Multivariate Bioclimatic Indices Modelling: A Coregionalised Approach

Xavier BARBER, David CONESA, Antonio LÓPEZ-QUÍLEZ, and Javier MORALES

A methodological approach for modelling the spatial multivariate distribution of multiple bioclimatic indices is presented. The value of the indices is modelled by means of a Bayesian conditional coregionalised linear model. Elicitation of prior distributions and approximation of posterior distributions of the parameters in the proposed model are also discussed. A posterior predictive distribution and a spatial bioclimatic probability distribution for each bioclimatic index are obtained. This allows researchers to obtain the probability of each location belonging to different bioclimates. The presented methodology is applied in a practical setting showing that the spatial bioclimatic probability distributions are more realistic than the ones obtained in the univariate setting, while providing an interesting tool in the context of climate change.

Key Words: Bioclimatology; Coregionalised models; Multivariate Bayesian spatial models; Spatial prediction; Spatial bioclimatic probability distribution.

1. INTRODUCTION

Bioclimatology is the ecological science concerned with the interactions between climate and living organisms over an extended period of time. As far as plants are concerned, Bioclimatology can be used to establish better habitats and to manage plant resources and landscape more efficiently. In addition, it can be used to forecast the production of agricultural and forestry resources to combat hunger and to determine future vegetation scenarios.

In order to study the relationship between climate and vegetation, Rivas-Martínez (1994) and Rivas-Martínez and Rivas-Saenz (2017) have established the so-called Worldwide Bioclimatic Classification system (one of the most widely used classification systems). Interestingly, this classification system only requires readily available climate data, such as temperature and rainfall, from where bioclimatic indices are obtained by simple formulas.

Two examples of the use of this bioclimatic classification system are Peng (2000) and Monteiro-Henriques and Espírito-Santo (2011), where it is used to predict the consequences

Xavier Barber (✉) and Javier Morales, Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Elche, Spain (E-mail: *xbarber@umh.es*). David Conesa and Antonio López-Quílez, Dpt. Estadística i Investigació Operativa, Universitat de València, Valencia, Spain.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 24, Number 2, Pages 225–244
https://doi.org/10.1007/s13253-018-00345-z

of climate changes in species distribution. Other uses include managing and monitoring alien invasive species (Angetter et al. 2011), regional planning (Canu et al. 2015), exploring human influence in a particular area (del Arco et al. 2006; Conley et al. 2014), studying changes in landscape naturalness (Catorci et al. 2012), analysing the boundaries of bioclimatic units to support the limits of actual vegetation units (Barber et al. 2001), and studying the influence of climate on crop yield (Camps and Ramos 2012).

Bioclimatic classification systems assign a bioclimate (the basic unit of this classification) to any region of interest depending on the obtained value of the index. Consequently, having a good spatial representation of bioclimatic indices is important in describing the relationship between climate and the distribution of vegetation. Indeed, some of the above-mentioned bioclimatology studies use spatial statistics techniques, such as geostatistics (Robertson 1987; Rossi et al. 1992; Burrough 2001; Garzón-Machado et al. 2014).

Barber et al. (2017) described a univariate methodological approach for modelling the spatial distribution of bioclimatic indices. With this approach, one can also obtain the probability distribution of each bioclimatic index and the probability of each location belonging to different bioclimates. Their proposal is based on a hierarchical Bayesian geostatistical model for bioclimatic indices that incorporates both structured and unstructured random effects.

However, this approach does not take into account that spatial data are often multivariate. The fact that multiple bioclimatic indices can be measured at each spatial location clearly anticipates not only dependence between indices at each location but also association between indices across locations. Consequently, a multivariate geostatistical approach should be used to study the behaviour of bioclimatic indices [see Chiles and Delfiner (1999) and Wackernagel (2003) for multivariate geostatistics].

It is worth noting that even though there have been various approaches for performing multivariate geostatistics, all of them have faced the same problem, namely prediction over non-sampled locations. Among the different approaches that have been proposed to deal with this problem, we highlight here cross-variograms and cross-covariance functions (Lin 2002; Buttafuoco et al. 2005; Genton and Kleiber 2015), to use kriging with external drift and/or cokriging (Goovaerts 2000; Verfaillie et al. 2006), or the one in Cressie and Zammit-Mangion (2016) that proposes a conditional approach to multivariate spatial covariance models. In our case, we will focus here in the one introduced in Schmidt and Gelfand (2003), that is, a Bayesian coregionalisation approach for multivariate data based on intrinsic correlation. This modelling was previously described in Gelfand et al. (2010) and in Banerjee et al. (2014), and it is based on linear models of coregionalisation (Grzebyk and Wackernagel 1994; Wackernagel 2003).

Our interest here is to model the spatial multivariate distribution of multiple bioclimatic indices, and to do so, we propose to model them by means of a Bayesian conditional coregionalised linear model (Schmidt and Gelfand 2003). This selection is based on the easy interpretation of the parameters of the model and the fact that we can reparameterise the variance/covariance structure of the model (Yan et al. 2007), achieving so a more efficient estimation and prediction processes.

Two main results may be obtained from this modelling: the posterior predictive distribution of each bioclimatic index (clearly based on the information of the remaining indices) and, more importantly, the spatial bioclimatic probability distribution of each bioclimatic index, which allows researchers to obtain the probability of each location belonging to different bioclimates.

It is finally worth noting that, as is usual in Bayesian complex models, the resulting posterior distributions of the parameters have no closed expression, and so, numerical approaches are needed to approximate them. In our case, we present an implementation that considers MCMC methods to estimate the parameters, but also efficient tools that improve prediction performance. In particular, prediction is implemented using distributed programming, reducing thus computation time.

The remainder of this article is organised as follows. After this introduction, in Sect. 2, we describe the two bioclimatic indices that will be used throughout the paper. In Sect. 3, we present the Bayesian coregionalisation model for bioclimatic indices and introduce the reparameterisation for more efficient estimation and prediction. Section 4 provides hints about prior elicitation, as this is not easy in the present context. In Sect. 5, we explain how to perform inference and prediction for these indices. In Sect. 6, we apply this methodology in a real setting. In particular, we obtain the predictive distributions of two bioclimatic indices on the island of Cyprus, using the altitude and climate information (temperature and rainfall) from 775 spot elevations and 59 meteorological stations. Section 7 is devoted to compare the behaviour of the presented model with respect a modelling in which predictions are done independently for each index. The last section concludes the paper and presents lines for future research.

2. BIOCLIMATIC INDICES

The so-called Worldwide Bioclimatic Classification System by Rivas-Martínez and Rivas-Saenz (2017) is one of the most widely used Bioclimatic Classification Systems. It encompasses five macrobioclimates (Tropical, Mediterranean, Temperate, Boreal and Polar). These are further subdivided into twenty-seven bioclimates and five bioclimatic variants. This classification is based on bioclimatic indices, values obtained by simple mathematical expressions involving certain climatic parameters and factors commonly used to characterise the climate of a region. With this classification, it is possible to recognise climatically homogeneous areas that may have similar vegetation types (species or communities).

As already mentioned, the fact that we are concerned with multiple bioclimatic indices anticipates possible association between them. This is the key to the spatial multivariate distribution of multiple bioclimatic indices that will be presented in the following sections. As an example, we now introduce two relevant correlated indices, namely the Ombrothermic and the Continentality index. They will be used throughout the paper to illustrate our findings. We note that all results presented here could also be applied in any other bioclimatic index from any classification.

The Ombrothermic index (OI) relates rainfall and temperature in an area. It is defined as

$$\mathrm{OI} = \frac{10P}{T},$$

where P is the sum of the average rainfall (in mm) of the months whose average temperature is above zero degrees Celsius and T is the sum of monthly average temperatures above zero degrees Celsius, expressed in tenths of a degree.

The Continentality index (CI) is defined as the annual variation of temperature

$$\text{CI} = T_{\max} - T_{\min},$$

where $T_{\max}$ is the average temperature of the warmest month and $T_{\min}$ is the average temperature of the coldest month.

As these indices can be biased by rare climatic events, researchers usually consider the average of the indices over a period of at least 25 years.

3. MULTIVARIATE MODELLING FOR BIOCLIMATIC INDICES

In what follows, we present a method for modelling multiple bioclimatic indices based on the conditional coregionalisation approach (Schmidt and Gelfand 2003).

