# Regression using hybrid Bayesian networks: modelling landscape - socioeconomy relationships

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

[a] *Informatics and Environment Laboratory, Dept. of Biology and Geology, University of Almería, Spain*
[b] *Dept. of Mathematics, University of Almería, Spain*

**Abstract**

Modelling environmental systems becomes a challenge when dealing directly with continuous and discrete data simultaneously. The aim in regression is to give a prediction of a response variable given the value of some feature variables. Multiple linear regression models, commonly used in environmental science, have a number of limitations: (1) all feature variables must be instantiated to obtain a prediction, and (2) the inclusion of categorical variables usually yields more complicated models. Hybrid Bayesian networks are an appropriate approach to solve regression problems without such limitations, and they also provide additional advantages. This methodology is applied to modelling landscape - socioeconomy relationships for different types of data (continuous, discrete or hybrid). Three models relating socioeconomy and landscape are proposed, and two scenarios of socioeconomic change are introduced in each one to obtain a prediction. This proposal can be easily applied to other areas in environmental modelling.

*Keywords:* Continuous Bayesian Networks, Mixtures of Truncated Exponentials, Regression, Landscape, Socioeconomic Structure

*Corresponding author. Tel.: +34 950214650; fax: +34 950015167.
*Email addresses:* rosa.ropero@ual.es (R. F. Ropero), aguilera@ual.es (P. A. Aguilera), afalvarez@ual.es (A. Fernández), rrumi@ual.es (R. Rumí)

## 1. Introduction

Bayesian networks (BNs) (Pearl, 1988; Jensen and Nielsen, 2007) play a relevant role in modelling of complex systems in which relationships between variables are subject to uncertainty. BNs provide an efficient framework for reasoning in terms of updating information about unobserved variables when some new information is incorporated into the system (Jensen et al., 1990; Shenoy and Shafer, 1990). Variables are modelled by means of probability distributions; therefore risk and uncertainty can be estimated more accurately than in other models (Uusitalo, 2007; Liao et al., 2010; Liu et al., 2012; Chen and Pollino, 2012).

Their graphical interpretation, based on nodes and arcs, allows stakeholders to easily understand the relationships between variables and refine the learned model manually by adding or removing arcs (even variables) from the graph to better represent reality (Voinov and Bousquet, 2010). Most available data in environmental sciences are continuous or hybrid (discrete and continuous), and while BNs can manage them, the limitations are too restrictive in many cases (Nyberg et al., 2006). The most widely-used solution in environmental modelling is to discretise the variables, accepting a loss of information (Bromley et al., 2005; Landuyt et al., 2013; Li et al., 2013; Nash et al., 2013; Sun and Müller, 2013). To date, several new solutions to this problem have been proposed, such as the *Conditional Gaussian* (CG) model (Lauritzen, 1992; Lauritzen and Jensen, 2001), the *Mixture of Truncated Exponentials* model (MTE) (Moral et al., 2001), the *Mixtures of Polynomials* model (MoP) (Shenoy and West, 2011) and the *Mixtures of Truncated Basis Functions* (MoTBFs) model (Langseth et al., 2012).

In the study of environmental systems, it is common to find problems in which the goal is to predict the value of a variable of interest depending on the values of some other observable variables. If the variable of interest is discrete, we are faced with a *classification problem*, whilst if it is continuous, it is usually called a *regression problem* (Hastie et al., 2009).

BNs have been proposed both for classification and regression purposes. Its main advantage with respect to other regression models is that it is not necessary to have a full observation of the features to give a prediction for the response variable; in addition, the model is usually richer from a semantic point of view.

Some restricted BNs models, such as the naïve Bayes (Minsky, 1963), have been applied to regression problems under the assumption that the

joint distribution of the features and the response variable is multivariate Gaussian (Gámez and Salmerón, 2005). If the normality assumption is not fulfilled, regression with naïve Bayes models has been approached using kernel densities to model the conditional distribution in the BN (Frank et al., 2000; Pérez et al., 2009), but the models obtained are too complex.

In a more general solution, we are interested in regression problems where the features can be either continuous or discrete. In this case, the joint distribution is not multivariate Gaussian in any case, due to the presence of discrete variables. To solve this problem, a naïve Bayes regression model based on the approximation of the joint distribution by an MTE was proposed (Morales et al., 2007).

Aguilera et al. (2011) reviewed the application of BNs in environmental modelling. Although there are few attempts in the literature to solve BNs-based regression problems in environmental science, these are focused on discrete response variables or Gaussian distributions unable to handle discrete and continuous variables simultaneously without constraints on the structure (Malekmohammadi et al., 2009; Pérez-Miñana et al., 2012). Moreover, in a more general setting, hybrid BNs have scarcely been applied in environmental modelling (Aguilera et al., 2010, 2013).

Territorial (landscape spatial pattern) and socioeconomic structures maintain a constant and reciprocal interaction since they are "co-evolving systems" (Norgaard, 1984; Turner et al., 1988; Schmitz et al., 2003; Lacitignola et al., 2007). Thus, socioeconomic processes, as drivers of change (Burgi et al., 2004), are the main cause of changes in land uses, *i.e.*, it determines the structure, function and dynamics of landscapes (Bicik et al., 2001; Wu and Hobbs, 2002). European agricultural landscapes have been undergoing significant changes associated with intense and rapid socio-economic changes (Nikodemus et al., 2005; Strijker, 2005). In Europe, and particularly in Spain, socioeconomic development has led to a notable migration of the rural population to the city, and the abandonment of the countryside.

Modelling environment-human relationships are becoming increasingly important and it has been applied in decision-making processes (Wang and Zhang, 2001; Serra et al., 2008; Milne et al., 2009; Celio et al., 2014). More specifically, the relationships between landscape structure and socioeconomy have been formalized through Multiple Linear Regression (MLR) (Schmitz et al., 2003, 2005). This procedure provides a dependence model with a limited number of socioeconomic variables, which themselves can account for much of the variation in the landscape structure.

There are no studies related to modelling landscape-socioeconomy relationship using hybrid BNs. Thus, our objective is to develop a regression model based on a hybrid BN that can be applied to study landscape-socioeconomy relationships. The article is organised as follows. Section 2 presents the theory behind the proposed hybrid BN-based regression model. In Section 3, the methodology proposed, using three types of data, is validated against MLR and a case study with three different landscape tendency changes is presented. Finally, Section 4 is devoted to draw some conclusions and future work.

## 2. Hybrid BNs-based regression

### *2.1. Bayesian networks*

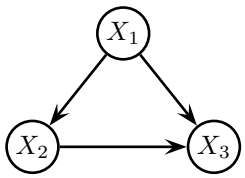
A *Bayesian network* (Jensen et al., 1990; Shenoy and Shafer, 1990) is a statistical multivariate model for a set of variables $\mathbf{X}$[1] appropriate for knowledge representation under uncertainty. It is explained in terms of two components: (1) *qualitative*, defined by means of a directed acyclic graph (DAG), in which arcs linking nodes determine the independence relations between them (see example in Fig. 1(a)); and (2) *quantitative*, specified using a conditional distribution $p(x_i \mid pa(x_i))$ for each variable $X_i$, $i = 1, \ldots, n$, given its parents in the graph, denoted as $pa(X_i)$ (see example in Fig. 1(b)).

The success of BNs stems from the fact that the DAG structure gives us information about which variables are relevant or irrelevant for some other variable of interest, which allows us to simplify, to a significant extent, the joint probability distribution (JPD) of the variables necessary to specify the model. In other words, BNs provide a compact representation of the JPD over all the variables, defined as the product of the conditional distributions attached to each node, so that

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid pa(x_i)). \tag{1}$$

For instance, the JPD associated to the network in Fig. 1, $p(x_1, x_2, x_3)$, is simplified as the product $p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_1, x_2)$.

[1]Uppercase letters denote random variables and boldfaced uppercase letters denote a set of variables, e.g. $\mathbf{X} = \{X_1, \ldots, X_n\}$. The domain of $\mathbf{X}$ is denoted as $\Omega_{\mathbf{X}}$. By lowercase letters $x$ (or $\mathbf{x}$) we denote some element of $\Omega_X$ (or $\Omega_{\mathbf{X}}$).




