Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections

Joaquín Bedia[a,1,*], Sixto Herrera[b], Jose Manuel Gutiérrez[a]

[a] *Instituto de Física de Cantabria, Universidad de Cantabria-CSIC. 39005 Santander, Spain*
[b] *Predictia Intelligent Data Solutions, S.L. CDTUC Fase A, Planta 2–203. Avda. los Castros s/n 39005 Santander, Spain*

Abstract

Global bioclimatic datasets are being widely used in ecological research to estimate the potential distribution of species using Climate Envelope Models (CEMs). These datasets are easily available and offer high resolution information for all land areas globally. However, they have not been tested rigorously in smaller regions, and their use in regional CEM studies may pose problems derived from their poor representation of local climate features. Moreover, these problems may be enhanced when using CEMs for future climate projections —a topic of current active research,— due to the uncertainty derived from the future altered climate scenarios.

In this paper we use distributional data of European beech (*Fagus sylvatica*) in Northern Iberian Peninsula to analyze the discrepancies of the

*Corresponding author
Email addresses: joaquin.bedia@unican.es (Joaquín Bedia), sixto@predictia.es (Sixto Herrera), gutierjm@unican.es (Jose Manuel Gutiérrez)
[1] Instituto de Física de Cantabria, Universidad de Cantabria-CSIC. Facultad de Ciencias, Avda. de los Castros 44 (Room 1068). 39005 Santander, Spain. Tel: (+34)942202064, Fax: (+34)942200935

Preprint submitted to Global and Planetary Change *February 8, 2013*

CEMs (predictive skill, variable importance and consistency using different predictor subsets) resulting from three alternative public, high-resolution climate datasets: A benchmarking regional climate dataset developed for the are of study (UC), the University of Barcelona Atlas for the Iberian Peninsula (UAB) and the worldwide WorldClim bioclimatic dataset (WC). The same CEM techniques (multiple logistic regression and multivariate adaptive regression splines) were applied to the different datasets, showing that the quality of the baseline climate has a great impact on the resulting models, as manifested by the different contributions of the bioclimatic predictors to the resulting models. Artifactual bioclimatic variables were found in some datasets, representing topographical features and spatial gradients, rather than true climatic patterns, thus significantly contributing to the models, although not for the right reasons. This causes a misleading model interpretation and problems for extrapolation in future climate conditions, as evidenced analyzing the future projections obtained using state-of-the-art regional climate projections from the ENSEMBLES project.
Keywords: Species distribution models, WorldClim, UAB Atlas, regional climate projection, impacts of climate change

---

1. Introduction

Climate Envelope Models (CEMs), also referred to as ecological niche models or species distribution models, are statistical predictive tools applied in ecological research to estimate the distribution of species, biological communities or habitats (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009). The use of these models is widespread throughout the ecological literature in a variety of application fields, such as biodiversity conservation (Wilting et al., 2010), invasive species propagation (Jeschke and Strayer, 2008) and impacts of climate change (Thuiller, 2003; Araújo et al., 2005), among others. Typically, these techniques use medium to high-resolution grids (several minutes to seconds of arc, see e.g. Kriticos et al., 2012) over the area of interest and combine observations of species occurrence with appropriate bioclimatic indicators defined at the grid box scale. The result is a predictive model assigning a probability of occurrence to each of the grid boxes as a function of the bioclimatic indicators.

The recent development of new global high-resolution bioclimatic datasets has broaden the scope of CEMs across different regions and continents and has also boosted their application in climate change impact studies (Peterson et al., 2002; Hijmans and Graham, 2006). The need for high-resolution input data in this context has been already highlighted by some authors, given the unability of coarse-resolution models to represent local refugia (e.g. Randin et al., 2009; Franklin et al., 2013). One of the most popular global bioclimatic products is the WorldClim dataset (Hijmans et al., 2005), which is widely being used because it is easily available and offers high resolution (∼1km) for all land areas globally. Other newer global interpolated products of similar characteristics have appeared recently in the literature (e.g. the new Climond dataset, Kriticos et al., 2012, which is partly based on WorldClim data), indicating the high demand of these kind of products in the last years.

However, these global datasets have not been tested rigorously in smaller regions, and their use in regional studies may pose problems derived from their poor representation of local climate features over certain areas. To date, most of the studies fail to explicitly analyze the sensitivity of CEMs to the baseline climate data (Peterson and Nakazawa, 2008; Soria-Auza et al., 2010), partly because of the lack of high-quality climate datasets —in many areas of the world— that may be confidently used as a reference. Moreover, in those studies applying CEMs for future climate projections, the defects of the baseline climatology may be enhanced by the uncertainty derived from the future climate scenarios (Beaumont et al., 2008; Wiens et al., 2009), thus seriously compromising the practical validity of the resulting projections for planners and adaption–strategists (see, e.g., Araújo and New, 2006).

In this study we present a sensitivity analysis of CEMs to different baseline climate datasets, using distributional data of a tree species —the European Beech (*Fagus sylvatica* L.), *Fagus* henceforth— in Northern Iberian Peninsula. In particular, we consider three different climate datasets encompassing a range of spatial extents, from global to regional: The WorldClim global database, the Atlas by the Universitat Autónoma de Barcelona for the Iberian Peninsula (Ninyerola et al., 2005), and the benchmark high–quality regional grid developed by the authors at the University of Cantabria for the region of study (Gutiérrez et al., 2010); hereafter we will refer to them as WC, UAB and UC respectively. The same CEM techniques (multiple logistic regression and multivariate adaptive regression splines) were applied to the different datasets, evaluating the resulting models in the light of their AUC and Cohen's κ, as the usual performance metrics in CEM studies.