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be a set of m bioclimatic indices and $\boldsymbol{s} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$ a subset of n locations in a region. Then, $\boldsymbol{Y}(\boldsymbol{s}) = (\boldsymbol{Y}(\boldsymbol{s}_1)^{\mathrm{T}}, \boldsymbol{Y}(\boldsymbol{s}_2)^{\mathrm{T}}, \ldots, \boldsymbol{Y}(\boldsymbol{s}_n)^{\mathrm{T}})^{\mathrm{T}}$ is a multivariate vector with length mn representing the values of all indices at all locations. Moreover, let $\boldsymbol{X}(\boldsymbol{s}_i)$ be the vector of regressors (environmental and other characteristics) for the location i. The multivariate spatial regression model can then be defined as

$$\boldsymbol{Y}(\boldsymbol{s}_i) = \left[\boldsymbol{I}_m \otimes \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{s}_i)\right]\boldsymbol{\beta} + \boldsymbol{W}(\boldsymbol{s}_i) + \boldsymbol{\varepsilon}(\boldsymbol{s}_i), \tag{1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m)$ is the regression coefficients vector (being each $\boldsymbol{\beta}_j$ a vector with appropriate dimension depending of the number of regressors); $\boldsymbol{W}(\boldsymbol{s})$ represents the spatial effect with cross-covariance function $\boldsymbol{C}_{\boldsymbol{W}}(\cdot, \cdot)$); and, finally $\boldsymbol{\varepsilon}(\boldsymbol{s})$ stands for the uncorrelated error vector, whose distribution is given by $\varepsilon_j(s_i) \sim \mathcal{N}(0, \tau_j^2)$, and represents the small-scale variability.

The cross-covariance function is a matrix-valued function defined for any pairs of locations,

$$\boldsymbol{C}_{\boldsymbol{W}}(\boldsymbol{s}_i, \boldsymbol{s}_k) = \left[\text{Cov}\left(\boldsymbol{W}_j(\boldsymbol{s}_i), \boldsymbol{W}_l(\boldsymbol{s}_k)\right)\right]_{j,l=1}^{m}. \tag{2}$$

There are several cross-covariance functions for multivariate geostatistics. Genton and Kleiber (2015) provide a detailed review of cross-covariance functions, while Cressie and Zammit-Mangion (2016) focus in the conditional approach. In our case, we define this cross-covariance function through a correlation function ρ and a positive definite matrix $\boldsymbol{D}$,

$$\boldsymbol{C}_{\boldsymbol{W}}(\boldsymbol{s}_i, \boldsymbol{s}_k) = \rho(\boldsymbol{s}_i, \boldsymbol{s}_k)\boldsymbol{D}. \tag{3}$$

Resultingly, the covariance matrix of $\boldsymbol{W}$ becomes $\boldsymbol{\Sigma}_{\boldsymbol{W}} = \boldsymbol{H} \otimes \boldsymbol{D}$, where $\boldsymbol{H}_{ik} = \rho(\boldsymbol{s}_i, \boldsymbol{s}_k)$.

In order to model multivariate bioclimatic indices, we focus on the corregionalisation idea, in particular, the one proposed by Schmidt and Gelfand (2003), which takes advantage of the Bayesian approach. In particular, we note that the above-mentioned spatial model can be expressed hierarchically (Gelfand 2012) by means of the following structure:

$$
\begin{aligned}
(I)\quad & \boldsymbol{Y}(\boldsymbol{s})|\boldsymbol{\beta}, \boldsymbol{W}, \boldsymbol{\Psi} \sim \mathcal{N}_{mn}\left(\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{s})\boldsymbol{\beta} + \boldsymbol{W}(\boldsymbol{s}), \boldsymbol{\Psi} \otimes \boldsymbol{I}_n\right) \\
(II)\quad & \boldsymbol{W}(\boldsymbol{s})|\boldsymbol{\theta} \sim \mathcal{N}_{mn}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{Y}}(\boldsymbol{s})\right) \\
(III)\quad & p(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\theta}),
\end{aligned}
\tag{4}
$$

where $\boldsymbol{\Psi} = \mathrm{diag}(\tau_1^2, \ldots, \tau_m^2)$ and $\boldsymbol{\theta}$ are the vector of parameters of the correlation functions associated with $\boldsymbol{W}(\boldsymbol{s})$. We note that the third level requires the elicitation of the prior distribution for all parameters of the model. Henceforth, for simplicity, we will denote $\boldsymbol{Y}(\boldsymbol{s})$, $\boldsymbol{X}(\boldsymbol{s})$ and $\boldsymbol{W}(\boldsymbol{s})$ by $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{W}$, respectively.

This model above presented can be rewritten in order to add some flexibility while retaining interpretability and computational tractability (Banerjee et al. 2014). Two ideas are behind, the first one being a characteristic of coregionalisation, that is, the joint distribution of the indices can be obtained by the product of the conditional distributions

$$
\begin{aligned}
& p\left[\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m\right] \\
& = p\left[\boldsymbol{Y}_m|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m-1}\right] p\left[\boldsymbol{Y}_{m-1}|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m-2}\right] \cdots p\left[\boldsymbol{Y}_2|\boldsymbol{Y}_1\right] p\left[\boldsymbol{Y}_1\right].
\end{aligned}
$$

The second one is that the matrix $\boldsymbol{D}$ that appears in the cross-covariance function in (3) can be decomposed as $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$, being $\boldsymbol{A}$ a full rank lower triangular matrix. This latter assumption implies that the cross-covariance function in (3) is separable (Genton and Kleiber 2015). Based on this two assumptions, expression (4) becomes

$$
\begin{aligned}
(I)\quad & \begin{cases}
\boldsymbol{Y}_1 \sim \mathcal{N}_n\left(\left(\boldsymbol{I}_m \otimes \boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}_1}\right) \\
\boldsymbol{Y}_2|\boldsymbol{Y}_1 \sim \mathcal{N}_n\left(\left(\boldsymbol{I}_m \otimes \boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{\gamma}_2 + \alpha_{2|1}\boldsymbol{Y}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}_2}\right) \\
\vdots \\
\boldsymbol{Y}_m|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m-1} \sim \mathcal{N}_n\left(\left(\boldsymbol{I}_m \otimes \boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{\gamma}_m + \sum_{j=1}^{m-1}\left(\alpha_{m|j}\boldsymbol{Y}_j\right), \boldsymbol{\Sigma}_{\boldsymbol{Y}_m}\right)
\end{cases} \\
(II)\quad & p\left(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2, \tau_m^2, \boldsymbol{\theta}\right),
\end{aligned}
\tag{5}
$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m)$ corresponds to the regressors coefficients, the vector $\boldsymbol{\alpha} = (\alpha_{2|1}, \alpha_{3|1}, \alpha_{3|2}, \ldots, \alpha_{m|1}, \ldots, \alpha_{m|m-1})$ corresponds to the coefficients relating each bioclimatic index to the previous ones, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)$ is the vector of parameters of the correlation functions. As in expression (4), this new second level requires elicitation of the prior of all parameters.

Moreover, in order to make the coregionalised model in (5) equivalent to the multivariate model in (4), and in line with Banerjee et al. (2014), it is necessary to assume that the matrix of regressors is identical for all bioclimatic indices and also that the first $m-1$ bioclimatic indices are purely spatial effects (implying that $\tau_1 = \ldots = \tau_{m-1} = 0$). That is, we assume that the variance–covariance matrices for the m bioclimatic indices are defined as

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{Y}_j} &= \sigma_j^2 \boldsymbol{H}_j(\boldsymbol{\theta}_j); \qquad j = 1, \ldots, m-1; \\
\boldsymbol{\Sigma}_{\boldsymbol{Y}_m} &= \sigma_m^2 \boldsymbol{H}_m(\boldsymbol{\theta}_m) + \tau_m^2 \boldsymbol{I}_n,
\end{aligned}
\tag{6}
$$

where $\boldsymbol{\sigma}^{\mathbf{2}} = (\sigma_1^2, \ldots, \sigma_m^2)$ are the spatially structured variabilities of the bioclimatic indices, τ_m^2 is the non-structured spatial variability of the last bioclimatic index, and $\boldsymbol{H}_j(\boldsymbol{\theta}_j)$ represents the correlation matrix between locations.