(a) Qualitative component.

$$
\begin{aligned}
&P(X_1 = 0) = 0.4 \\
&P(X_2 = 0 \mid X_1 = 0) = 0.4 \\
&P(X_2 = 0 \mid X_1 = 1) = 0.3 \\
&P(X_3 = 0 \mid X_1 = 0, X_2 = 0) = 0.2 \\
&P(X_3 = 0 \mid X_1 = 0, X_2 = 1) = 0.1 \\
&P(X_3 = 0 \mid X_1 = 1, X_2 = 0) = 0.7 \\
&P(X_3 = 0 \mid X_1 = 1, X_2 = 1) = 0.3
\end{aligned}
$$

(b) Quantitative component.

Figure 1: An example of a discrete Bayesian network with three binary variables. Note that $P(X_i = 1 \mid \ldots)$ is not specified as it can directly be computed as $1 - P(X_i = 0 \mid \ldots)$.

There are two approaches for learning a BN: automatic and manual (or a mixture of the two). The first approach involves using algorithms which, starting with a set of training data, calculate the optimum structure for them (Cooper and Herskovits, 1992; Spirtes et al., 1993). From here, the corresponding probability distributions are estimated. In contrast, using manual approximation, expert opinion is included as part of the process to indicate which variables are related and how strongly. This second option is often used when no training data are available or some of them are missing.

A BN can carry out an efficient reasoning for a given scenario under conditions of uncertainty. This is known as probability propagation or probabilistic inference. Hence, the objective is to obtain information about a set of variables of interest (unobserved variables) given known values of other variables (observed or evidenced variables). If we denote the evidenced variables as $\mathbf{E}$, and its values as $\mathbf{e}$, then we can calculate the posterior probability distribution, $p(x_i \mid \mathbf{e})$, for each variable of interest $X_i \notin \mathbf{E}$.

### *2.2. Mixtures of truncated exponentials*

BNs were originally proposed for handling discrete variables (see Fig. 1(b)) and today, a broad and consolidated theory can be found in the literature for this case (Jensen and Nielsen, 2007). However, in environmental systems, it is very common to find continuous and discrete domains simultaneously.

In a hybrid framework, the simplest and the most common solution is to discretise the continuous data and treat them as if they were discrete. Thus, existing methods for discrete variables can be easily applied. However,

discretisation of variables can lead to a loss in precision and this is why other approaches have received so much attention over the last few years.

So far, several approaches have been devised to represent probability distributions in hybrid BNs. The CG model is used extensively by researchers but it puts some restrictions on the network. It is only useful in situations where the joint distribution of the continuous variables, for each configuration of the discrete ones, follows a multivariate Gaussian. Moreover, CG models are not valid in frameworks where a discrete variable has continuous parents, even though some attempts to overcome this restriction have been addressed (Lerner et al., 2001).

On the other hand, the MOPs and MoTBFs models have been recently proposed as promising solutions, but they still have not been applied to solving regression problems.

Discretisation is equivalent to approximating a density by a mixture of uniforms, meaning that each interval is approximated by a constant function. Thus, the accuracy of the final model could be increased if, instead of constants, other functions with better fitting properties were used. A good choice are exponential functions, since they are closed under restriction, marginalisation and combination. This is the idea behind the so-called MTE model, explained next.

During the probability inference process, where the posterior distributions of the variables are obtained given some evidence, the intermediate functions are not necessarily density functions. Therefore, a general function called *MTE potential* needs to be defined as follows:

**Definition 1.** (MTE potential) *Let* $\mathbf{X}$ *be a mixed* $n$*-dimensional random vector. Let* $\mathbf{Z} = (Z_1, \ldots, Z_d)^{\mathsf{T}}$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_c)^{\mathsf{T}}$ *be the discrete and continuous parts of* $\mathbf{X}$*, respectively, with* $c + d = n$*. We say that a function* $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ *is a* Mixture of Truncated Exponentials potential (MTE potential) *if one of the next conditions holds:*

i. $\mathbf{Z} = \emptyset$ *and* $f$ *can be written as*

$$f(\mathbf{x}) = f(\mathbf{y}) = a_0 + \sum_{i=1}^{m} a_i \mathrm{e}^{\left\{\mathbf{b}_i^{\mathsf{T}}\mathbf{y}\right\}} \tag{2}$$

*for all* $\mathbf{y} \in \Omega_{\mathbf{Y}}$*, where* $a_i \in \mathbb{R}$ *and* $\mathbf{b}_i \in \mathbb{R}^c$*,* $i = 1, \ldots, m$*.*

ii. $\mathbf{Z} = \emptyset$ *and there is a partition* $D_1, \ldots, D_k$ *of* $\Omega_{\mathbf{Y}}$ *into hypercubes such that* $f$ *is defined as*

$$f(\mathbf{x}) = f(\mathbf{y}) = f_i(\mathbf{y}) \quad \text{if} \quad \mathbf{y} \in D_i,$$

*where each* $f_i$, $i = 1, \ldots, k$ *can be written in the form of Equation* (2).

iii. $\mathbf{Z} \neq \emptyset$ *and for each fixed value* $\mathbf{z} \in \Omega_{\mathbf{Z}}$, $f_{\mathbf{z}}(\mathbf{y}) = f(\mathbf{z}, \mathbf{y})$ *can be defined as in* ii.

**Definition 2.** (MTE density) *An MTE potential* $f$ *is an* MTE density *if*

$$\sum_{\mathbf{z} \in \Omega_{\mathbf{Z}}} \int_{\Omega_{\mathbf{Y}}} f(\mathbf{z}, \mathbf{y}) d\mathbf{y} = 1.$$

A *conditional MTE density* can be specified by dividing the domain of the conditioning variables and specifying an MTE density for the conditioned variable for each configuration of splits of the conditioning variables.

**Example 1.** *Consider two continuous variables* $Y_1$ *and* $Y_2$. *A possible conditional MTE density for* $Y_1$ *given* $Y_2$ *is the following:*

$$f(y_1 \mid y_2) = \begin{cases} 0.28 + 0.01\mathrm{e}^{1.03y_1} + 0.02\mathrm{e}^{0.01y_1} & \text{if } 0 \leq y_1 < 1,\ 1 \leq y_2 < 3, \\ 0.02 + 0.02\mathrm{e}^{1.01y_1} + 0.12\mathrm{e}^{0.09y_1} & \text{if } 1 \leq y_1 < 3,\ 1 \leq y_2 < 3, \\ 0.49 - 0.12\mathrm{e}^{0.59y_1} - 0.24\mathrm{e}^{-0.08y_1} & \text{if } 0 \leq y_1 < 1,\ 3 \leq y_2 < 4, \\ 0.07 - 0.02\mathrm{e}^{-0.23y_1} + 0.62\mathrm{e}^{-0.23y_1} & \text{if } 1 \leq y_1 < 3,\ 3 \leq y_2 < 4. \end{cases}$$

In the same way as in discretisation, the more intervals used to divide the domain of the continuous variables, the better the MTE model accuracy, but also more complex. Furthermore, in the case of MTEs, using more exponential terms within each interval substantially improves the fit to the real model, but again more complexity is assumed.

### *2.3. MTE Bayesian networks for regression*

Assume we have a set of variables $Y, X_1, \ldots, X_n$. Regression analysis consists in finding a model $g$ that explains the response variable $Y$ in terms of the feature variables $X_1, \ldots, X_n$, so that given a full observation of the features $x_1, \ldots, x_n$, a prediction about $Y$ can be obtained as $\hat{y} = g(x_1, \ldots, x_n)$.

A BN can be used as a regression model for prediction purposes if it contains a continuous response variable $Y$ and a set of discrete and/or continuous feature variables $X_1, \ldots, X_n$. Thus, in order to predict the value for $Y$ from $k$ observed features, with $k \leq n$, the conditional density

$$f(y \mid x_1, \ldots, x_n), \tag{3}$$

is computed, and a numerical prediction for $Y$ is given[2] using the expected value as follows:

$$\hat{y} = g(x_1, \ldots, x_n) = \mathbb{E}[Y \mid x_1, \ldots, x_n] = \int_{\Omega_Y} y f(y \mid x_1, \ldots, x_n) dy, \tag{4}$$

where $\Omega_Y$ represents the domain of $Y$.

Note that $f(y \mid x_1, \ldots, x_n)$ is proportional to $f(y) \times f(x_1, \ldots, x_n \mid y)$, and therefore, solving the regression problem would require a distribution to be specified over the $n$ variables given $Y$. The associated computational cost can be very high. However, using the factorisation determined by the network, the cost is reduced. Although the ideal would be to build a network without restrictions on the structure, usually this is not possible due to the limited data available. Therefore, networks with fixed and simple structures are used.