A first comparison of the three datasets reveals deficiencies of WorldClim, which fails to properly represent all precipitation bioclimate indices over the region. We show how these artifactual indices are actually representing topographical features and spatial gradients as a result of the underlying interpolation process, rather than true climatic patterns, thus significantly contributing to the CEM models, although not for the right reasons. This causes a misleading interpretation of the resulting models and problems for extrapolation in future climate conditions. The problems found are beyond the deficiencies reported for WorldClim precipitation in mountainous areas (Hijmans et al., 2005; Tadić, 2010).

In order to estimate the sensitivity of the resulting CEMs in future climate scenarios, we considered the regional projections given by the ensemble of Regional Climate Models (RCMs) from the EU-funded ENSEMBLES project for the A1B scenario (Jacob et al., 2007). First, the *delta* method was applied to obtain the future climate projections —and the corresponding derived bioclimatic indices,— adding the differences to the three different baseline climatologies (see, e.g., Räisänen, 2007; Zahn and von Storch, 2010, for a description and application of delta method). Then, future projections of species distributions were obtained by applying the different CEM models to the corresponding future bioclimatic indices. The resulting projections differed markedly —particularly for WorldClim,— highlighting the inadequacy of high resolution worldwide climate datasets for their application in regional climate change studies. However, when only temperature-related bioclimatic variables —more robust across the different datasets— were included, the projections were in relatively good agreement for all the datasets.

This paper is organized as follows. In Sec. 2 the area of study and datasets used in the paper are presented. Sec. 3 describe the different methodologies applied, including CEM modeling algorithms, model assessment. The main results and discussion are presented in Sec. 4. Finally, some conclusions are given in Sec. 5.

2. Area of Study and Datasets

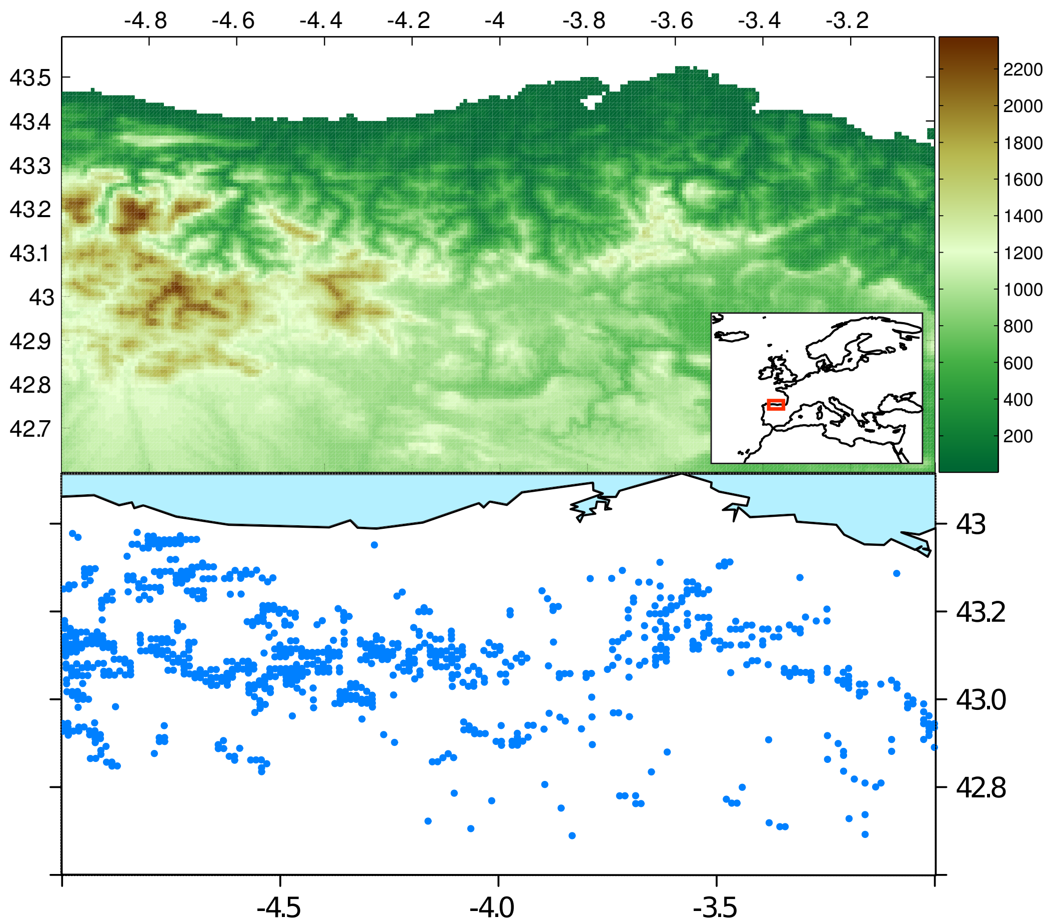

Figure 1: Location of the study area. In the top panel the orography of the target area is represented at a 1km resolution (meters above sea level). In the lower panel, the distribution of *Fagus* is shown at a 1km pixel resolution.