A flexible family of correlation matrices that generalises many of the widely used covariance models in spatial statistics is due to Matérn (1986). Even though other parameterisations are possible for this family, our choice is that proposed by Handcock and Wallis (1994), in which the isotropic correlation function has the following form

$$\rho(h, \phi, \nu) = 2^{\mu-1} \left(2h\nu^{1/2}/\phi\right)^{\nu} K_\nu \frac{\left(2h\nu^{1/2}/\phi\right)}{\Gamma(\nu)},$$

where $\boldsymbol{h} = \|\boldsymbol{s}_i - \boldsymbol{s}_k\|$ and K_ν is the modified Bessel function of the second kind and order $\nu > 0$ (Abramowitz and Stegun 1964, section 9.6). The correlation matrix $\boldsymbol{H}_j(\boldsymbol{\theta}_j)$ depends on two parameters: ϕ_j and ν_j. ϕ_j is the range parameter, which controls the rate of decay of the correlation between observations as distance increases. ν_j controls the behaviour of the correlation function for observations that are separated by small distances. Our choice for this parameterisation is mainly based on the fact that the interpretation of ϕ_j is largely independent of ν_j, reducing the posterior correlation between them.

The model in (5) provides the joint distribution of the bioclimatic indices via the conditional distributions. However, following Yan et al. (2007), in order to avoid the identifiability problem of spatial and non-spatial variability, we reparameterise the spatially structured variability of each bioclimatic index, σ_j, and the non-structured spatial variabilities of each bioclimatic index, τ_j, using two new parameters, namely κ_j and ξ_j^2. In particular, $\xi_j^2 = \sigma_j^2 + \tau_j^2$ now represents the total variability of the random effects, whereas $\kappa_j = \tau_j^2/\xi_j^2$ stands for the proportion of the non-spatial variability to the total variability.

Taking into account that the $m-1$ first bioclimatic indices should be purely spatial effects, the variances κ_j and ξ_j^2 can be written as

$$\begin{aligned} &\xi_j^2 = \sigma_j^2,\ j = 1, \ldots, m-1; \quad \xi_m^2 = \sigma_m^2 + \tau_m^2; \\ &\kappa_j = 0,\ j = 1, \ldots, m-1; \quad \kappa_m = \tau_m^2/\left(\sigma_m^2 + \tau_m^2\right). \end{aligned} \tag{7}$$

This allows us to rewrite the variance–covariance matrices in (6) as

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{Y}j} &= \xi_j^2 \boldsymbol{H}_j(\boldsymbol{\theta}_j); \quad j = 1, \ldots, m-1; \\ \boldsymbol{\Sigma}_{\boldsymbol{Y}m} &= \xi_m^2 \left[(1-\kappa_m)\boldsymbol{H}_m(\boldsymbol{\theta}_m) + \kappa_m \boldsymbol{I}_n\right]. \end{aligned} \tag{8}$$

Even though the changes in model (5) do not seem relevant, the advantage appears when elicitating the prior distribution on the second level of the model $p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta})$. The natural conjugate prior for a multivariate normal distribution is the inverse Wishart distribution, which is multivariate. By contrast, with the coregionalised model, we need only specify the prior distributions for the structured variability of each bioclimatic index, ξ_j^2. For the proportion of non-structured variability, κ_m, associated with the conditional distribution of $\boldsymbol{Y}_m$, we specify all univariate distributions.

4. SELECTION OF PRIOR DISTRIBUTIONS

Since we are using the Bayesian paradigm, it is necessary to elicit the prior distribution of the vector of parameters involved in the model. When it is possible, we always make use of previous information based on expert opinion. In situations where there is no expert opinion, our choice is to use non-informative prior distributions.

A usual assumption for selecting the prior is to consider independence of the parameters:

$$p(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta}) = p(\boldsymbol{\gamma}) p(\boldsymbol{\alpha}) p(\boldsymbol{\xi}^2) p(\kappa_m) p(\boldsymbol{\theta}).$$

This assumption allows us to express our knowledge for each of these parameters independently. To this end, we must provide both their distributions and the values of their hyperparameters. Usual choices for the distribution of $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ under a complete ignorance premise are either Gaussian distributions or non-informative improper distributions. In our case, for mathematical simplicity, we work with the latter assuming independence of all the components, that is,

$$p(\boldsymbol{\gamma}) = p(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m) = \prod_{i=1}^{m} p(\boldsymbol{\gamma}_i) \propto 1,$$

$$p(\boldsymbol{\alpha}) = p(\alpha_{2|1}, \alpha_{3|1}, \alpha_{3|2}, \ldots, \alpha_{m|1}, \ldots, \alpha_{m|m-1}) = \prod_{i=2}^{m} \prod_{j=1}^{i} p(\alpha_{i|j}) \propto 1\,.$$

In line with this, for the proportion parameter κ_m, a natural choice under a complete ignorance setting is to use a uniform distribution between 0 and 1, that is, $\kappa_m \sim \mathcal{U}(0, 1)$.

As for the correlation parameters for the Matérn function, following Stein (1999) and Finley et al. (2015), we use a uniform prior distribution for ϕ_j defined by

$$p(\phi_j) = \mathcal{U}\left(\frac{1}{d_1}, \frac{1}{d_2}\right),$$

where d_1 (resp., d_2) is the maximum (resp., minimum) distance between two locations. Moreover, the selection for the prior of the smoothing parameter is again a uniform distribution: $\nu_j \sim \mathcal{U}(0.05, 1.95)$. As mentioned above, these two parameters are largely independent.

Finally, following Barber et al. (2017), information about ξ_j^2 can be incorporated in the scale parameters of different distributions by considering that the observed values of the index on a set of locations is a priori uniformly distributed between $\left(Y_j^{\min}, Y_j^{\max}\right)$, being $Y_j^{\max}$ and $Y_j^{\min}$ the highest and lowest value of the index in the region of study (according the Rivas-Martinez classification). The corresponding variability of this uniform distribution (the most disadvantageous option, as this would imply that all the regions have the same orographic features) is

$$\mathrm{Var}(Y_j) = \frac{\left(Y_j^{\max} - Y_j^{\min}\right)^2}{12}, \tag{9}$$

the maximum value of which (denoted by $V_j^{\max}$) would be an upper bound of the variability index. Then, a prior distribution is constructed by matching the range of variability $\left(a, V_j^{\max}\right)$ with the quantile 0.95 of any chosen distribution (Chambers and Dunstan 1986; Strupczewski et al. 2007). That is,

$$0.95 = \int_a^{V_j^{\max}} f(y|\boldsymbol{\delta})\mathrm{d}y, \tag{10}$$

where f is the chosen prior distribution and $\boldsymbol{\delta}$ the vector of its corresponding parameters. Since variability is always positive, a can be chosen to be as small as possible (in our case, we work with $a = 0.001$). Usual options for the prior f along with their corresponding scale parameters are: $p(\xi^2) \sim \mathcal{U}(0.001, b)$, uniform over the variance; $p(\xi) \sim \mathcal{U}(0.001, \sqrt{b})$, uniform over the standard deviation; an Inverse Gamma, $p(\xi^2) \sim \mathcal{IG}(2, \beta)$; or a half-Cauchy, $p(\xi^2) \sim \mathcal{HC}(\delta)$. Our choice, the second one with scale parameter $b = \frac{V_{\max} - 0.00005}{0.95}$, is based mainly on the easiness for implementing it. Moreover, it is worth mentioning that a sensitivity analysis for these priors showed that the posterior is not too sensitive to the prior choice.

5. INFERENCE AND PREDICTION

Once we have defined the coregionalised model and shown how to select the prior distributions, the Bayesian learning process is applied to obtain the posterior distribution

$$p\left(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta} | \boldsymbol{Y}, \boldsymbol{X}\right) \propto \ell\left(\boldsymbol{Y}|\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta}, \boldsymbol{X}\right) p\left(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta}\right), \tag{11}$$

where the likelihood is

$$\begin{aligned}
\ell\left(\boldsymbol{Y}|\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\xi}^2, \kappa_m, \boldsymbol{\theta}, \boldsymbol{X}\right) &= \mathcal{N}_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}1}\right) \times \\
&\quad \cdots \\
&\quad \times \mathcal{N}_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_j + \alpha_{2|1}\boldsymbol{Y}_1 + \cdots + \alpha_{j|j-1}\boldsymbol{Y}_{j-1}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_j}\right) \times \\
&\quad \cdots \\
&\quad \times \mathcal{N}_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_m + \sum_{j=1}^{m-1}\left(\alpha_{m|j}\boldsymbol{Y}_j\right), \boldsymbol{\Sigma}_{\boldsymbol{Y}_m}\right).
\end{aligned}$$

This posterior distribution contains all our knowledge about the indices. However, as is usual in complex spatial models, it does not have an analytic expression. Therefore, numerical approximations are needed. Interestingly, the fact that we have analytical expressions of the conditional distributions allows us to use Gibbs sampling, one of the most widely used options of Markov Chain Monte Carlo techniques (Gamerman and Lopes 2006).