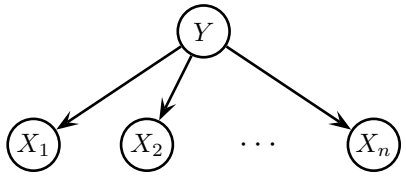



Figure 2: Structure of a *naïve Bayes* model.

The extreme case is the so-called *naïve Bayes* (NB) structure (Friedman et al., 1997; Duda et al., 2001). It consists of a BN with a single root node and a set of features having only one parent (the root node). The NB model structure is shown in Fig. 2.

[2]Note that in the BN framework, a prediction of $Y$ can be obtained even when some of the variables are not observed.

Its name comes from the naive assumption that the features $X_1, \ldots, X_n$ are considered independent given $Y$. This strong independence assumption is somehow compensated by the reduction in the number of parameters to be estimated from data, since in this case, it holds that

$$f(y \mid x_1, \ldots, x_n) \propto f(y) \prod_{i=1}^{n} f(x_i \mid y), \tag{5}$$

which means that, instead of one $n$-dimensional conditional distribution, $n$ one-dimensional conditional distributions are estimated. Despite this extreme independence assumption, the results are competitive with respect to other models.

However, if some variables are highly correlated, the error in the regression would decrease if any dependence between them could be included in the network (*i.e.*, links between features). There are several structures in which each feature is permitted to have more parents beside $Y$, for instance, TAN (Friedman et al., 1997), FAN (Lucas, 2002), $k$DB (Sahami, 1996) or AODE (Webb et al., 2005). These models are richer but an increase of complexity is assumed instead, both in the structure and the probability learning. In this study, the scarcity of data does not allow the use of complex structures and that is why we opted for the NB structure to develop the regression model.

In any case, regardless of the structure employed, it is necessary that the joint distribution for $Y, X_1, \ldots, X_n$ follows a model for which the computation of the density in Equation (3) can be carried out efficiently. As we are interested in models able to simultaneously handle discrete and continuous variables without any restriction in the structure developed, the approach that best meets these requirements is the MTE model.

Regarding inference, the posterior MTE distribution, $f(y \mid x_1, \ldots, x_n)$, will be computed using the Variable Elimination algorithm (Li and D'Ambrosio, 1994; Dechter, 1996; Zhang and Poole, 1996).

For learning the model, we follow the approach of Morales et al. (2007) to estimate the corresponding conditional distributions. Let $X_i$ and $Y$ be two random variables, and consider the conditional density $f(x_i \mid y)$. The idea is to split the domain of $Y$ by using the equal frequency method with three intervals. Then, the domain of $X_i$ is also split using the properties of the exponential function, which is concave, and increases over its whole domain (see Rumí et al. (2006)). Accordingly, the partition consists of a series of

intervals whose limits correspond to the points where the empirical density changes between concavity and convexity or decrease and increase. In case of models with more than one conditioning variable, see Moral et al. (2003) for more details.

At this point, a 5-parameter MTE is fitted for each split of the support of $X$, which means that in each split there will be 5 parameters to be estimated from data:

$$f(x) = a_0 + a_1 e^{a_2 x} + a_3 e^{a_4 x}, \quad \alpha < x < \beta\,, \tag{6}$$

where $\alpha$ and $\beta$ define the interval in which the density is estimated.

The reason to use the 5-parameter MTE lies in its ability to fit the most common distributions accurately, while the model complexity and the number of parameters to estimate is low (Cobb et al., 2006). The estimation procedure is based on least squares (Romero et al., 2006; Rumí et al., 2006).

A natural way to obtain the predicted value from the distribution is to compute its expectation. Thus, the expected value of a random variable $X$ with a density defined as in Equation (6) is computed as

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{\alpha}^{\beta} x(a_0 + a_1 e^{a_2 x} + a_3 e^{a_4 x}) dx \\
&= a_0 \frac{\beta^2 - \alpha^2}{2} + \frac{a_1}{a_2^2}((a_2\beta - 1)e^{a_2\beta} - (a_2\alpha - 1)e^{a_2\alpha}) + \\
&\quad \frac{a_3}{a_4^2}((a_4\beta - 1)e^{a_4\beta} - (a_4\alpha - 1)e^{a_4\alpha}).
\end{aligned}$$

If the density is defined by different intervals, the expected value would be the sum of the expression above for each part.

*2.4. Feature selection*

It is well known in prediction problems (particularly in regression) that, in general, including more variables does not necessarily increase the model accuracy. It can happen that some variables are not informative for the prediction task, and therefore including them in the model provides noise to the predictor. Also, unnecessary variables increase the number of parameters required to be determined from data. There are different approaches to the feature selection problem:

- The *filter* approach, which in its simplest formulation, consists in establishing a ranking of the variables according to some measure of relevance with respect to the class variable, usually called *filter measure*. Then, a threshold for the ranking is selected and those variables below that threshold are discarded.

- The *wrapper* approach proceeds by constructing several models with different sets of feature variables, and selecting the model that gives the highest accuracy.

- The *filter-wrapper* approach (Ruiz et al., 2006) is a mixture of the above two options. First of all, the variables are sorted using a filter measure and then, using that order, they are included only if they increase the accuracy of the current model.

The problem of selecting the features to be included in the MTE model was addressed by Morales et al. (2007) following a filter-wrapper approach. The accuracy of the model is measured using the root mean squared error (rmse) (Witten and Frank, 2005) between the actual values of the response variable, $y_1, \ldots, y_n$, and those predicted by the model, $\hat{y}_1, \ldots, \hat{y}_n$, for the records in a test database. In practice, this external test set is rarely available as such. Instead, the original dataset is randomly divided into two sets, one for learning the model, and the other for testing it. Thus, the rmse is obtained as

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{7}$$

The mutual information between two random variables $X$ and $Y$ is defined as

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}dxdy, \tag{8}$$

where $f_{XY}$ is the joint density for $X$ and $Y$, $f_X$ is the marginal density for $X$ and $f_Y$ is the marginal for $Y$.

The mutual information has been successfully applied as a filter measure in classification problems with continuous features (Pérez et al., 2006). For MTEs, the computation of Equation (8) cannot be obtained in closed form. We will therefore use the estimation procedure proposed by Morales et al. (2007), which is based on the estimator

$$\hat{I}(X,Y) = \frac{1}{m}\sum_{i=1}^{m}\left(\log_2 f_{X|Y}(X_i \mid Y_i) - \log_2 f_X(X_i)\right), \tag{9}$$

for a sample of size $m$, $(X_1, Y_1), \ldots, (X_m, Y_m)$, drawn from $f_{XY}$.

### *2.5. The naïve Bayes model for regression*

The task of estimating this model from data is simplified since the underlying structure is fixed beforehand, as in Fig. 2. The detailed steps for its construction can be found in the appendix, Algorithm 1.

This procedure includes all the available features in the model. The version in which the features are filtered and selected is called *selective*. We follow the filter-wrapper approach presented in Section 2.4. The steps for its construction are detailed in Algorithm 2, and graphically shown with an example in Fig. 3. The main idea is to start with a model containing the class variable and one feature variable, which is the node with the highest mutual information with respect to the response variable ($Y$ and $X_{(1)}$). Afterwards, the remaining variables are included in the model in sequence, according to their mutual information with respect to $Y$. In each step, if the included variable reduces the error defined in Equation (7), it is kept. Otherwise, it is discarded.

## 3. Case study

### *3.1. Methodology*

#### *3.1.1. Study area*

The study area is located in southeastern Spain, straddling parts of Almería and Granada provinces (Fig. 4). It covers around 500,000 ha. It has a spatially and temporally irregular rain pattern. Spatially, rainfall ranges from 300 mm in the South, to 700 mm in the highland area, increasing to 850 mm in wet years. This rain patterns has configured a particular cultural landscape (García-Latorre and Sánchez-Picón, 2001).