The area of analysis in this study is centered in Northern Iberian Peninsula, with a bounding box of coordinates $42.60°N$, $-5.00°E$ to $43.60°N$, $-2.99°E$ (Fig. 1). In the Iberian Peninsula, *Fagus* forests are mainly found in the Northern mountain areas although they reach the Iberian and Central Ranges at some particular locations (Costa et al., 1998). *Fagus* has a restricted niche in the study area, linked to mountainous areas mostly between 400 and 1400m above sea level. 95% of the presence localities used in this study lie within this elevation range, showing a very clear unimodal distribution. The sharp interaction between *Fagus* and the climatic gradient in the study area has motivated the choice of this species, which is expected to be modeled with higher accuracy than other generalist species (Brotons et al., 2004; Araújo and Guisan, 2006; Tsoar et al., 2007). This region is determined by the extent of the UC climate grid, which has the smallest geographical extent of the datasets used in this study.

2.1. Species distribution data

The information on *Fagus* distribution was obtained from the Forest Map produced by the Third National Forest Inventory (MARM, 2006). The original vector map was filtered so that all polygons containing the target species were retained and then rasterized to a pixel size of 0.0083° (aprox. 1km), leading to a total of *ca.* 900 localities of presence within the study area (Fig. 1, bottom panel).

Most probabilistic modeling methods require absence points —in addition to occurrences— for training (see e.g. Bedia et al., 2011). Since we lacked of real absences, we generated synthetic background points assigning them a value of zero (absence). Occurrence data define the conditions under which the species is more likely to be present, whereas background data establishes the environmental domain of the study (Hijmans et al., 2012). Thus, background points do not represent actual absences, and they are randomly generated in an equal number to the presences, following some authors who suggest that intermediate prevalences produce better results (McPherson et al., 2004; Allouche et al., 2006; Mateo et al., 2010). In addition, we set a buffer radius of 2000m around known presences, in order to minimize false negatives due to cartographic inaccuracies inherent to the delineation of vectorial maps (Graham et al., 2008).

2.2. Baseline climate datasets

For the sake of conciseness, we only introduce their main characteristics of the climate datasets used in this study, with some emphasis in the description of the more recent benchmarking UC dataset. The interested reader is referred to the published documentation of these datasets for further details on their construction.

WorldClim (WC, Hijmans et al., 2005) is a global temperature and precipitation dataset with a spatial resolution of 30 arc–seconds (aprox. 1km), obtained applying a thin–plate spline smoothing interpolation algorithm to a large number of weather stations throughout the world, covering most of Earth's for approximately 50 years (1950–2000). This dataset is freely available for download from internet (`http://www.worldclim.org`).

The climate surfaces of the University of Barcelona Atlas (UAB, Ninyerola et al., 2005) were calculated by multiple regression and residual analysis, introducing as covariates a relatively simple set of variables: altitude, slope, different indices used to describe distance to the sea, solar radiation and terrain curvature. Temperature and precipitation data for the period 1950–

2000 were obtained from the national network of the Spanish Meteorology Agency (AEMET), and from the literature in the case of Portugal. The UAB dataset is provided at a very high resolution (200m) for the entire Iberian Peninsula, and is available for download from the internet (`http://opengis.uab.es/wms/iberia/mms/index.htm`).

The high resolution climate grid developed for Cantabria and surrounding territories by the University of Cantabria (UC, Gutiérrez et al., 2010), is based on the same AEMET stations network than UAB. Data from 148 (62) stations were used for precipitation (temperature), respectively, after a process of data quality control, within the period 1950-2003. All data series were required to have a minimum of 10 years with less than 10% of missing values, and they were tested for relative homogeneity (Alexandersson, 1986; Alexandersson and Moberg, 1997) and absolute homogeneity (SNHT method, Khaliq and Ouarda, 2007), after discarding outliers. The performance of different techniques was tested, namely thin–plate splines, angular distance weighting and kriging (Krige, 1951), obtaining best results with the latter one. This method has been widely used in climate research (Atkinson and Lloyd, 1998; Biau et al., 1999; Haylock et al., 2008) and provides high flexibility for covariate introduction and uncertainty analysis. In the case of the precipitation, a two–step interpolation process was conducted: first, precipitation occurrence was interpolated using *indicator kriging* (Juang and Lee, 1998); then, the amount of precipitation was interpolated using ordinary kriging, assigning values of 0 to all 'dry' points. Thus, the frequency distribution of precipitation for both occurrence and amount was optimally fit. In the calculation of uncertainty, the dependency among observations was incorporated following Yamamoto (2000). The final 1km–resolution grid was obtained by regression–kriging (Hengl et al., 2007), introducing a set of basic covariates describing terrain characteristics including, elevation, distance to coastline, and topographic blocking effects. The interpolated temperature and precipitation were subject to expert revision by meteorologists of AEMET based on their deep knowledge on the climate of this region (Cano, 1999), leading to final refinement by elimination of some coastal weather stations with systematic errors, not detected in the previous stage of automated data quality control.