In order to present the conditionals, we introduce a slightly different notation for the parameters of the model. In particular if $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_m)$ represents the complete parameter vector, where

$$\begin{aligned}
\boldsymbol{\Delta}_1 &= \left(\boldsymbol{\gamma}_1, \xi_1^2, \boldsymbol{\theta}_1\right) \\
\boldsymbol{\Delta}_j &= \left(\boldsymbol{\gamma}_j, \alpha_{j|j-1}, \ldots, \alpha_{2|1}, \xi_j^2, \boldsymbol{\theta}_j\right) \quad \text{for } j = 2, \ldots, m-1 \\
\boldsymbol{\Delta}_m &= \left(\boldsymbol{\gamma}_m, \alpha_{m|m-1}, \ldots, \alpha_{2|1}, \xi_m^2, \kappa_m, \boldsymbol{\theta}_m\right),
\end{aligned}$$

the conditional distribution of $\boldsymbol{\Delta}_j$ given $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{j-1}$ for $j = 1, \ldots, m$ is

$$\begin{aligned}
p(\boldsymbol{\Delta}_1|\boldsymbol{Y}_1, \boldsymbol{X}) &\propto N_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}1}\right) p(\boldsymbol{\Delta}_1) \\
p(\boldsymbol{\Delta}_2|\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{X}) &\propto N_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_2 + \alpha_{2|1}\boldsymbol{Y}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}2}\right) p(\boldsymbol{\Delta}_2) \\
&\vdots \\
p(\boldsymbol{\Delta}_m|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{m-1}, \boldsymbol{Y}_m, \boldsymbol{X}) &\propto N_n\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\gamma}_m + \sum_{j=1}^{m-1}\left(\alpha_{m|j}\boldsymbol{Y}_j\right), \boldsymbol{\Sigma}_{\boldsymbol{Y}m}\right) p(\boldsymbol{\Delta}_m).
\end{aligned} \tag{12}$$

Although other options could have been used to perform MCMC, in our case, we work with WinBUGS (Lunn et al. 2000), a flexible software application for performing the Bayesian analysis of complex statistical models that also provides great flexibility in specifying the variance–covariance matrix given at the first level of hierarchy. Nevertheless, the standard Gibbs and Metropolis algorithms for computing posterior simulations from hierarchical models can have convergence problems due to the redundant parameters (that involve reparameterisation and overparameterisation). Gelman et al. (2008) provide alternative strategies that can be used in these cases.

It is also worth noting that the fact that the model in (12) can include covariates implies the use of a model selection procedure in order to check their relative relevance. Although other procedures could be used, this selection process could be done by means of the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002).

As we are within the Bayesian paradigm, prediction is reduced to obtaining the posterior predictive distributions. In particular, if $\boldsymbol{Y}^p$ represents the values of a vector of m bioclimatic indices in a new set of locations with observed covariates $\boldsymbol{X}^p$, then the posterior predictive distribution of the new values $\boldsymbol{Y}^p$ (conditional to the observed ones, henceforth, $\boldsymbol{Y}^o$) is

$$p\left(\boldsymbol{Y}^p|\boldsymbol{Y}^o, \boldsymbol{X}^o, \boldsymbol{X}^p\right) = \int p(\boldsymbol{Y}^p|\boldsymbol{X}^p, \boldsymbol{\Delta})p(\boldsymbol{\Delta}|\boldsymbol{Y}^o, \boldsymbol{X}^o)d\boldsymbol{\Delta}. \tag{13}$$

A simulated sample from this multivariate posterior predictive distribution can be obtained via the composition method taking into account that

$$\begin{aligned}
p\left(\boldsymbol{Y}^p|\boldsymbol{X}^p, \boldsymbol{\Delta}\right) &= p\left(\boldsymbol{Y}_1^p|\boldsymbol{X}^p, \boldsymbol{\Delta}_1\right) \times p\left(\boldsymbol{Y}_2^p|\boldsymbol{X}^p, \boldsymbol{Y}_1^p, \boldsymbol{\Delta}_2\right) \times \cdots \times p\left(\boldsymbol{Y}_m^p|\boldsymbol{X}^p, \boldsymbol{Y}_1^p, \ldots, \boldsymbol{Y}_{m-1}^p, \boldsymbol{\Delta}_m\right) \\
&= \mathcal{N}_n\left(\left(\boldsymbol{X}^p\right)^{\mathrm{T}}\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}_{\boldsymbol{Y}_1}\right) \times \cdots \times \mathcal{N}_n\left(\left(\boldsymbol{X}^p\right)^{\mathrm{T}}\boldsymbol{\gamma}_m + \alpha_{2|1}\boldsymbol{Y}_1^p + \cdots + \alpha_{m|m-1}\boldsymbol{Y}_{m-1}^p, \boldsymbol{\Sigma}_{\boldsymbol{Y}_m}\right),
\end{aligned}$$

with $\boldsymbol{\Sigma}_{\boldsymbol{Y}_j}$ given in (8).

This method, even though it is effective, can be computationally expensive when the number of covariates and/or new locations to predict is large. A good option to reduce this computational cost is the use of distributed computation that allows us to increase performance when executing matrix calculations. In fact, parallelisation allows us to reduce computation time by nearly 80% compared with sequential approaches (226.5 s per simulation using *R* matrix computations; 147 s using spBayes; and 17 s using a non-uniform memory access parallel system with 20 cores). This clearly helps if one needs to work with a large number of new locations using a large number of samples from the posterior distribution.

As stated in Barber et al. (2017), the most valuable information one can obtain from the predictive distributions is the probability of each location belonging to different bioclimates. One method of presenting these probabilities is via a discrete probability distribution for each location, which is called "spatial bioclimatic probability distribution." This is represented by a single figure consisting of different graphs showing the probability of belonging to each bioclimate.

6. MULTIVARIATE BIOCLIMATIC CLASSIFICATION OF THE ISLAND OF CYPRUS

In what follows, we illustrate the use of the previously presented approach through an application that analyses the bioclimatic classification in Cyprus, an island country in the Mediterranean, the flora of which is relatively rich and diverse. In fact, there are about 1800 taxons (species, subspecies and varieties), 132 of which are considered to be endemic. Notably, 19 plant species (18 endemics and 1 rare) are included on the list of the strictly protected plants of the Bern Convention, which was signed by the Cyprus Government. This number represents one sixth of all protected plants in Europe.

We use the information gathered from 59 meteorological stations distributed all over the island. This information was provided by the Meteorological Service of the Ministry of Agriculture, Natural Resources and Environment. The Service also provided meteorological data on mean monthly and annual precipitation as well as temperature (mean daily maximum and minimum, mean monthly maximum and minimum, etc.). Moreover, the coordinates and altitude for each meteorological station were provided.

As shown in Fig. 1, there is a strong relationship between the Ombrothermic index and altitude. Moreover, there is a relationship between the logarithm (a transformation used to achieve linearity) of the Ombrothermic index (LOI) and the Continentality index (CI) after adjusting by altitude. Based on these two facts, we propose a joint analysis of both indices by means of the coregionalised linear model introduced above including altitude as a covariate (that express a large-scale effect) and considering the logarithm of the Ombrothermic index to be a purely spatial effect. The reason underneath this ordering (and not the one relating LOI respect to CI) is merely climatic: there could be different Ombrotypes for a single value of Continentality, but not the other way around (Rivas-Martínez and Rivas-Saenz 2017).