Landscape is characterized by an altitude gradient from the sea level to the peak of the study area (Sierra Nevada and Filabres mountains with more than 2,000 meters a.s.l.). The lowland part of the study area, named "Campo de Dalías" has an extension of more than 18,000 ha covered by greenhouses. In contrast, the middle to high altitude agricultural landscape

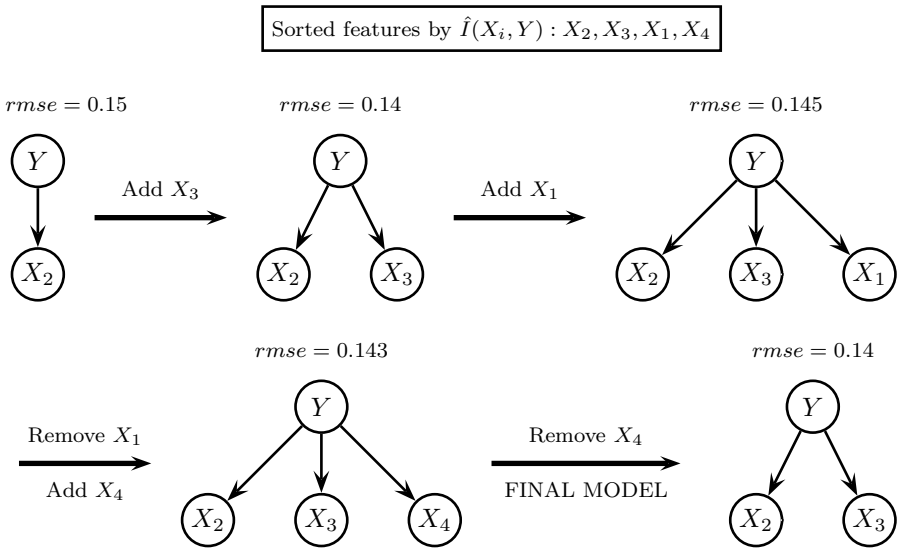



Figure 3: Example of feature selection in a NB regression model. First, features are sorted in a decreasing order using its mutual information with respect to $Y$. Then, their inclusion is checked step by step. Note that only the inclusion of $X_2$ and $X_3$ reduces the error. Finally, the procedure selects two out of four variables to be part of the final model.

is a patchwork of olive and almond groves, grapevines, subsistence croplands, and forest (conifers and oak).

The study area contains 90 municipalities, those from "Campo de Dalías" area are quite densely populated with a high degree of migration. Socioeconomic activities are linked to primary (intensive agriculture with greenhouses) and tertiary sector, which is related to the development of intensive agriculture (e.g. large numbers of banks and shopping centers). The municipalities in the middle to high altitudes are less populous and more prone to depopulation through emigration; there is less primary sector activity and, in some cases, rural tourism is more pronounced than in the lowland area.

*3.1.2. Data collection*

Table 1 shows the selected socioeconomic variables which are representative of the socioeconomic structure of the territory (Schmitz et al., 2005; Aranzabal et al., 2008). Data were obtained per municipality in 2007 from the Andalusian Statistical Institute (Andalusian Regional Government).

Using the landscape typologies described by Schmitz et al. (2005), we

Table 1: Socioeconomic variables. National emigration/immigration refers to people who emigrate/immigrate to/from other places in Spain, while foreign emigration/immigration refers to emigrants/immigrants to/from other countries.

| Socioeconomic variables | Unit |
|---|---|
| Total population | No. people |
| Ageing[a] | % of people |
| Natural increase[b] | Value of Natural increase |
| Male index[c] | No. males / No. females |
| Primary sector | No. employees |
| Secondary sector | No. employees |
| Tertiary sector | No. employees |
| Unemployed | No. unemployed |
| National emigration | % of people |
| Foreign emigration | |
| National immigration | |
| Foreign immigration | |
| Illiterate | % of people |
| Primary studies | |
| Secondary studies | |
| Higher studies | |

[a]Ageing is defined as the percentage of the population older than 65.

[b]Natural increase refers to the difference between the number of births and deaths.

[c]Male index is included since in the last decades in Spain, rural areas presented more male population than female (Camarero et al., 2009).

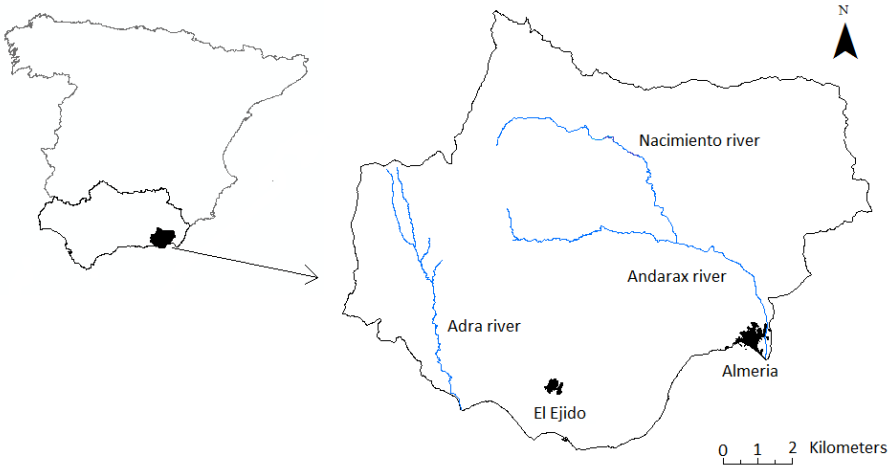



Figure 4: Study area.

selected three types of landscape: scrubland (dense and sparse scrubland), agricultural Mediterranean landscape (heterogeneous traditional croplands with olive trees and grapevine), and native forest (oak trees). Corresponding landscape data (percentages per municipalities) were obtained from Land Use and Land Cover shape file[3] using ArcGis v.9.3.1 (ESRI, 2006).

*3.1.3. Model learning*

Model learning was addressed using Elvira software (Elvira-Consortium, 2002) and involves performing feature selection, which is influenced by two issues. Firstly, as shown in Section 2.4, mutual information cannot be analytically computed, but it must be estimated from a simulated sample instead. If this sample size is *small*, the selected features can vary between different executions (Fernández et al., 2007). Secondly, the scarcity of data (only 90 instances) implies that the selected features strongly depends on the random test selected from the original dataset. To solve both problems, Algorithm 2 was run twenty times and the variables appearing at least 75% of the time were chosen. Accordingly, three continuous regression models were learned, one for each landscape (Figs. 5, 6 and 7) following the methodology described in Section 2.

In order to compare the performance of the continuous model (presented

[3] `http://www.juntadeandalucia.es/medioambiente/site/web/rediam`

Table 2: Intervals of socioeconomic and land use variables included in the discrete model. * refers to those variables discretised in the hybrid model. $k$-means method is used to discretise the variables.

| Socioeconomic variables | Intervals |
|---|---|
| Total population* | [98, 9519) [9519, 47510) [ 47510, 186651] |
| Ageing* | [ 6.76, 19.13) [ 19.13, 29.77 ) [ 29.77, 48.44 ] |
| Natural increase* | [ -29, 46 ) [ 46, 528) [ 528, 982 ] |
| Male index | [ 0.49, 0.72 ) [ 0.72, 0.88 ) [ 0.88, 1.04] |
| Tertiary sector* | [ 0.0, 2095.5 ) [ 2095.5, 8041.5 ) [ 8041.5, 11819.0 ] |
| Unemployed* | [ 2.0, 975.5 ) [ 975.5, 7936.5 ) [ 7036.5, 12645.0 ] |
| National emigration* | [ 0.78, 5.38) [ 5.38, 9.61 ) [ 9.61, 19.35 ] |
| Foreign emigration | [ 0.0, 0.24) [ 0.24, 1.58 ) [ 1.58, 3.87 ] |
| National immigration | [ 0.0, 5.38 ) [ 5.38, 16.63 ) [ 16.63, 28.21 ] |
| Foreign immigration | [ 0.0, 1.18 ) [1.18, 3.14 ) [ 3.14, 6.70 ] |
| Primary studies | [ 4.03, 16.85 ) [ 16.85, 28.96 ) [ 28.96, 43.0 ] |
| Secondary studies* | [ 14.87, 25.04) [ 25.04, 32.21) [ 32.21, 45.07 ] |
| *AML* | [ 4.45, 23.39 ) [ 23.39, 44.49 ) [44.49, 80.93 ] |
| *Scrubland* | [ 0.0, 5.20 ) [ 5.20, 15.94 ) [ 15.94, 29.75 ] |
| *Native forest* | [ 0.0, 18.11 ) [ 18.11, 36.71 ) [ 36.71, 67.35 ] |

above), against other alternatives, a hybrid and a discrete model were learned using Algorithm 1 with the same set of variables selected for the continuous case. In this way, the comparison is more reliable as the model structure remains fixed, and it makes more sense from an environmental point of view. In the hybrid approach, half of the variables were discretised (see Table 2). Note that, as explained in Section 2.2 the CG model cannot be applied in this situation, since there are some discrete feature variables with a continuous parent (the response variable). Several discretisation methods (equal frequency, equal width and $k$-means) were tested to obtaining the hybrid and the fully discrete model (including the response variable). Finally, the $k$-means algorithm with three intervals was used as it reported the best results in terms of rmse.