Thus, UC and UAB are constructed upon the same network of stations and using a similar methodology, based on multiple linear regression with a residual adjustment by means of an interpolation process. The main differences between them lie the level of detail at which the resulting surfaces have been checked for quality, due to their different geographical coverages, and in the type of covariates introduced into the models. In this sense, UAB uses an input orography of 200m resolution, and introduces terrain curvature among other covariates, thus leading to a fine-grain level of detail that is then propagated into the climate surface by the regression model. On the other hand, WC is based on thin–plate splines, considering a simple set of covariates (longitude, latitude and elevation), which are applied to a much more sparse network of observations, provided its worldwide coverage.

In this work we consider the set of 19 bioclimatic indices provided by WorldClim, which are commonly used in ecological modeling (see Table 1). To allow for full spatial comparability among the three datasets (UC, UAB and WC), the original layers were re-projected to geographical coordinates and resampled to match the same 1km regular grid. For UC and UAB, the bioclimatic indices were derived from the precipitation and temperature layers provided in those cases. The common baseline period 1950–2000 was selected for the three datasets based on their temporal overlapping. The resulting bioclimatic indices are compared in Table 1 and partially displayed in Fig. 3.

2.3. Future climate projections

In order to calculate future projections of species distributions using CEMs, we considered the state-of-the-art regional projections given by seven Regional Climate Models (RCMs, Table 2) from the EU-funded ENSEMBLES project (van der Linden and Mitchell, 2009). These RCMs were run over a limited domain covering Europe with a horizontal resolution of 25km, driven at the boundaries by different GCM simulations under the A1B emission scenario (Nakićenović, 2000). However, it has been recently recognized that the outputs of the RCMs cannot be used directly for impact studies, since they may contain important biases resulting from different physics and parameterizations involved in their formulation (Winkler et al., 1997). To alleviate this problem, we applied the so-called 'delta' method (see, e.g., Räisänen, 2007; Zahn and von Storch, 2010) or 'change factors' (Winkler et al., 1997) and, thus, the baseline climatological values are modified at a grid-box level by a change factor, obtained as the difference/ratio of the temperature/precipitation values between a future period (e.g. 2071-2100) and the control period (1970-1999 in this study). We computed the altered future bioclimatic indices for the periods 2011–2040, 2041–2070 and 2071–2100, based on the climate change signals for precipitation, minimum and maximum temperature values. The mean ensemble increments (and deviations) obtained for the region of study for the periods 2011–2040, 2041–2070 and 2071–2100 were, respectively, -32.2 (47.6), -93.8 (32.2) and -173.3 (82.5) mm/year for precipitation; 0.80 (0.18), 1.76 (0.19) and 2.54 (0.12) $^{\circ}C$ for minimum temperature, and 0.92 (0.17), 1.98 (0.14) and 2.94 (0.22) $^{\circ}C$ for maximum temperature.

3. Methods

3.1. CEM modeling algorithms

CEMs were originally constructed using a number of probabilistic algorithms, namely generalized linear models, support vector machines, artificial neural networks, maximum entropy, and multivariate adaptive regression splines (see Bedia et al., 2011, for a comparative analysis of this techniques in the framework of species distribution modeling). All methods yielded similar results, with slight to moderate differences in the resulting probabilistic distributions, leading to the same overall conclusions. We selected generalized linear models (GLMs) as the preferred technique to use, given that the focus of this study is to analyze the uncertainties derived from the baseline climatology, rather than the inherent differences stemming from the use of different modeling algorithms. Although non-linear techniques may lead to models of improved predictive accuracy (Elith et al., 2006; Bedia et al., 2011), on the other hand they may eventually obscure the actual contribution of each variable proven their higher complexity. With this regard, GLMs provide a flexible and robust framework for assessing the statistical significance of the explanatory variables and the estimation of their importance (see Section 3.4), providing a simple and sound model interpretability at a low computational cost (see Guisan et al., 2002, for an overview of GLMs in the context of species distribution modeling). Nevertheless, throughout the manuscript we will also present some results corresponding to the Multivariate Adaptive Regression Splines models (MARS, Friedman, 1991) as an example of non-linear technique, that illustrates the consistency of the results regardless of the modeling technique applied. MARS is a non–parametric method for regression which approximates the underlying function through a set of adaptive piecewise linear regressions, known as *basis functions*. More details on this method are presented in (Bedia et al., 2011).

3.2. Correlation analysis

The high inter-dependence between some of the bioclimatic variables used as predictors (Table 1) gives raise to the issues of redundancy and multicollinearity, negatively affecting variable selection and model interpretability due to the drastic changes in model parameter values, and also hampering the ability of the model for extrapolation (Brauner and Shacham, 1998), central in climate change studies. In order to avoid redundancy, we eliminated from the analysis the bioclimatic variables yielding correlation values above 0.95 (Spearman's rho coefficient) in the pairwise cross-correlation matrix of each dataset (*intra-dataset* correlations). The threshold of 0.95 is conservative, and it was chosen in order to keep other variables that, although also highly correlated, may still provide some useful additional information. Moreover, in the next step, collinear variables have been set aside of subsequent analyses (Sec 3.3). In addition, we also computed pairwise correlations between datasets (*inter-dataset* correlations) as a first exploratory analysis of the consistency of the different climatologies.