Particularising Eq. (5), the conditional coregionalised model for CI and LOI becomes

$$\begin{aligned} &\text{(I)} \quad \begin{cases} \boldsymbol{Y}_{\text{LOI}} \sim \mathcal{N}_n\left(\boldsymbol{X}^{\text{T}}\boldsymbol{\gamma}_{\text{LOI}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_{\text{LOI}}}\right) \\ \boldsymbol{Y}_{\text{CI}}|\boldsymbol{Y}_{\text{LOI}} \sim \mathcal{N}_n\left(\boldsymbol{X}^{\text{T}}\boldsymbol{\gamma}_{\text{CI}} + \alpha_{\text{CI}|\text{LOI}}\boldsymbol{Y}_{\text{LOI}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}_{\text{CI}}}\right) \end{cases} \\ &\text{(II)} \quad p\left(\boldsymbol{\gamma}_{\text{LOI}}, \boldsymbol{\gamma}_{\text{CI}}, \alpha_{\text{CI}|\text{LOI}}, \xi^2_{\text{LOI}}, \xi^2_{\text{CI}}, \kappa_{\text{CI}}, \boldsymbol{\theta}_{\text{LOI}}, \boldsymbol{\theta}_{\text{CI}}\right), \end{aligned} \tag{14}$$

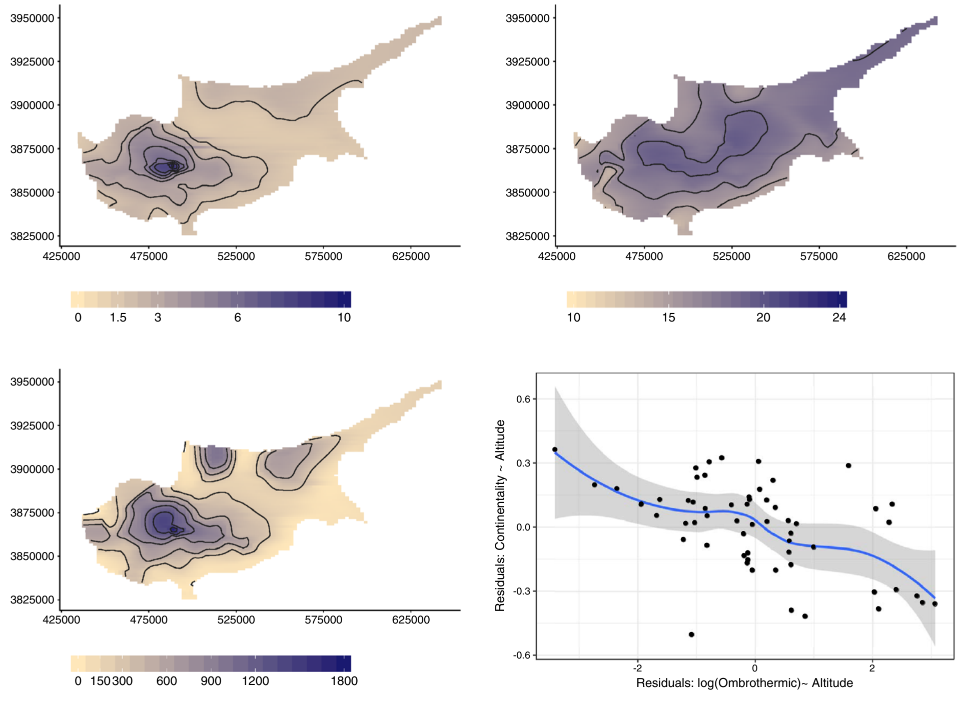

Figure 1. Spatial distribution of the Ombrothermic index (upper left) and Continentality index (upper right) obtained by means of a spline interpolation using the meteorological information gathered from the 59 stations. Spatial distribution of the altitude obtained similarly (down left). Relationship between the residuals of Continentality and the logarithm of the Ombrothermic index after adjusting by altitude (down right).

where the corresponding variance–covariance matrices are given by

$$\begin{aligned}\boldsymbol{\Sigma}_{Y_{\mathrm{LOI}}} &= \xi^2_{\mathrm{LOI}}\boldsymbol{H}(\boldsymbol{\theta}_{\mathrm{LOI}})\\ \boldsymbol{\Sigma}_{Y_{\mathrm{CI}}} &= \xi^2_{\mathrm{CI}}\left[(1-\kappa_{\mathrm{CI}})\boldsymbol{H}(\boldsymbol{\theta}_{\mathrm{CI}}) + \kappa_{\mathrm{CI}}\boldsymbol{I}_n\right],\end{aligned}$$

with $\boldsymbol{\theta}_{\mathrm{LOI}} = (\phi_{\mathrm{LOI}}, \nu_{\mathrm{LOI}})$ and $\boldsymbol{\theta}_{\mathrm{CI}} = (\phi_{\mathrm{CI}}, \nu_{\mathrm{CI}})$. In this model, ξ^2_{LOI} and $\xi^2_{\mathrm{CI}}(1-\kappa_{\mathrm{CI}})$ represent the structured variability associated with each index, whereas $\xi^2_{\mathrm{CI}}\kappa_{\mathrm{CI}}$ stands for the spatially non-structured variability associated with the Continentality index. More importantly, $\alpha_{\mathrm{CI|LOI}}$ reflects the influence of the Ombrothermic index on the Continentality index.

The resulting posterior distribution of the parameters (denoted by $\boldsymbol{\Delta}$) is given by:

$$\begin{aligned}p(\boldsymbol{\Delta}|\boldsymbol{Y}_{\mathrm{CI}}, \boldsymbol{Y}_{\mathrm{LOI}}, \boldsymbol{X}) &\propto p(\boldsymbol{Y}_{\mathrm{CI}}, \boldsymbol{Y}_{\mathrm{LOI}}|\boldsymbol{\Delta}, \boldsymbol{X})p(\boldsymbol{\Delta})\\ &\propto p(\boldsymbol{Y}_{\mathrm{CI}}, \boldsymbol{Y}_{\mathrm{LOI}}|\boldsymbol{\Delta}_{\mathrm{CI}}, \boldsymbol{\Delta}_{\mathrm{LOI}}, \boldsymbol{X})p(\boldsymbol{\Delta}_{\mathrm{CI}}, \boldsymbol{\Delta}_{\mathrm{LOI}})\\ &\propto p(\boldsymbol{Y}_{\mathrm{CI}}|\boldsymbol{Y}_{\mathrm{LOI}}, \boldsymbol{X}, \boldsymbol{\Delta}_{\mathrm{CI}})p(\boldsymbol{\Delta}_{\mathrm{CI}})p(\boldsymbol{Y}_{\mathrm{LOI}}|X, \boldsymbol{\Delta}_{\mathrm{LOI}})p(\boldsymbol{\Delta}_{\mathrm{LOI}})\end{aligned}$$

where, based on the hints given in Sect. 4, we use the following product of priors for the parameters $\boldsymbol{\Delta}_{\mathrm{LOI}}$

$$\begin{aligned}&p(\boldsymbol{\gamma}_{\mathrm{LOI}}, \xi_{\mathrm{LOI}}, \phi_{\mathrm{LOI}}, \nu_{\mathrm{LOI}})\\ &\quad\propto 1 \times \mathcal{U}(0.001, \sqrt{0.54}) \times \mathcal{U}(1/100000, 1/1818) \times \mathcal{U}(0.05, 1.95),\end{aligned}$$

Table 1. Mean and standard deviation of the posterior distribution and 95% credible intervals of the parameters for the joint Ombrothermic (LOI) and Continentality (CI) using a coregionalisation approach.

	Mean	SD	$q_{2.5}$	$q_{97.5}$
$\gamma_{0,\,\mathrm{LOI}}$	0.538	0.0696	0.406	0.676
$\gamma_{1\,,\mathrm{LOI}}$	0.000718	0.0000806	0.000561	0.000879
ξ^2_{LOI}	0.0466	0.0199	0.0288	0.0972
ϕ_{LOI}	0.00011	0.000024	0.000068	0.00016
ν_{LOI}	1.47	0.275	1.03	1.93
$\gamma_{0,\mathrm{CI}\mid\mathrm{LOI}}$	16.8	0.587	15.7	18.0
$\gamma_{1,\,\mathrm{CI}\mid\mathrm{LOI}}$	0.00435	0.00109	0.00217	0.00652
$\alpha_{\mathrm{CI}\mid\mathrm{LOI}}$	− 2.21	0.996	− 4.14	− 0.209
$\xi^2_{\mathrm{CI}\mid\mathrm{LOI}}$	1.61	0.563	1.04	3.11
$\kappa_{\mathrm{CI}\mid\mathrm{LOI}}$	0.163	0.146	0.0191	0.568
ϕ_{CI}	0.00017	0.000095	0.000069	0.00046
ν_{CI}	1.46	0.333	0.62	1.93

and the following product for the parameters $\boldsymbol{\Delta}_{\mathrm{CI}}$

$$
\begin{aligned}
&p\left(\boldsymbol{\gamma}_{\mathrm{CI}}, \alpha_{\mathrm{CI}\mid\mathrm{LOI}}, \xi_{\mathrm{CI}}, \kappa_{\mathrm{CI}}, \phi_{\mathrm{CI}}, \nu_{CI}\right) \\
&\quad \propto 1 \times 1 \times \mathcal{U}(0.001, \sqrt{8.77}) \times \mathcal{U}(0, 1) \times \mathcal{U}(1/100000, 1/1818) \times \mathcal{U}(0.05, 1.95),
\end{aligned}
$$

where, in line with Barber et al. (2017), the prior for ξ in both cases is a Uniform distribution whose upper value is obtained by solving Eq. (10).

