P., Doncaster, C. P., Eigenbrod, F., Florke, M., Jeffers, E., Mackay, A., Nykvist, B., Poppy, G. M., 2014. Safe and just operating spaces for regional social-ecological systems. Global Environmental Change 28, 227–238.

Duda, R. O., Hart, P. E., Stork, D. G., 2001. Pattern classification. Wiley Interscience.

Elvira-Consortium, 2002. Elvira: An Environment for Creating and Using Probabilistic Graphical Models. In: Proceedings of the First European Workshop on Probabilistic Graphical Models. pp. 222–230.
URL http://leo.ugr.es/elvira

ESRI, 2006. ArcMap Version 10.0. Environmental Systems Research Institute (ESRI), Redlands, CA.

Fernandes, J., Irigoien, X., Goikoetxea, N., Lozano, J. A., Inza, I., Pérez, A., Bode, A., 2010. Fish recruitment prediction using robust supervised classification methods. Ecological Modelling 221, 338–352.

Fernández, A., Gámez, J. A., Rumí, R., Salmerón, A., 2014. Data clustering using hidden variables in hybrid Bayesian networks. Progress in Artificial Intelligence 2(2), 141–152.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Machine Learning 29, 131–163.

Gasca, A. M., 2014. Guía de escenarios regionalizados de cambio climático sobre España a partir de los resultados del IPCC-AR4. AEMET, Ministerio de Agricultura, Alimentación y Medio Ambiente.

Gieder, K. D., Karpanty, S. M., Fraser, J. D., Catlin, D., Gutierrez, B. T., Plant, N. G., Turecek, A. M., Thieler, E. R., 2014. A Bayesian network approach to predicting nest presence of the federally-threatened piping plover (Charadrius melodus) using barrier island features. Ecological Modelling 276, 38–50.

Giménez-Casalduero, F., Gomariz-Castillo, F. J., Calvín, J. C., 2011. Hierarchical classification of marine rocky landscape as management tool at southeast Mediterranean coast. Ocean & Coastal Management 54, 497–506.

Haberl, H., Winiwarter, V., Andersson, K., Ayres, R., Boone, C., Castillo, A., Cunfer, G., Fischer-Kowalski, M., Freudenburg, W., Furman, E., Kaufmann, R., Krausmann, F., Langthaler, E., Lotze-Campen, H., Mirtl, M., Redman, C. L., Reenberg, A., Wardell, A., Warr, B., Zechmeister, H., 2006. From LTER to LTSER: Conceptualizing the Socioeconomic Dimension of Long-term Socioecological Research. Ecology and Society 11(2), 1–34.

Hamilton, S. H., Pollino, C. A., Jakeman, A. J., 2015. Habitat suitability modelling of rare species using Bayerian networks: Model evaluation under limited data. Ecological Modelling 299, 64–78.

Hufnagl-Eichiner, S., Wolf, S. A., Drinkwater, L. E., 2011. Assessing social-ecological coupling: Agriculture and hypoxia in the Gulf of Mexico. Global Environmental Change 21, 530–539.

Jackson, L. E., Pulleman, M. M., Brussaard, L., Bawa, K. S., G., B. G., Cardoso, I. M., Ruiter, P. C., García-Barrios, L., Hollander, A. D., Lavelle, P., Ouédraogo, E., Pascual, U., Setty, S., Smukler, S. M., Tscharntke, T., Van Noordwijk, M., 2012. Social-ecological and regional adaptation of agrobiodiversity management across a global set of research regions. Global Environmental Change 22, 623–639.

Jensen, F. V., Lauritzen, S. L., Olesen, K. G., 1990. Bayesian updating in causal probabilistic networks by local computation. Computational Statistics Quarterly 4, 269–282.

Jensen, F. V., Nielsen, T. D., 2007. Bayesian Networks and Decision Graphs. Springer.

Keenan, T., Serra, J., Lloret, F., Ninyerola, M., Sabate, S., 2011. Predicting the future of forests in the Mediterranean under climate change, with niche- and process- based models: CO2 matters! Global Change Biology 17, 565–579.

Kelly, R., Jakeman, A. J., Barreteau, O., Borsuk, M., ElSawah, S., Hamilton, S., Henriksen, H. J., Kuikka, S., Maier, H., Rizzoli, E., Delden, H., Voinov, A., 2013. Selecting among five common approaches for integrated environmental assessment and management. Environmental Modelling & Software 47, 159–181.