3.3. Multicollinearity analysis and variable selection

After the elimination of highly correlated variables, the resulting non-redundant datasets were further checked for multicollinearity. Among the different approaches available for detecting multicollinearity (see Brauner and Shacham, 1998, for an overview), we have followed the classical method based on the condition number of the normal matrix, which has been extensively used for collinearity diagnosis (Brauner and Shacham, 1998). In the absence of multicollinearity, the eigenvalues, condition indices and condition number of the predictors matrix will all equal one. As collinearity increases, eigenvalues will be both greater and smaller than one (eigenvalues close to zero indicate a multicollinearity problem), and the condition indices and the condition number will increase, leading to an unstable model definition.

The simplest approach to circumvent multicollinearity consists of dropping all collinear variables. However, in order to avoid inferential problems derived from arbitrarily dropping/retaining predictors (Graham, 2003), we have followed a sequential data-driven modeling approach: first, the variable attaining the highest predictive performance (in terms of AUC) is retained. Then, the remaining variables are tested for collinearity, setting a maximum allowable condition number below 30. Those variables producing condition numbers above the threshold of 30 are dropped, and the selection procedure is iteratively repeated until no more candidate variables remain. The main disadvantage of this approach is that no critical value for the condition number has been established to indicate harmful collinearity (Brauner and Shacham, 1998). The value chosen has been suggested by (Cohen et al., 2003), and represents a "rule of thumb" criterion, that we deemed appropriate in this case after checking the the low cross-correlation values of the resulting datasets (Fig.2) and their spatial distribution (Fig.3). We followed this variable selection procedure for each dataset (UC, AUB and WC) leading to three different sets of bioclimatic predictors, subsequently used in the following analyses.

3.4. Variable importance assessment

In order to estimate variable importance in the context of logistic regression, we have applied the method of hierarchical partitioning, by which the independent effect of each variable is calculated by comparing the fit of all models containing a particular variable to the fit of all nested models lacking that variable (Chevan and Sutherland, 1991). For instance, for variable X_1, its importance I would be calculated as follows:

$$I_{x1} = \sum_{i=0}^{k-1} \frac{\sum(r^2_{y,X_1X_h} - r^2_{y,X_h})/\binom{k-1}{i}}{k} \tag{1}$$

where X_h is any subset of i predictors from which X_1 is excluded. As a result, the variance shared by two or more correlated predictors can be partitioned into the variance attributable to each predictor. This method provides a robust assessment of variable importance and has been shown to outperform other methods used for variable importance estimation in the context of regression analysis, after the removal of spurious variables (Murray and Conner, 2009).

3.5. Model assessment

We performed a k-fold cross-validation of the models, with k=10 stratified randomly splitted subsets of presence/absence, each of them containing an approximately equal number of presences and absences (50%), following the criteria presented in Section 2.1. Model skill was assessed by computing the ROC curves for each model and calculating the corresponding AUCs. We also computed Cohen's κ using prevalence as the probability cutoff threshold ($P = 0.5$).

All the analyses were conducted in the `R` language and environment for statistical computing (R Development Core Team, 2012).

4. Results and Discussion

4.1. Correlation analysis of bioclimatic variables

The intra-dataset pairwise correlation analysis identified some redundant variables, common to the three datasets (Fig. 2a-b). As a result, BIO1, 6 and 11, based on temperature data, were in all cases highly cross-correlated ($\rho > 0.95$), and only BIO11 was retained. Regarding precipitation, variables BIO12 and 13 (redundant with BIO16) and BIO17 (redundant with BIO14) were dropped for the same reason. There is a high number of temperature-related bioclimatic variables highly correlated with precipitation ones in the WC dataset, whereas these correlations are lower and less frequent in the case of UAB and UC. As an example, unlike UC and UAB, BIO5 of WC shows a very high correlation with BIO14 and BIO17 (Fig. 2b). This constitutes a first note of warning on the problems with the precipitation variables in WC.

The inter-dataset pairwise correlations revealed remarkable differences between the bioclimatic variables among datasets. The lack of consistency between datasets is more accentuated for WC than for UAB, with regard to the UC data (Fig. 2c-d). There is a general good agreement between precipitation variables of UAB and UC, but there is scarce correspondence in the case of WC, highlighting again the problems derived from precipitation data in WC. These differences become apparent in the spatial distribution of the bioclimatic variables displayed in Fig. 3. For instance, BIO14 (precipitation of the driest month) has a comparable spatial distribution for UC and UAB. Although UAB exhibits a fine-grain level of detail that seems not realistic in this case, it does not significantly alter the overall spatial pattern, preserving a high level of agreement with UC (ρ=0.91, $rmse$=4.8). On the contrary, BIO14 of WC has a markedly different spatial distribution and magnitude (ρ=0.64, $rmse$=13.7). Similar results are obtained for BIO15, which in the case of WC is strongly correlated with the topography, and unlike UC and UAB, with negative sign (Fig. 4). With regard to the temperature-related bioclimatic variables, BIO9 (mean temperature of the driest quarter) is the most similar among datasets, evidencing a close relationship with the orography in all cases (Fig. 4). On the other hand, BIO3 (isothermality) and BIO5 (maximum temperature of the warmest month), are not correlated at all with orography in UC, but they are in UAB and WC. Moreover, in the case of BIO3, the signs of the correlation of UAB and WC are opposite.