Table 1 presents the mean of the posterior distribution of the parameters of the model and their corresponding 95% credible intervals. These posterior distributions were obtained by simulation using WinBUGS (Lunn et al. 2000). Each posterior distribution was approximated from 15,000 (5000 from each of three simulation chains) simulated values (obtained after discarding 10000 simulations from a burn-in period that ensured convergence).

It can be seen that both indices are related to altitude (both $\gamma_{1,\mathrm{LOI}}$ and $\gamma_{1,\mathrm{CI}\mid\mathrm{LOI}}$ are different from 0). Moreover, the small values of the posterior distribution of the proportion of the non-spatial variability to the total variability, $\kappa_{\mathrm{CI}\mid\mathrm{LOI}}|$data, indicate that the Continentality index has a strong spatial behaviour. Moreover, the negative values of the coefficient that relates the Continentality index to the logarithm of the Ombrothermic index, $\alpha_{\mathrm{CI}\mid\mathrm{LOI}}$, indicate that the relationship between both indices is negative, in line with what can be seen in Fig. 1. Consequently, the use of a coregionalised spatial model is necessary. As it can be seen in Table 1, the posterior mean of ϕ_{LOI} is 0.00011, that corresponds to an effective range of 9 km, while the posterior mean of ϕ_{CI} is 0.00017, corresponding to an effective range of 6 km. Finally, it is also worth noting that as usual in this context, identifiability of the parameter ν is not completely achieved, as shown by a large credible interval.

After obtaining the posterior distribution of the parameters, we perform prediction of both indices over the entire island. To this end, we use the altitude of 775 spot elevations and the information from 59 meteorological stations. Using distributed computation, obtaining posterior predictive distributions takes around 45–55 min (it could last more than 6 h without this approach).

Figure 2 shows the mean and standard deviation of the posterior predictive distribution of the Ombrothermic index and the mean and standard deviation of the posterior predictive distribution of the Continentality index. Notably, both figures show the topography of the island and the lack of information in the north of the island (in accordance with the high standard deviation of the posterior predictive).

As commented above, the most valuable information one can obtain from the predictive distributions is the probability of each location belonging to different bioclimates. Figure 3 represents the spatial bioclimatic probability distribution for the Ombrothermic and the Continentality index. This figure consists of different graphs showing the probability of belonging to different bioclimates. This corresponds to the probability of the index to be between the values stated in Rivas-Martínez and Rivas-Saenz (2017) for each bioclimate. Notably, this figure provides the boundaries between the types of bioclimates for each index. These boundaries define the areas that in the future may change from one bioclimate to another. This is useful in studying potential climate changes and their effects on the vegetation of a region.

The results for the Ombrothermic index (Fig. 3) show that most of the island is included in the arid bioclimate subtype. The central mountain is in the semi-arid interval. However, some plains belong to the hyper-arid subtype. Only the highest peak of the island belongs to the dry subtype. More importantly, these distributions fit the landscape change gradually, demonstrating the usefulness of these bioclimatic boundaries. As far the Continentality index (Fig. 3) is concerned, we note that the interior and the coast of the island are clearly shown as two different continental subtypes. Climate changes can be seen in smooth gradient areas, providing so an interesting tool in the context of climate change.

7. A COMPARISON OF UNIVARIATE VERSUS MULTIVARIATE BIOCLIMATIC CLASSIFICATION OF THE ISLAND OF CYPRUS

In order to show the relevance of the proposed model, we now present a detailed comparison of the results obtained when fitting both a multivariate (coregionalised) model and an univariate analysis (the one that assumes independence) of the dataset presented in the paper. In particular, we focus here in comparing the estimation obtained for the parameters governing both modellings, and differences between the probabilities of belonging to different climate subtypes. Finally, we also include a leave-one-out cross-validation to show the improvement of the multivariate approach.

Table 2 displays, for the logarithm of the Ombrothermic index (LOI), the posterior mean and the credible interval of the parameters in both modellings. As it can be seen, the results for the fixed effects are similar, while only slightly differences appear in the correlation function parameters: ϕ, ν and ξ^2.

Differences between both modellings can be appreciated in the same Table, that shows for the Continentality index (CI) the posterior mean and credible intervals of the parameters in both modellings. Similarly to LOI, there are not big differences in the fixed effects parameters. But note that the total variability ξ^2 has decreased considerably when using the

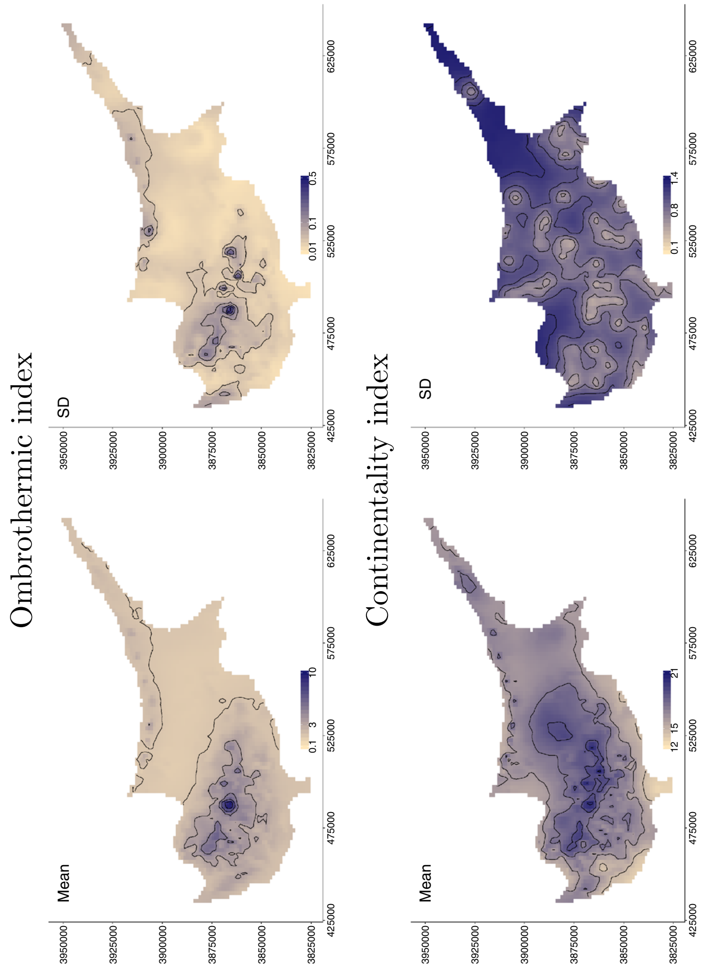

Figure 2. Mean (left) and standard deviation (right) of the posterior predictive distribution of the Ombrothermic and Continentality Index.