Keshtkar, A. R., Slajegheh, A., Sadoddin, A., Allan, M. G., 2013. Application of Bayesian networks for sustainability assessment in catchment modeling and management (Case study: The Hablehrood river catchment). Ecological Modelling 268, 48–54.

Kotova, T., Miklyaeva, I. M., Ogureeva, G. N., Suslova, E. G., Shvergunova, L. V., 2000. Experience in Mapping the Ecological State of the Plant Cover. Russian Journal of Ecology 31, 318–323.

Lahr, J., Kooistra, L., 2010. Environmental risk mapping of pollutants: State of the art and communication aspects. Science of the Total Environment 408, 3899–3907.

Lambin, E., Turner, B., Geist, H., Agbola, Angelsen, S., Bruce, A., Coomes, O., Dirzo, R., Fischer, G., Folke, C., George, P., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E., Mortimore, M., Ramakrishnan, P., Richards, F., Slanes, H., Steffen, W., Stone, G., Svedin, U., Veldkamp, T., Vogel, C., Xu, J., 2001. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11, 261–269.

Langmead, O., McQuatters-Gollop, A., Mee, L. D., Friedrich, J., Gilbert, A. J., Gomoiu, M. T., Jackson, E. L., Knudsen, S., Minicheva, G., Todorova, V., 2009. Recovery or decline of the northwestern Black Sea: A societal choice revealed by socio-ecological modelling. Ecological Modelling 220, 2927–2939.

Lauritzen, S. L., 1992. Propagation of probabilities, means and variances in mixed graphical association models. Journal of the American Statistical Association 87, 1098–1108.

Lauritzen, S. L., 1995. The EM algorithm for graphical association models with missing data. Computational Statistics and Data Analysis 19, 191 – 201.

Liedloff, A., Smith, C. S., 2010. Predicting a tree change in Australiaś tropical savannas: Combining different types of models to understand complex ecosystem behaviour. Ecological Modelling 221, 2565–2575.

Liu, K. F. R., Lu, C. F., Chen, C. W., Shen, Y. S., 2012. Applying Bayesian belief networks to health risk assessment. Stochastic Environmental Research & Risk Assessment 26, 451–465.

Liu, R., Zhang, K., Zhang, Z., Brothwick, A. G. L., 2014. Land-use suitability analysis for urban development in Beijing. Journal of Environmental Management 145, 170–179.

Marcot, B., 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecological Modelling 230, 50–62.

Martín de Agar, P., de Pablo, C. L., Pineda, F., 1995. Mapping the ecological structure of a territory: a case study in Madrid (central Spain). Environmental Management 19(3), 345–357.

Méndez-Jiménez, M., 2012. Estudio Básico de Adaptación al Cambio Climático.

Moral, S., Rumí, R., Salmerón, A., 2001. Mixtures of Truncated Exponentials in Hybrid Bayesian Networks. In: ECSQARU'01. Lecture Notes in Artificial Intelligence. Vol. 2143. Springer, pp. 156–167.

Myers, N., Mittenmeier, R. A., Mittenmeier, C. G., da Fonseca, G. A. B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853 – 858.

Niederscheider, M., Kuemmerle, T., Muller, D., Erb, K., 2014. Exploring the effects of drastic institutional and socio-economic changes on land system dynamics in Germany between 1883 and 2007. Global Environmental Change 28, 98–108.

Nieto, J., Linares, P., 2011. Cambio Global España 2020/50. Energía, Economía y Sociedad. Centro Complutense de Estudios e Información Medioambiental.

Quisthoudt, K., Adams, J., Rajkaran, A., Dahdouh-Guebas, F., Koedam, N., Randin, C., 2013. Disentangling the effects of global climate and regional land-use change on the current and future distribution of mangroves in South Africa. Biodiversity and Conservation 22, 1369–1390.

Rapinel, S., Clément, B., Magnanon, S., Sellin, V., Hubert-Moy, L., 2014. Identification and mapping of natural vegetation on a coastal site using a Worldview-2 satellite image. Journal of Environmental Management 144, 236–246.