Therefore, the correlation analysis revealed important inconsistencies between datasets. The largest deviations are exhibited by WC, with some bioclimatic patterns that seem more related with orography than with the actual climatic features of the study area, as represented by UC. This is specially true in the case of precipitation, as none of the bioclimatic variables is able to approximate the UC and UAB precipitation pattern, which in general terms are more similar than WC. However, regardless of their dependence on temperature or precipitation, the most differing bioclimatologies correspond to those related with climatic variability (BIO2 and 3 for temperature, and BIO15 for precipitation). In this case, UAB also failed to approximate the UC climatologies.

4.2. Variable selection and importance in the models

The large differences among the bioclimatic datasets, with intra-dataset dependencies and correlations of varying nature and magnitude, prevents from the use of a common subset of variables for the development of the CEM models, from which an overall assessment of variable importance can be made. Thus, we applied the variable selection procedure independently to each dataset, which yielded the predictor combinations (or *subsets* hereafter) presented in Table 3.

In all cases, the first variables chosen (based on their maximization of model AUC), were related with temperature. These were BIO9 (mean temperature of the driest quarter) in the case of UC and WC, and BIO5 (maximum temperature of the warmest month) in the case of UAB, both related with the temperature regime during summer in the study area. In the case of BIO9, due to its strong control by orography (Fig. 4), the differences of UAB and WC with UC are minor. In the case of BIO5, WC shows a 2°C mean bias, although the spatial pattern is well preserved in general terms. Variable BIO14 (Precipitation of the driest month) was included in the three subsets of predictors.

For the sake of conciseness, in the analysis of variable importance we will display only the results of the UC subset, provided that the overall results and conclusions are similar when the UAB and WC subsets are used instead. The variable importance given to temperature–related variables is quite high in the case of the UC model, and also in the case of UAB, whereas WC models tend to give larger importance to precipitation-related variables, notably BIO14 (Fig. 5).

The variable importance in the models evidences that temperature is an important variable for modeling *Fagus* distribution, which implies a strong orographic component, as highlighted in Fig. 4. Nevertheless, there is an important fraction of the variability explained by precipitation in the UC model (BIO16), a variable that is weakly correlated with the elevation in the study area, and therefore the added value of precipitation for *Fagus* modeling should not be disregarded. As a result, some variables very correlated with topography are very important for *Fagus* CEMs. In the case of precipitation variables of WC, this relationship with the orography is not justified by a real climatic phenomenology, but rather by a side effect of the interpolation algorithm. The same applies to some temperature-related bioclimatic variables, like BIO2 and BIO5, that both UAB and WC include with preference in their models, and which exhibit large differences with the UC benchmark.

4.3. Predictive skill of the models

For the assessment of CEM predictive skill, we computed the AUC and Cohen's κ of the 10-fold cross validation models, considering for each dataset its own subset of predictors (Table 3), thus maximizing the predictive skill in each case. All models achieved fairly high AUC and Cohen's κ values, typically attributed to predictive systems with a good discrimination ability (Swets, 1988; Landis and Koch, 1977). The results corresponding to Cohen's κ are comparable to those obtained by AUC, and thus, for the sake of brevity, we will refer only to AUC hereafter. In addition, the results achieved by the more sophisticated MARS algorithm are also displayed, evidencing its better performance in terms of AUC (Fig. 6), although in relative terms, the results are similar to GLMs.

As previously shown, some precipitation variables have a large weight in the WC model, even though they do not correspond to the actual precipitation pattern in the study area. However, this had no apparent effect on the CEM skill, which was similar in the three datasets, with a slightly better performance of GLM in the case of UAB (median > 0.90 considering the k=10 models of the k-fold cross validation. Fig. 6, lower panel). Given that the largest differences between datasets are in precipitation, we also computed CEMs using temperature variables only (indicated in Table 3 without the asterisk). In this case, the results were more similar across datasets, with a very slight loss of skill, more apparent for MARS than for GLM models, probably due to the non-linearities between both types of variables that MARS is able to capture. The inclusion of precipitation improved the predictive skill of the UAB and UC models, confirming the added value of precipitation for *Fagus* modeling, previously indicated in the independent effects analysis (Sec. 4.2). On the contrary, the removal of precipitation variables in the WC model did not produce any changes the AUC, evidencing that precipitation variables of WC provide little or no improvement at all in CEM skill, once temperature-related ones are used.

4.4. CEM predictions and uncertainty

As it can be expected from the similar predictive skills attained by the UC, UAB and WC *Fagus* CEMs, the probabilistic maps yielded similar results in terms of spatial distribution of *Fagus* potentiality (Fig. 7a), although some fine-grain details, previously shown in the bioclimatic predictors, are now apparent in the predicted distributions of UAB. The sharp transitions between presence and absence in UAB and WC (probability threshold of 0.5), contrasts with the smooth probabilistic spatial prediction of the UC model. In order to test the robustness of these models to changes in the predictor combinations, we alternatively constructed CEMs using the three different variable subsets (Table 3) for each climate dataset, and computed the standard deviation of the resulting distribution maps. We found that UC yielded very similar distributions in all cases, whereas the spread of the predictions was larger in the case of UAB and WC (Fig. 7b), showing the robustness of the UC models to changes in the predictor combinations.