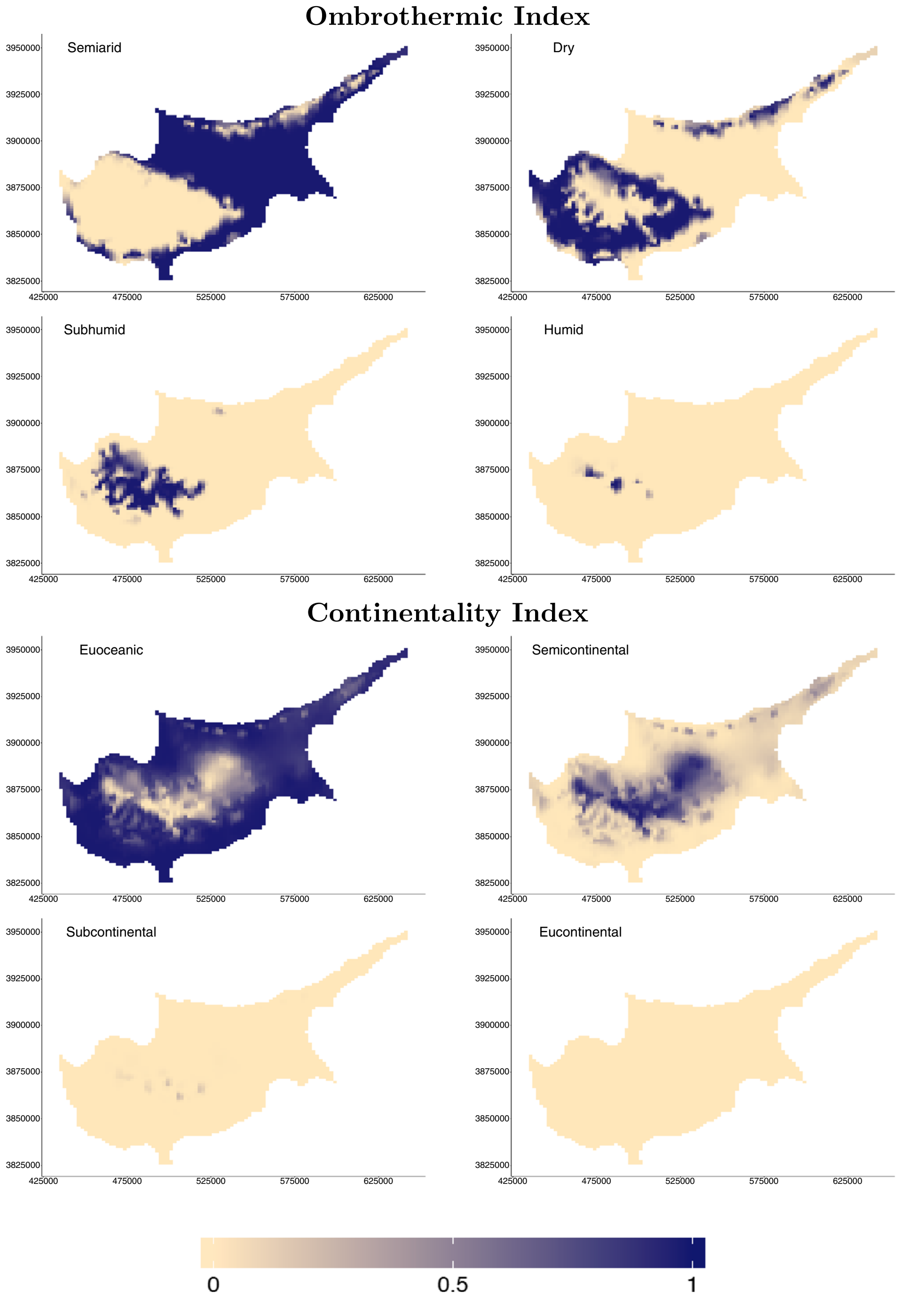

Figure 3. Spatial bioclimatic probability distribution of the threshold of Ombrotype and Continentality horizon values based on the indices.

Table 2. Comparison of the posterior mean and 95% credible intervals of the parameters in both univariate and multivariate modellings for the Ombrothermic and for Continentality index.

	Univariate LOI			Multivariate LOI		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%
γ_0	0.53905	0.40120	0.68741	0.53876	0.33139	0.68681
γ_1	0.00072	0.00056	0.00088	0.00072	0.00056	0.00088
ϕ	0.00011	0.00006	0.00016	0.00011	0.00007	0.00016
ν	1.47488	1.04652	1.90247	1.47507	1.02601	1.92702
ξ^2	0.02916	0.02973	0.09245	0.04660	0.02884	0.09707
	Univariate CI			Multivariate CI|LOI		
	Mean	2.50%	97.50%	Mean	2.50%	97.50%
γ_0	15.59666	13.17928	16.79024	16.84249	15.70436	18.00025
γ_1	0.00218	0.00099	0.00334	0.00435	0.00217	0.00652
ϕ	0.00009	0.00002	0.00029	0.00017	0.00007	0.00046
ν	1.47328	1.02092	1.92507	1.46572	0.62749	1.92523
κ	0.21241	0.04405	0.52121	0.16328	0.01909	0.57342
ξ^2	3.23533	1.25901	7.96605	1.61045	1.04097	3.11005

multivariate approach. This happens because the CI takes benefit of the information present in the LOI.

The usefulness of the multivariate modelling can also be appreciated when observing the mean squared prediction error from a leave-one-out cross-validation method. Note that the mean squared prediction error decreases for the CI when using a multivariate approach (from 2.38 in the Univariate case to 1.17 in the multivariate). Interestingly, the results for the LOI are opposite, as the mean squared prediction error increases when using a multivariate approach (from 0.18 in the univariate case to 0.32 in the multivariate). This latter result is the expected one, as in the multivariate case the LOI is modelled with a purely spatial effect. Indeed, the advantages of the multivariate approach appear when predicting the CI jointly with the LOI. If the interest is to predict the LOI alone, the best option should be to use the univariate approach.

Finally, Fig. 4 shows the spatial bioclimatic probability distribution of a particular Ombrotype (*dry*) and one Continentality subtype (*Semicontinental*) based on the results obtained when using an univariate approach and a multivariate approach. Similarly as above, there are no differences for the Ombrothermic index, although differences appear when comparing the results for the Continentality index, as results are more realistic from a bio-climatological point of view. The reason underneath is that the slopes indicating the changes between locations with probability 0 and probability 1 are smoother in the multivariate case.

8. CONCLUDING REMARKS

This work presented a methodological approach for modelling the spatial multivariate distribution of multiple bioclimatic indices with the aim of establishing a bioclimatic clas-

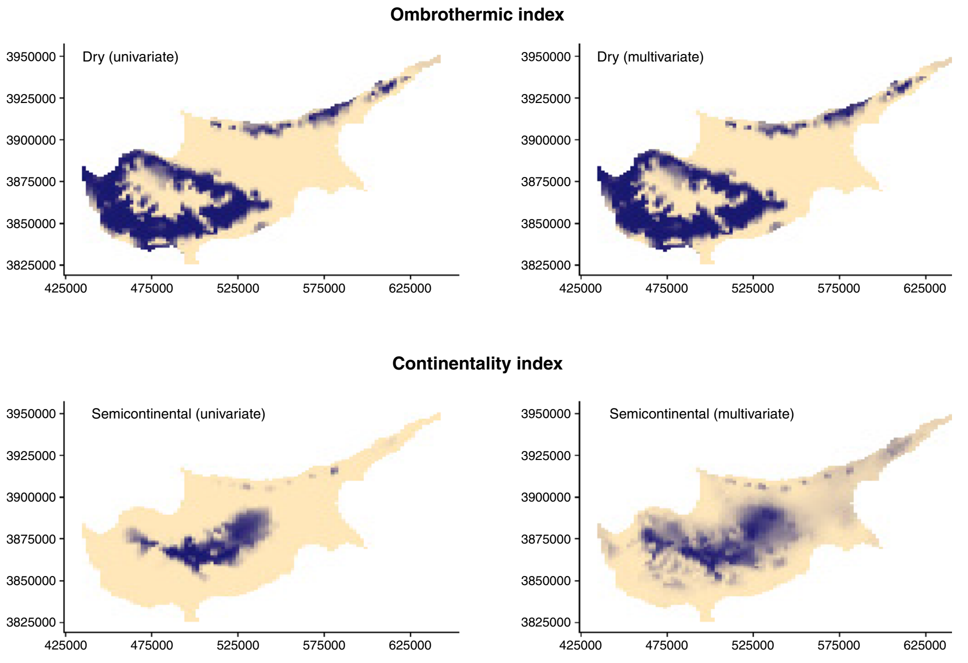

Figure 4. Spatial bioclimatic probability distribution of the Ombrotype *dry* (top) and Continentality subtype *Semicontinental* (bottom) based on the indices when using an univariate approach (left) and a multivariate approach (right).

sification. The main idea underneath the proposal is to use a Bayesian conditional coregionalised linear model. The complete inferential process is reasonable in terms of computational efficiency due to the use of distributed programming, allowing the predictive process to be carried out in a large number of locations. Using this modelling proved to be necessary in those situations in which there is a relationship between indices, an issue that previous works did not take into account.