It should be remembered that a fully discrete model is mainly oriented towards classification and not to regression. Consequently, in order to compare this model with the hybrid and continuous cases, the rmse specified in Equation (7) needs to be re-computed for the discrete version as:

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - ca(\hat{c}_i))^2}\,, \tag{10}$$

where $ca(\hat{c}_i)$ is the class average for the predicted category after propagating the records in the discrete case, and $y_i$ is the actual continuous value for the response variable. Note that, once the data are discretised, the original continuous values are still necessary to compute this version of the rmse.

Direct (+) and inverse (-) relationships between each feature and the response variable were analysed. Two variables, $X$ and $Y$, are considered to have a direct relationship if an increase (or decrease) in the value of $X$ implies an increase (or decrease) in the expected value of the posterior distribution of $Y$. In contrast, an inverse relationship means that when the value of $X$ increases (or decreases), the expected value of $Y$ decreases (or increases). In order to check the sign of the relationships, for each feature, 10 equidistant values from its domain (including the minimum and maximum) were used as evidences for carrying out different propagations on the model. Thus, 10 expected values (means) of each posterior distribution for the response variable gave us information about the type of relationship (+ or -).

*3.1.4. Validation of the model*

The model was tested using $k$-fold cross-validation (Stone, 1974). It is a widely used technique in Artificial Intelligence to validate models. The aim is to check how predictive a model is when confronted with data that have not been previously used for learning the model. It is based on the holdout method in which the data set is separated into two complementary sets, one for learning ($D_l$) and another for testing ($D_t$). In this way, we can estimate the error of a model built from $D_l$ according to set $D_t$, using the formula in Equation (7).

To reduce variability, the data set is initially divided into $k$ subsets, and the holdout method is repeated $k$ times. Each time, one of the $k$ subsets is used as $D_t$ and the other $k-1$ subsets are put together to form $D_l$. Then the average error across all $k$ trials is computed. For the case study presented in this paper, we set the value of $k$ to 10.

Finally, the validation was conducted by comparing our BN-based proposals (continuous, hybrid, and discrete models), with a MLR implemented in `R` software (R Development Core Team, 2012), since it is the most common regression solution used in environmental sciences.

The MLR model can also be applied in the presence of categorical variables, usually by transforming them into dummy variables. In particular, each categorical variable with $k$ states has to be converted into $k-1$ binary variables, one for each category of the variable. However, the interpretation of the regression coefficients for the categorical variables is different from the continuous ones. Another disadvantage of this hybrid MLR approach is that the manual construction of dummy variables can be laborious and even error prone, especially in the case of many categories. On the other hand, Bayesian networks naturally include categorical and continuous variables in the same model using the MTE distributions without the need of creating new variables.

Hence, the idea is just to overview that the BN-based solution is coherent, and not to provide an exhaustive comparison of the two approaches.

*3.1.5. Scenarios of socioeconomic change*

Two scenarios of socioeconomic change were proposed (Table 3). The first scenario shows a positive socioeconomic development which involves an increase in the variables related to population, migration movement, study level (mainly in secondary and higher studies), and primary and tertiary economic sector. The second scenario shows a negative socioeconomic change. It involves a decrease in study level and primary and tertiary sector while an increase in emigration rates and unemployment.

Both scenarios represent general tendencies in the socioeconomic structure. As each regression model has a different subgroup of socioeconomic variables selected during the pre-processing step, the evidence is only introduced in those variables included in the corresponding model.

*3.2. Model validation results*

Table 4 shows the results in terms of the rmse (see Equations (7) and (10)) when comparing the four approaches for the three variables of interest. As expected, the errors for the proposed continuous method are smaller than the other approaches. It is well-known that discretising data implies loss of information as demonstrated in the errors obtained for the hybrid and discrete approaches. Finally, MLR obtains a significantly larger error than the BN-based approaches.

In any case, the goal of this paper is not only to compare the models above, but to present different ways to solve a regression problem in environmental modelling that carry fewer limitations, and which depend on the nature of the

Table 3: Scenarios of socioeconomic change. Minimum and maximum values refer to the minimum and maximum value found in the data set. Percentage changes are taken from Schmitz et al. (2005) and Aranzabal et al. (2008).

| Scenario | Variables involved | % Change |
|---|---|---|
| Positive socioeconomic change | Foreign immigration | Maximum value |
| | National emigration | +50% |
| | Tertiary sector | +60% |
| | Primary sector | +80% |
| | Higher studies | +15% |
| | Secondary studies | +30% |
| | Natural increase | +70% |
| | Ageing | Minimum value |
| Negative socioeconomic change | National emigration | Maximum value |
| | Higher studies | -70% |
| | Natural increase | Minimum value |
| | Primary sector | -20% |
| | Tertiary sector | -80% |
| | Total population | -50% |
| | Ageing | +80% |
| | Secondary studies | -40% |
| | Unemployment | Maximum value |

Table 4: Root mean squared error for the four BN-based regression models and the MLR. 10 fold-cross-validation is used to reduce variability.

| Model | *Native forest* | *AML* | *Scrubland* |
|---|---|---|---|
| Continuous BN | 6.47 | 14.73 | 18.60 |
| Hybrid BN | 6.74 | 14.89 | 19.13 |
| Discrete BN | 6.98 | 16.07 | 26.72 |
| MLR | 8.81 | 19.92 | 29.47 |

available data (continuous, hybrid and discrete). For a detailed comparison of BN-based regression models, see Morales et al. (2007).

### *3.3. Scenario results*

The previous section suggests the continuous approach to be the most appropriate one for modelling this problem as it has the lowest rmse. For this reason, only results from the continuous models are presented here.

Results are presented according to three different settings: *a priori* and *a posteriori* (two scenarios). The *a priori* information is obtained purely from the probability distributions learned from data, *i.e.*, the current data description. On the other hand, information *a posteriori* is computed by carrying out inference considering the proposed scenarios as evidence introduced into the model.

Figs. 5, 6, 7 show the qualitative part of the BNs developed and the direct and inverse relationships with the selected socioeconomic variables.

Fig. 8 shows the probability distributions *a priori* without introducing any scenario (black line), and the posterior probability distributions (blue and red line) of the variables after introducing the two socioeconomic scenarios according to Table 3.

Table 5 shows some statistics for each variable in the three different settings specified above.

#### *3.3.1. Agricultural Mediterranean Landscape (AML)*

*A priori*, *AML* (Fig. 5) is related to a socioeconomic structure characterized by a sparse total population with a low male index, but a positive natural increase. Educative level is medium with a high unemployment rate. National immigration is low. Variable *F*oreign immigration has a peculiar behaviour, since its middle and low values are related to agriculture workforce (mainly coming from northern Africa), and have a direct relationship

with *AML*. On the other hand, its high values are more related to retired population coming mostly from northern Europe, who have a second home in the area, and this has a inverse relationship with *AML*.

A negative scenario means a decrease in total population, natural increase and study level variables (as specified in Table 3). It entails a rural abandonment in which elder and non-qualified population, dedicated to traditional agricultural activities, stay in the area. Therefore, it involves an increase in the mean of the posterior probability distribution of *AML* (Table 5).

On the other hand, a positive scenario supposes larger values in natural increase, foreign immigration and study level (as specified in Table 3). Thus, the interests in other economic sectors, like in the tertiary sector, is increased. It involves a decrease in the mean of the posterior probability distribution of *AML*.