4.5. Future distribution forecasting

Future *Fagus* distributions were computed using the models obtained in the previous section, but driven by the regional scenarios described in Sec. 2.3, calculated according to the delta method. The future distributions corresponding to each RCM projections were computed individually, and the mean and standard deviation of the resulting ensemble was computed in a grid box basis (Fig. 8). Note that in the future maps presented, especially in the case of WC, the native grid of the RCMs is noticeable. This is the "true" resolution of the climate change signal provided by the ENSEMBLES RCMs (∼25km), added to the baseline climatology applying the delta method. Thus, the resulting squared tessellation is not an artifact, but the real resolution at which projections can be realistically provided in this case. We prefer to keep it instead of smoothing the maps by means of an interpolation process, as this would represent an added source of uncertainty to the projections. In addition, by preserving the original resolution of the climate change signal, the spatial consistency of UC and UAB models when deltas are applied is highlighted, as opposite to WC, which also constitutes an indication of the lack of robustness of WorlClim in the representation the climate in the region of analysis.

In general, future distributions using UC and UAB datasets are similar, and represent the expected trend of *Fagus* retreat in its southern European limit of distribution, in accordance with previous studies on this species (Kramer et al., 2010; Felicísimo et al., 2010). In contrast, future range projections produced by WC do not follow a logical pattern, in the sense that a very sudden decline in potentiality is projected for the first period (2011-2040), that is reverted during the second period (2041-2050). In addition, the uncertainty (i.e., the standard deviation of the ensemble) associated to WC projections is very large, a clear symptom of an unreliable future projection.

Note, however, that when only temperature-related bioclimatic variables —more robust across the different datasets— are considered in the modeling process, the projections obtained with the resulting CEMs are in relatively good agreement for all the datasets and similar to the full-variable results obtained in the case of UC. This gives some extra evidence of the instability caused in the future projections by the deficiencies of the baseline climate in the CEM modeling process.

5. Conclusions

We found that the precipitation of WorldClim does not correspond to the actual climatic conditions in the study area, neither in the spatial pattern represented, nor in its magnitude. On the contrary, the UAB dataset was able to preserve both characteristics, although other problems derived from the inclusion of fine-grain covariates in the regression models were noticeable in some bioclimatic maps and in the resulting CEMs. Even though temperatures had a similar spatial distribution in all datasets –with an important negative bias in the case of maximum temperatures in WC–, some of the derived bioclimatic variables, such as the mean diurnal temperature range and the isothermality, showed large differences. With this regard, our results evidence the reliance of these bioclimatic variables on the orography, attributable to the interpolation methods used to build the climatologies.

As a result, in spite of the large differences among datasets and the high importance attained by precipitation-related variables in the WC model, their respective CEMs were able to skillfully predict current *Fagus* distribution in all cases, attaining similar model performances after the cross–validation tests, and consistent results independently of the modeling algorithm used. Nevertheless, in the case of UAB and WC, this comes at the cost of a misleading model interpretation and a lack of robustness of the resulting CEMs with the introduction of new predictor combinations. With regard to future projections, as far as the climate change signals in the delta method are not added to true climatic features, but on statistical artifacts highly related to the topography, the resulting future maps obtained using WC become unreliable due to the large spread of the forecasts, yielding non-robust projections.

Modelers should be aware of the limitations imposed by the poor representation of regional climate that global datasets perform at some areas of the world. Due to the lack of adequate high-resolution data for validation in many areas of the world, the problems derived from the use of WorldClim for CEM development at a regional/local scale might not be readily apparent, given that model skill, as determined by the commonly applied performance metrics, is not necessarily as bad as to discard the models. However, we warn about the potentially misleading interpretability of the resulting models and their inadequacy for climate change studies, which seriously impair their practical applicability in biodiversity management and conservation planning.

Finally, we want to emphasize that the aim of this study is to warn about the critical importance of accurate input climate data for CEM analysis and interpretability, and subsequent extrapolation to future climate conditions, and not the estimation of the current/future bioclimatic potentiality of *Fagus*, that would require accounting for other sources of uncertainty beyond the scope of this paper (see, e.g. Fronzek et al., 2011).

6. Aknowledgements

This research has received funding from the European Union's Seventh Framework Programme under grant agreements 243888 (FUME Project) and from the CICYT project EXTREMBLES (CGL2010-21869). We thank two anonymous reviewers, who provided insightful comments that greatly improved the original manuscript.

References

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. J. Climatol. 6, 661–675.

Alexandersson, H., Moberg, A., 1997. Homogenization of swedish temperature data. Part I: Homogeneity test for linear trends. Int. J. Climatol. 17, 35–54.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688.

Araújo, M.B., New, M., 2006. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Glob. Change Biol. 11, 1504–1513.

Atkinson, M., Lloyd, C.D., 1998. Mapping precipitation in Switzerland with ordinary and indicator kriging. Journal of Geographic Information and Decision Analysis 2, 65–76.

Beaumont, L.J., Hughes, L., Pitman, A.J., 2008. Why is the choice of future climate scenarios for species distribution modelling important? Ecology Letters 11, 1135–1146.