The reparameterisation of the proportion of spatial and non-spatial variability to total variability (Yan et al. 2007) allowed us to avoid the identifiability problem of spatial and non-spatial variability. Moreover, this proposal provides an easier method for eliciting the prior distributions. In this respect, we have shown how to incorporate our prior knowledge about the parameters via their prior distributions taking into account the particular characteristics of bioclimatic indices. Sensitivity analysis showed that there is no dependence on the selected prior.

Even though posterior predictive distributions of each index proved to be satisfactory measures of the information about the bioclimates, their value is that they provide the probability of each location belonging to different bioclimates. Thus, these spatial bioclimatic probability distributions are the main result of this modelling, as they provide a powerful method not only in studies about the effects on the vegetation of a region but also in landscape management, particularly in establishing future policies or resource management.

Moreover, the dataset that was analysed has shown that this multivariate approach not only shows that the relationship between the indices must be taken into account, but also (and this is even more important) is in accordance with the conditions of the region under study. This suggests that the proposed model could be adapted to other regions without difficulty.

Finally, bioclimate changes over time suggest the incorporation of a temporal component in the model. This may be useful in studies on climate change.

ACKNOWLEDGEMENTS

This work was partially supported by research Grants MTM2016-77501-P and TEC2016-81900-REDT from the Spanish Ministry of Economy and Competiveness. We thank Dr. J.J. López-Espín, Miguel Hernández University-Center for Operations Research, for assistance with parallel computing code.

[Received September 2017. Accepted November 2018. Published Online January 2019.]

REFERENCES

Abramowitz, M., Stegun, I.: Handbook of mathematical functions: with formulas, graphs, and mathematical tables. National Bureau of Standards (1964)

Angetter, L., Loetters, D., Roedder, D.: Climate niche shift in invasive species: the case of the brown anole. Biological Journal of the Linnean Society **104**(4), 943–954 (2011)

Banerjee, S., Carlin, B., Gelfand, A.: Hierarchical Modeling and Analysis for Spatial Data, Second Edition. Chapman and Hall/CRC, Boca Raton (2014)

Barber, A., Tun, J., Crespo, M.: A new approach on the bioclimatology and potential vegetation of the Yucatan peninsula (Mexico). Phytocoenologia **31**(1), 1–31 (2001)

Barber, X., Conesa, D., López-Quílez, A., Mayoral, A., Morales, J., Barber, A.: Bayesian hierarchical models for analysing the spatial distribution of bioclimatic indices. SORT-Statistics and Operations Research Transactions **1**(2), 277–296 (2017)

Burrough, P.: GIS and geostatistics: Essential partners for spatial analysis. Environmental and Ecological Statistics **8**(4), 361–377 (2001)

Buttafuoco, G., Castrignano, A., Busoni, E., Dimase, A.: Studying the spatial structure evolution of soil water content using multivariate geostatistics. Journal of Hydrology **311**(1), 202–218 (2005)

Camps, J., Ramos, M.: Grape harvest and yield responses to inter-annual changes in temperature and precipitation in an area of north-east Spain with a Mediterranean climate. International Journal of Biometeorology **56**(5), 853–864 (2012)

Canu, S., Rosati, L., Fiori, M., Motroni, A., Filigheddu, R., Farris, E.: Bioclimate map of sardinia (italy). Journal of Maps **11**(5), 711–718 (2015)

Catorci, A., Foglia, M., Tardella, F., Vitanzi, A., Sparvoliand, D., Gatti, R., Galli, P., Paradisi, L.: Map of changes in landscape naturalness in the Fiastra and Salino catchment basins (central Italy). Journal of Maps **8**(1), 97–106 (2012)

Chambers, R., Dunstan, R.: Estimating distribution functions from survey data. Biometrika **73**(3), 597–604 (1986)

Chiles, J., Delfiner, P.: Geoestatistics: Modeling Spatial Uncertainty. Wiley, New York (1999)

Conley, A.K., Fuller, D.O., Haddad, N., Hassan, A.N., Gad, A.M., Beier, J.C.: Modeling the distribution of the west nile and rift valley fever vector culex pipiens in arid and semi-arid regions of the middle east and north africa. Parasites & Vectors **7**(1), 289 (2014). https://doi.org/10.1186/1756-3305-7-289

Cressie, N., Zammit-Mangion, A.: Multivariate spatial covariance models: a conditional approach. Biometrika **103**(4), 915–935 (2016) del Arco, M., Perez-de Paz, P.L., Acebes, J.R., Gonzalez-Mancebo, J.M., Reyes-Betancort, J.A., Bermejo, J.A., de Armas, S., Gonzalez-Gonzalez, R.: Bioclimatology and climatophilous vegetation of Tenerife (Canary Islands). Annales Botanici Fennici **43**(3), 167–192 (2006)

Finley, A., Banerjee, S., Gelfand, A.: spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models. Journal of Statistical Software **63**(13), 1–28 (2015)

Gamerman, D., Lopes, H.: Markov chain Monte Carlo: stochastic simulation for Bayesian inference. CRC Press, Boca Raton (2006)

Garzón-Machado, V., Otto, R., del Arco Aguilar, M.J.: Bioclimatic and vegetation mapping of a topographically complex oceanic island applying different interpolation techniques. International Journal of Biometeorology **58**(5), 887–899 (2014)

Gelfand, A.: Hierarchical modeling for spatial data problems. Spatial statistics **1**, 30–39 (2012)

Gelfand, A., Diggle, P., Guttorp, P., Fuentes, Handbook of spatial statistics. CRC press (2010)

Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S.: A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics **2**(4), 1360–1383 (2008)

Genton, M.G., Kleiber, W.: Cross-covariance functions for multivariate geostatistics. Statistical Science **30**(2), 147–163 (2015)

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology **228**(1-2), 113–129 (2000)

Grzebyk, M., Wackernagel, H.: Multivariate analysis and spatial/temporal scales: real and complex models. In: XVIIth International Biometrics Conference, pp. 19–33. International Biometrics Society, Ontario (1994)

Handcock, M., Wallis, J.: An approach to statistical spatial-temporal modeling of meteorological fields. Journal of the American Statistical Association **89**, 368–390 (1994)

Lin, Y.: Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. Environmental Geology **42**(1), 1–10 (2002)

Lunn, D., Thomas, A., Best, N., Spiegelhalter, D.: WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing **10**(4), 325–337 (2000)

Matérn, B.: Spatial Variation, Second Edition. Springer-Verlang, Berlin (1986)

Monteiro-Henriques, T., Espírito-Santo, M.D.: Climate change and the outdoor regional living plant collections: an example from mainland Portugal. Biodiversity and Conservation **20**(2), 335–343 (2011). https://doi.org/10.1007/s10531-010-9864-3

Peng, C.: From static biogeographical model to dynamic global vegetation model: a global perspective on modelling vegetation dynamics. Ecological Modelling **135**(1), 33–54 (2000)

Rivas-Martínez, S.: Bioclimatic classification system of the Earth. Folia Botanica Matritensis **12**, 1–23 (1994)

Rivas-Martínez, S., Rivas-Saenz, S.: Worldwide Bioclimatic Classification System. Phytosociological Research Center, Spain (2017)

Robertson, G.: Geostatistics in ecology: interpolating with known variance. Ecology **68**(3), 744–748 (1987)

Rossi, R., Mulla, D., Journel, A., Franz, E.: Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecological Monographs **62**(2), 277–314 (1992)

Schmidt, A.M., Gelfand, A.E.: A Bayesian coregionalization approach for multivariate pollutant data. Journal of Geophysical Research: Atmospheres **108**(D24) (2003)

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**(4), 583–639 (2002)

Stein, M.: Interpolation of spatial statistics data: some theory for kriging. Springer-Verlang, New York (1999)

Strupczewski, W., Kochanek, K., Weglarczyk, S., Singh, V.: On robustness of large quantile estimates to largest elements of the observation series. Hydrological Processes **21**(10), 1328–1344 (2007)

Verfaillie, E., Van Lancker, V., Van Meirvenne, M.: Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Continental Shelf Research **26**(19), 2454–2468 (2006)

Wackernagel, H.: Multivariate Geostatistics: an introduction with applications, 3rd edition. Springer, New York (2003)

Yan, J., Cowles, M., Wang, S., Armstrong, M.: Parallelizing MCMC for Bayesian spatiotemporal geostatistical models. Statistics and Computing **17**(4), 323–335 (2007)