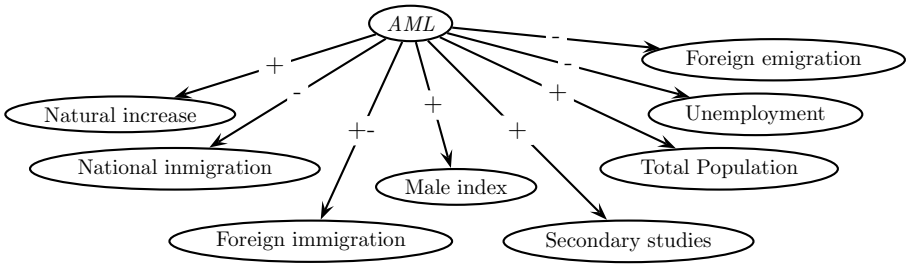



Figure 5: Regression model for variable *Agricultural Mediterranean Landscape* (*AML*) after the feature selection using Algorithm 2. Direct and inverse relationships between each feature and the variable *AML* are labelled with "+" and "-", respectively, on the corresponding arc. *F*oreign immigration is labelled with "+-" as it has a peculiar behaviour, low and middle values in its domain present a direct relationship with *AML*, however high values have a inverse relationship.

### *3.3.2. Scrubland*

*A priori*, *Scrubland* (Fig. 6) is related to a socioeconomic structure characterized by an ageing population with low study levels. Unemployment is considerable and national emigration prevails over international. In this context, tertiary sector is the main economic activity.

A negative scenario means an increase in ageing, national emigration and unemployment; and a decrease in secondary studies and tertiary sector (Table 3). It causes rural abandonment and entails an increase in the mean of the posterior probability distribution of *Scrubland*.

On the other hand, a positive scenario entails a decrease in variable *A*geing to its minimum and an increase in secondary studies, national emigration and tertiary sector (as specified in Table 3). Population growth and a increase in secondary study levels cause a higher interest in other economic activities replacing scrubland with other land uses. It, therefore, involves a decrease in the mean of the posterior probability distribution of *Scrubland*.

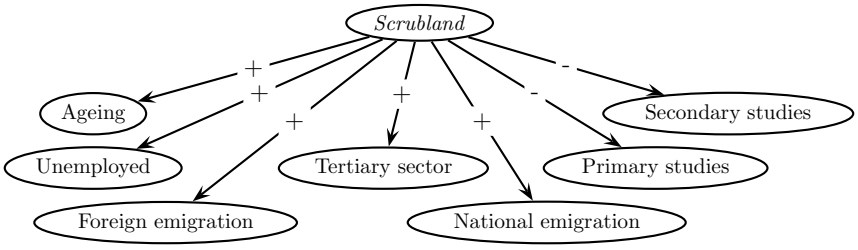



Figure 6: Regression model for variable *Scrubland* after the feature selection using Algorithm 2. Labels "+" and "-" have the same meaning as in Fig. 5.

*3.3.3. Native forest*

*A priori*, *Native forest* (with oak trees) (Fig. 7) is related to a socioeconomic structure characterized by low population with primary studies, but a positive value in natural increase variable. Moreover, the tertiary sector is well-developed and there is a national migration (immigration and emigration).

A negative scenario entails a decrease in the value of natural increase variable, total population and tertiary sector (rural tourism), whilst there is an increase in national emigration. Rural abandonment and low tourism levels entail less interest in maintaining native forest. It involves a slight decrease in the mean of the posterior probability distribution of *Native forest*.

On the other hand, a positive scenario means an increase in the tertiary sector, national emigration and natural increase variables. Population growth and greater touristic activities lead to an improvement in infrastructure not only for tourism, but also for residents. It entails the replacement of native forest with land uses related to those improvements. It involves a decrease in the mean of the posterior probability distribution of *Native forest*.

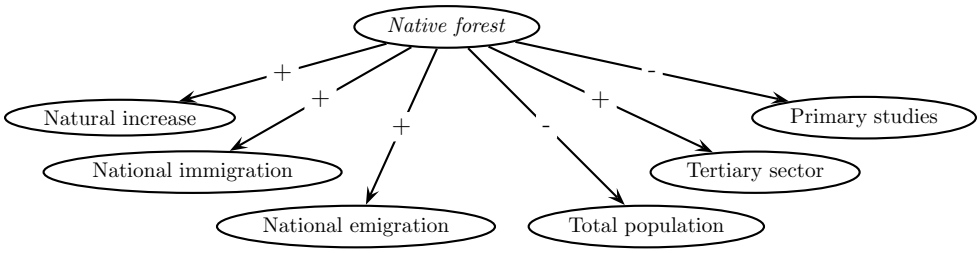



Figure 7: Regression model for variable *Native forest* after the feature selection using Algorithm 2. Labels "+" and "-" have the same meaning as in Fig. 5.

Table 5: Mean and standard deviation values *a priori* and in each scenario of socioeconomic changes. *AML* refers to *Agricultural Mediterranean Landscape*.

| | *A priori* | | Negative Scenario | | Positive Scenario | |
|---|---|---|---|---|---|---|
| | Mean | Sd | Mean | Sd | Mean | Sd |
| *AML* | 21.50 | 15.16 | 29.65 | 16.18 | 15.62 | 5.17 |
| *Native forest* | 7.66 | 6.71 | 6.32 | 5.79 | 3.81 | 3.48 |
| *Scrubland* | 39.62 | 18.73 | 49.98 | 16.22 | 26.96 | 18.52 |

### *3.3.4. Discussion of the results*

A positive scenario is defined as a development in the socioeconomic structure involving the decrease in *AML*, *Scrubland* and *Native forest*. Despite the population growth, traditional agricultural landscape is reduced, but it does not cause an increase in *Scrubland*. Moreover, the interest in native vegetation decreases causing a fall in *Native forest*. This new scenario can promote the development of new landscape typologies (Schmitz et al., 2005).

On the other hand, a negative scenario describes a similar situation to a rural abandonment. In such conditions, both *AML* and *Scrubland* tend to increase, while *Native forest* undergoes a slight reduction. Agricultural Mediterranean landscape, as a heterogeneous landscape, only in rural areas is kept, where elderly people cultivate small patches of traditional croplands (Schmitz et al., 2003). A lower level of education and fewer job opportunities mean a restriction in the number of economic activities, so that several patches are abandoned promoting the increase of *Scrubland* (Camarero et al., 2009). In that situation, traditional activities related to the maintenance of native forest are somehow forgotten, so the surface area of native forest is slightly reduced (Jiménez-Herrero et al., 2011).

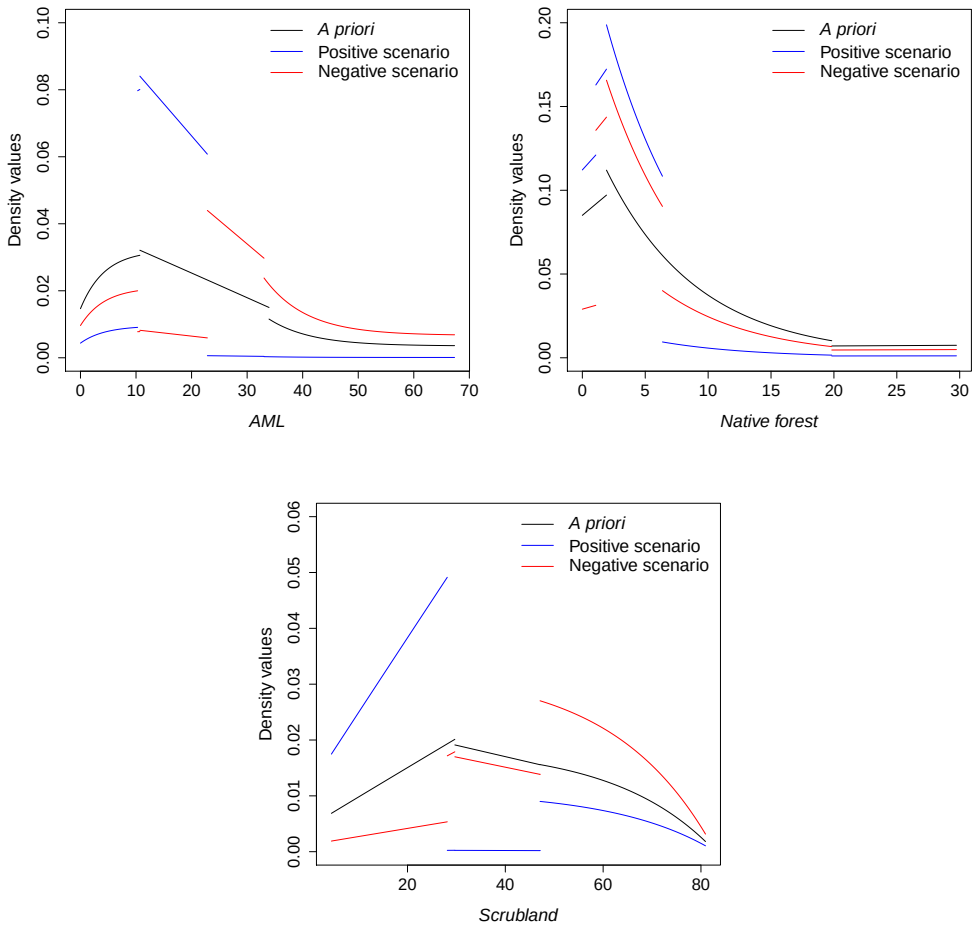



Figure 8: Probability distributions of the landscape variables in each continuous regression model. Three density functions are displayed: *a priori* (without introducing any scenario to the model), and *a posteriori* (introducing both positive and negative scenarios according to Table 3). Note that density functions are defined as a piecewise function using MTEs (for more information, see Example 1). By comparing the probability mass of the three densities, the impact of introducing different scenarios can be analysed.