Renken, H., Mumby, P. J., 2009. Modelling the dynamics of coral reef macroalgae using a Bayesian belief network approach. Ecological Modelling 220, 1305–1314.

Ropero, R. F., Aguilera, P., Rumí, R., 2014a. Soft Clustering based on Hybrid Bayesian networks in Socioeclogical Cartography. In: Hybrid Artificial Intelligence Systems, 9th International Conference, HAIS, Salamanca, Spain. pp. 607–617.

Ropero, R. F., Aguilera, P. A., Fernández, A., Rumí, R., 2014b. Regression using hybrid Bayesian networks: Modelling landscape-socioeconomy relationships. Environmental Modelling & Software 57, 127–137.

Rubidge, E., Monahan, W., Parra, J., Cameron, S., Brashares, J., 2011. The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. Global Change Biology 17, 696–708.

Ruiz-Labourdette, D., Martínez, F., Martín-López, B., Montes, C., Pineda, F., 2011. Equilibrium of vegetation and climate at the European rear edge. A reference for climate change planning in mountainous Mediterranean regions. Int.J. Biometeorol 55, 285–301.

Rumí, R., Salmerón, A., 2007. Approximate probability propagation with mixtures of truncated exponentials. International Journal of Approximate Reasoning 45, 191–210.

Rumí, R., Salmerón, A., Moral, S., 2006. Estimating mixtures of truncated exponentials in hybrid Bayesian networks. Test 15, 397–421.

Schmitz, M., Pineda, F., Castro, H., Aranzabal, I. D., Aguilera, P., 2005. Cultural landscape and socioeconomic structure. Environmental value and demand for tourism in a Mediterranean territory. Consejería de Medio Ambiente. Junta de Andalucía. Sevilla.

Shenoy, P. P., Shafer, G., 1990. Axioms for probability and belief functions propagation. In: Shachter, R., Levitt, T., Lemmer, J., Kanal, L. (Eds.), Uncertainty in Artificial Intelligence, 4. North Holland, Amsterdam, pp. 169–198.

Smith, G., Brennan, R. E., 2012. Losing our way with mapping: Thinking critically about marine spatial planning in Scotland. Ocean & Coastal Management 29, 210–216.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., Miller, H., 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P., 2013. Climate Change 2013. The Physical Science Basis. Working Group I Contribution to the fifth Assessment Report of the Intergovermental Panel on Climate Change. WMO, UNEP.

Strand, G. H., 2011. Uncertainty in classification and delineation of landscapes: A probabilistic approach to landscape modeling. Environmental Modelling & Software 26.

Tanner, M. A., Wong, W. H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–550.

Tiller, R., Gentry, R., Richards, R., 2013. Stakeholder driven future scenarios as an element of interdisciplinary management tools; the case of future offshore aquaculture development and the potential effects on fishermen in Santa Barbara, California. Ocean & Coastal Management 73, 127–135.

Trincsi, K., Pham, T. T. H., Turner, S., 2014. Mapping mountain diversity: Ethnic minorities and land use land cover change in Vietnam's borderlands. Land Use Policy 41, 484–497.

Turner, B., Kasperson, R., Matson, P., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., Schiller, A., 2003. A framework for vulnerability analysis in sustainability science. PNAS 100, 8074–8079.

Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling 203, 312–318.

Varis, O., Kuikka, S., 1999. Learning Bayesian decision analysis by doing: lessons from environmental and natural resources management. Ecological Modelling 119, 177–195.

Vezeanu, C., Grigor-Pop, O., Gruia, R., Marculescu, A., 2010. Geospatial techniques in the cartography and management of habitats in Piatra Craiului National Park. Environmental Engineering and Management Journal 9, 1611–1617.

Villamagna, A. M., Mogollón, B., Angermeier, P. L., 2014. A multi-indicator framework for mapping cultural ecosystem services: The case of freshwater recreational fishing. Ecological Indicators 45, 255–265.

Virah-Sawmy, M., Gillson, L., Willis, K. J., 2009. How does spatial heterogeneity influence resilience to climatic changes? Ecological dynamics in southeast Madagascar. Ecological Monographs 79(4), 557–574.

Wilson, D. S., Stoddard, M. A., Puettmann, K. J., 2008. Monitoring amphibian populations with incomplete survey information using a Bayesian probabilistic model. Ecological Modelling 214, 210–218.