## 4. Conclusions

This work presents MTE-based hybrid BNs as a tool for solving regression problems in environmental sciences, using the modelling of landscape - socioeconomy relationship in southern Spain as our study case. Three landscape tendency changes are studied under two socioeconomic scenarios. Advantages of a BN-based regression with respect to other solutions are presented below.

The proposed BN is able to deal with different types of data, totally continuous, totally discrete, and hybrid, in which continuous and discrete variables are both allowed in the model. One of the most common preprocessing tasks in data mining is discretisation, which involves loss of information, and therefore returns less accurate results (see Table 4). Thus, discretisation should be avoided when solving a regression problem. In the case of the available data being fully discrete (including the response variable), even if the problem could be solved with the methodology proposed, it is more appropriate to address it using classification models, which are specially proposed to handle this type of data.

A further advantage of using BN for regression is that a probability distribution obtained, and therefore any statistics of interest can be computed. In this way the information about uncertainty is richer as shown in Fig. 8.

Not all features must be instantiated to obtain a prediction, *i.e.*, information about the response variable can be obtained even if only partial information about the features is available. In the case study, the proposed scenarios of social change do not include all the features as we see in Section 3.1.5. It allows scenarios of change to be designed and the behaviour of the response variable to be checked. Also, probabilistic information can be extracted from other non-evidenced variables. This cannot be done with other regression solutions.

Validation in Section 3.1.4 reports better results in terms of error for the BN-based solutions vs. MLR, obtaining the continuous model the lowest error. As shown in Section 2.2, MTEs split the probability densities into pieces to better fit the real density determined from data, whilst other standard solutions use only one function (see for instance Morales et al. (2007); Fernández and Salmerón (2008); Fernández et al. (2010)). However, the number of parameters to be determined from data is high for MTEs. Thus, although more complexity in learning and inference is assumed, the results in terms of error are better. If data are limited, a BN solution might be poor

and other solutions could be more appropriate.

The versatility of BNs allows information from different sources to be included in the model, including expert knowledge. It is important to refine the model during the learning stage, for instance, when some data are missing or not at all representative.

For these reasons, MTE-based BNs are considered a novel approach to solve regression in environmental problems, in which continuous and discrete variables can be treated in the same model simultaneously without any restriction nor data preprocessing.

There are some issues that could be considered for future work. Data instances in real problems, and especially in environmental sciences, usually have certain implicit dependence among them. In this field, this is more common since data are collected from nearby areas that may be ecologically related. Although BNs learning requires independent data, they are mostly being applied to cases in which this restriction is not totally satisfied. In the current study, the fact that data are collected per municipality in a heterogeneous area alleviates this problem. However, it should be further investigated.

Another consideration is how the relationship between the three landscapes might somehow be modelled. This issue lends itself to the application of BNs-based multi-regression techniques where multiple predictions can be assigned to each data instance at the same time.

## Acknowledgements

This work has been supported by the Spanish Ministry of Science and Innovation through project TIN2010-20900-C04-02, by the Regional Ministry of Economy, Innovation and Science (Junta de Andalucía) through projects P08-RNM-03945 and P11-TIC-7821, and by ERDF funds. R.F. Ropero is being supported by the Spanish Ministry of Education, Culture and Sport through a FPU research grant, AP2012-2117. We would like to thank the anonymous reviewers for their valuable comments.

## Appendix A. Algorithms

## References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., Salmerón, A., 2011. Bayesian networks in environmental modelling. Environmental Modelling

**Algorithm 1:** MTE-NB regression model

**Input** : A database $D$ with variables $X_1, \ldots, X_n, Y$.
**Output**: A NB model with root variable $Y$ and features $X_1, \ldots, X_n$, with joint distribution of class MTE.

1 Construct a new network $\mathcal{G}$ with nodes $Y, X_1, \ldots, X_n$.
2 Insert the links $Y \rightarrow X_i$, $i = 1, \ldots, n$ in $\mathcal{G}$.
3 Estimate an MTE density for $Y$, and a conditional MTE density for each $X_i$, $i = 1, \ldots, n$ given $Y$.
4 Let $\mathcal{P}$ be the set of estimated densities.
5 Let NB be a Bayesian network with structure $\mathcal{G}$ and distributions $\mathcal{P}$.
6 **return** NB.

& Software 26, 1376–1388.

Aguilera, P. A., Fernández, A., Reche, F., Rumí, R., 2010. Hybrid Bayesian network classifiers: Application to species distribution models. Environmental Modelling & Software 25 (12), 1630–1639.

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.

Aranzabal, I. D., Schmitz, M. F., Aguilera, P. A., Pineda, F. D., 2008. Modelling of landscape changes derived from the dynamics of socio-ecological systems. A case in a semiarid Mediterraneam landscape. Ecological Indicators 8, 672–685.

Bicik, I., Jelecek, L., Stepanek, V., 2001. Land use changes and their social driving forces in Czechia in the 19th and 20th centuries. Land Use Policy 18(1), 65–73.

Bromley, J., Jackson, N. A., Clymer, O. J., Giacomello, A. M., Jensen, F. V., 2005. The use of Hugin$^R$ to develop Bayesian networks as an aid to integrated water resource planning. Environmental Modelling & Software 20, 231–242.

Burgi, M., Hersperger, A. M., Schneeberger, N., 2004. Driving forces of landscape change - current and new directions. Landscape Ecology 19, 857–868.

**Algorithm 2:** Selective MTE-NB regression model

**Input** : Variables $X_1, \ldots, X_n, Y$ and a database $D$ for variables $X_1, \ldots, X_n, Y$.

**Output**: Selective NB predictor for the variable $Y$.

```
for i := 1 to n do
    Compute Î(X_i, Y).
Let X_(1), ..., X_(n) be a decreasing order of the feature variables according to Î(X_i, Y).
Divide the database D into two sets, one for learning the model (D_l) and the other for testing its accuracy (D_t).
Using Algorithm 1, construct a NB model M with variables Y and X_(1) from database D_l.
Let rmse(M) be the estimated accuracy of model M using D_t.
for i := 2 to n do
    Let M_1 be the NB predictor obtained from Algorithm 1 for variables of M and X_(i).
    Let rmse(M_1) be the estimated accuracy of model M_1 using D_t.
    if rmse(M_1) ≤ rmse(M) then
        M := M_1.
return M.
```


Camarero, L., Cruz, F., González, M., del Pino, J. A., Oliva, J., Sampedro, R., 2009. La población rural de España. De los desequilibrios a la sostenibilidad social. Tech. rep., Obra Social. Fundación la Caixa.

Celio, E., Koellner, T., Grêt-Regamey, A., 2014. Modeling land use decisions with bayesian networks: Spatially explicit analysis of driving forces on land use change. Environmental Modelling & Software 52, 222–233.

Chen, S., Pollino, C., 2012. Good practice in bayesian network modelling. Environmental Modelling & Software 37, 134–145.

Cobb, B. R., Shenoy, P. P., Rumí, R., 2006. Approximating probability density functions with mixtures of truncated exponentials. Statistics and Computing 16, 293–308.

Cooper, G. F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9, 309–347.

Dechter, R., 1996. Bucket elimination: A unifying framework for probabilistic inference algorithms. In: Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence. pp. 211–219.

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 9.2. Environmental Systems Research Institute (ESRI), Redlands, CA.

Fernández, A., Morales, M., Salmerón, A., 2007. Tree augmented naïve Bayes for regression using mixtures of truncated exponentials: Applications to higher education management. IDA'07. Lecture Notes in Computer Science 4723, 59–69.

Fernández, A., Nielsen, J. D., Salmerón, A., 2010. Learning Bayesian networks for regression from incomplete databases. International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 18, 69–86.

Fernández, A., Salmerón, A., 2008. Extension of Bayesian network classifiers to regression problems. In: Geffner, H., Prada, R., Alexandre, I. M., David, N. (Eds.), Advances in Artificial Intelligence - IBERAMIA 2008. Vol. 5290 of Lecture Notes in Artificial Intelligence. Springer, pp. 83–92.

Frank, E., Trigg, L., Holmes, G., Witten, I. H., 2000. Technical note: Naive Bayes for regression. Machine Learning 41, 5–25.

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

Gámez, J. A., Salmerón, A., 2005. Predicción del valor genético en ovejas de raza manchega usando técnicas de aprendizaje automático. In: Actas de las VI Jornadas de Transferencia de Tecnología en Inteligencia Artificial. Paraninfo, pp. 71–80.

García-Latorre, J., Sánchez-Picón, A., 2001. Dealing with aridity: socio-economic structures and environmental changes in an arid Mediterranean region. Land Use Policy 18, 53–64.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer.

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.

Jiménez-Herrero, L. M., Álvarez-Uría, P., de la Cruz-Leiva, J. L., 2011. Biodiversidad en España. Base de la Sostenibilidad ante el Cambio Global. Observatorio de la Sostenibilidad en España.

Lacitignola, D., Petrosillo, I., Cataldi, M., Zurlini, G., 2007. Modelling socio-ecological tourism-based systems for sustainability. Ecological Modelling 206, 191–204.

Landuyt, D., Broekx, S., D'hondt, R., Engelen, G., Aertsens, J., Goethals, P. L., 2013. A review of bayesian belif networks in ecosystem service modelling. Environmental Modelling & Software 46, 1–11.

Langseth, H., Nielsen, T. D., Rumí, R., Salmerón, A., 2012. Mixtures of Truncated Basis Functions. International Journal of Approximate Reasoning 53 (2), 212–227.

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., Jensen, F., 2001. Stable local computation with conditional Gaussian distributions. Statistics and Computing 11, 191–203.

Lerner, U., Segal, E., Koller, D., 2001. Exact inference in networks with discrete children of continuous parents. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI-01). pp. 319–32.

Li, D., Yang, H. Z., Liang, X. F., 2013. Prediction analysis of wastewater treatment system using a bayesian network. Environmental Modelling & Software 40, 140–150.

Li, Z., D'Ambrosio, B., 1994. Efficient inference in Bayes networks as a combinatorial optimization problem. International Journal of Approximate Reasoning 11, 55–81.

Liao, Y., Wang, J., Guo, Y., Zheng, X., 2010. Risk assessment of human neural tube defects using a Bayesian belief network. Stochastic Environmental Research & Risk Assessment 24, 93–100.

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.

Lucas, P., 2002. Restricted Bayesian network structure learning. In: Proceedings of the 1st European Workshop on Probabilistic Graphical Models (PGM'02). pp. 217–232.

Malekmohammadi, B., Kerachian, R., Zahraie, B., 2009. Developing monthly operating rules for a cascade system of reservoirs: Application of Bayesian networks. Environmental Modelling & Software 24, 1420–1432.

Milne, E., Aspinall, R. J., Vldkamp, T. A., 2009. Integrated modelling of natural and social systems in land change science. Landscape Ecology 24, 1145–1147.

Minsky, M., 1963. Steps towards artificial intelligence. Computers and Thoughts, 406 – 450.

Moral, S., Rumí, R., Salmerón, A., 2001. Mixtures of Truncated Exponentials in Hybrid Bayesian Networks. In: Benferhat, S., Besnard, P. (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Vol. 2143 of Lecture Notes in Artificial Intelligence. Springer, pp. 156–167.

Moral, S., Rumí, R., Salmerón, A., 2003. Approximating conditional MTE distributions by means of mixed trees. ECSQARU'03. Lecture Notes in Artificial Intelligence 2711, 173–183.

Morales, M., Rodríguez, C., Salmerón, A., 2007. Selective naïve Bayes for regression using mixtures of truncated exponentials. International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 15, 697–716.

Nash, D., Waters, D., Buldu, A., Wu, Y., Lin, Y., Yang, W., Song, Y., Shu, J., Qin, W., Hannah, M., 2013. Using a conceptual bayesian network to investigate environmental management of vegetale production in the lake taihu region of china. Environmental Model 46, 170–181.

Nikodemus, O., Bell, S., Grine, I., Liepins, I., 2005. The impact of economic, social and politic factors on the landscape structure of the Vidzeme Uplands in Latvia. Landscape and Urban Planning 70(1/2), 57–67.

Norgaard, R. B., 1984. Coevolutionary development potential. Land Economics 60 (2), 160–173.

Nyberg, J. B., Marcot, B. G., Sulyma, R., 2006. Using Bayesian belief networks in adaptive management. Canadian Journal of Forest Research 36, 3104–3116.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems. Morgan-Kaufmann (San Mateo).

Pérez, A., Larrañaga, P., Inza, I., 2006. Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naïve Bayes. International Journal of Approximate Reasoning 43, 1–25.

Pérez, A., Larrañaga, P., Inza, I., 2009. Bayesian classifiers based on kernel density estimation: Flexible classifiers. International Journal of Approximate Reasoning 50 (2), 341–362.

Pérez-Miñana, E., Krause, P. J., Thornton, J., 2012. Bayesian Network fot the management of greenhouse gas emissions in the British agricultural sector. Environmental Modelling & Software 35, 132–148.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Romero, V., Rumí, R., Salmerón, A., 2006. Learning hybrid Bayesian networks using mixtures of truncated exponentials. International Journal of Approximate Reasoning 42, 54–68.

Ruiz, R., Riquelme, J., Aguilar-Ruiz, J. S., 2006. Incremental wrapper-based gene selection from microarray data for cancer classification. Pattern Recognition 39, 2383–2392.

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

Sahami, M., 1996. Learning limited dependence Bayesian classifiers. In: Second International Conference on Knowledge Discovery in Databases. pp. 335–338.

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.

Schmitz, M. F., Aranzabal, I. D., Aguilera, P. A., Rescia, A., Pineda, F., 2003. Relationships between landscape typology and socioeconomic structure. Scenarios of change in Spanish cultural landscapes. Ecological Modelling 168, 343–356.

Serra, P., Pons, X., Saurí, D., 2008. Land-cover and land-use change in a mediterranean landscape: A spatial analysis of driving forces integrating biophysical and human factors. Applied Geography 28(3), 189–209.

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.

Shenoy, P. P., West, J. C., 2011. Inference in hybrid Bayesian networks using mixtures of polynomials. International Journal of Approximate Reasoning 52 (5), 641–657.

Spirtes, P., Glymour, C., Scheines, R., 1993. Causation, prediction and search. Vol. 81 of Lecture Notes in Statistics. Springer Verlag.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological) 36 (2), 111–147.

Strijker, D., 2005. Marginal lands in Europe. Causes of decline. Basic and Applied Ecology 6, 99–106.

Sun, Z., Müller, D., 2013. A framework for modelling payments for ecosystem services with agent-based models, bayesian belief networks and opinion dynamics models. Environmental Modelling & Software 45, 15–28.

Turner, R. K., Lorenzoni, I., Beaumont, N., Bateman, I. J., Langford, I., McDonald, A., 1988. Coastal management for sustainable development: analysing environmental and socio-economic changes on the UK coast. Geographical Journal 164(3), 269–281.

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

Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling & Software 24, 1268–1281.

Wang, Y., Zhang, X., 2001. A dynamic modelling approach to simulating socioeconomic effects on landscapes changes. Ecological Modelling 140 (1-2), 141–162.

Webb, G. I., Boughton, J. R., Wang, Z., 2005. Not So Naive Bayes: Aggregating One-Dependence Estimators. Machine Learning 58, 5–24.

Witten, I. H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). Morgan Kaufmann.

Wu, J., Hobbs, R., 2002. Key issues and research priorities in landscape ecology: and idiosyncratic synthesis. Landscape Ecology 17, 335–365.

Zhang, N. L., Poole, D., 1996. Exploiting causal independence in Bayesian network inference. Journal of Artificial Intelligence Research 5, 301–328.