Bedia, J., Busqué, J., Gutiérrez, J.M., 2011. Predicting plant species distribution across an alpine rangeland in northern Spain: a comparison of probabilistic methods. Applied Vegetation Science 14, 415–432.

Biau, G., Zorita, E., von Storch, H., Wackernagel, H., 1999. Estimation of precipitation by kriging in the EOF space of the sea level pressure field. Journal of Climate 12, 1070–1085.

Brauner, N., Shacham, M., 1998. Role of range and precision of the independent variable in regression of data. Aiche Journal 44, 603–611.

Brotons, L., Thuiller, W., Miguel, B., 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27, 437–448.

Cano, R., 1999. Atlas climático de la Región Cantábrica. Nota Técnica CMT/CAS. Instituto Nacional de Meteorología. In Spanish.

Chevan, A., Sutherland, M., 1991. Hierarchical partitioning. The American Statistician 45, 90–96.

Christensen, O., Drews, M., Christensen, J., Dethloff, K., Ketelsen, K., Hebestadt, I., Rinke, A., 2006. The HIRHAM regional climate model version 5. Technical Report. Danish Meteorological Institute. Copenhagen, Denmark.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, New Jersey, USA. 3rd edition.

Collins, M., Booth, B., Harris, G., Murphy, J., Sexton, D., M., W., 2006. Towards quantifying uncertainty in transient climate change. Climate Dynamics 27, 127–147.

Costa, M., Morla, C., Sainz, H., 1998. Los bosques ibéricos. Una interpretación geobotánica. Planeta, Barcelona, Spain. In spanish.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Leathwick, J.R., 2009. Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics 40, 677–697.

Felicísimo, A.M., Muñoz, J., Villalba, C.J., Mateo, R.G., 2010. Impactos, vulnerabilidad y adaptación al cambio climático de la flora española. Technical Report. Universidad de Extremadura, Real Jardín Botánico (CSIC), Oficina Española de Cambio Climático. In Spanish.

Franklin, J., Davis, F., Ikegami, M., Syphard, A., Flint, L., Flint, A., Hannah, L., 2013. Modeling plant species distributions under future climates: how fine scale do climate projections need to be? Global Change Biology 19, 473–483.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Annals of Statistics 19, 1–67.

Fronzek, S., Carter, T., Luoto, M., 2011. Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-arctic palsa mires. Natural Hazards and Earth System Sciences 11, 2981–2995.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., The NCEAS Predicting Species Distributions Working Group, 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45, 239–247.

Graham, M., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84, 2809–2815.

Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157, 89–100.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Gutiérrez, J.M., Herrera, S., San Martín, D., Sordo, C., Rodríguez, J.J., Frochoso, M., Ancell, R., Fernández, J., Cofiño, A.S., Pons, M.R., Rodríguez, M.A., 2010. Escenarios Regionales Probabilísticos de cambio climático en Cantabria: Termopluviometría. Gobierno de Cantabria-Consejería de Medio Ambiente y Universidad de Cantabria, Santander, Spain. In Spanish.

Haylock, M.R., Hofstra, N., Klein-Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. Journal of Geophysical Research 113.

Hengl, T., Heuvelink, G., Rossiter, D., 2007. About regression-kriging: From equations to case studies. Computers and Geosciences 33, 1301–1315.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology 12, 2272–2281.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2012. dismo: Species distribution modeling. R package version 0.7-23.

Jacob, D., Barring, L., Christensen, O.B., Christensen, J.H., de Castro, M., Deque, M., Giorgi, F., Hagemann, S., Lenderink, G., Rockel, B., Sanchez, E., Schaer, C., Seneviratne, S.I., Somot, S., van Ulden, A., van den Hurk, B., 2007. An inter-comparison of regional climate models for Europe: model performance in present-day climate. Climatic Change 81, 31–52.

Jacob, D., Van den Hurk, B., Andrae, U., Elgered, G., Fortelius, C., Graham, L., Jackson, S., Karstens, U., Kopken, C., Lindau, R., Podzun, R., Rockel, B., Rubel, F., Sass, B., Smith, R., Yang, X., 2001. A comprehensive model inter-comparison study investigating the water budget during the BALTEX–PIDCAP period. Meteorology and Atmospheric Physics 77, 19–43.

Jeschke, J.M., Strayer, D.L., 2008. Usefulness of bioclimatic models for studying climate change and invasive species, in: YEAR IN ECOLOGY AND CONSERVATION BIOLOGY 2008. BLACKWELL PUBLISHING, 9600 GARSINGTON RD, OXFORD OX4 2DQ, OXEN, ENGLAND. volume 1134 of *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, pp. 1–24.

Juang, K., Lee, D., 1998. Simple indicator kriging for estimating the probability of incorrectly delineating hazardous areas in a contamined site. Environmental Science & Technology 32, 2487–2493.

Khaliq, M.N., Ouarda, T.B.M.J., 2007. On the critical values of the Standard Normal Homogeneity Test (SNHT). International Journal of Climatology 27, 681–687.

Kjellström, E., Bärring, L., Gollvik, S., Hansson, U., Jones, C., Samuelsson, P., Rummukainen, M., Ullerstig, A., Willén, U., Wyser, K., 2005. A 140–year simulation of European climate with the new version of the Rossby Centre regional atmospheric climate model (RCA3). Rep. Meteorol. Climatol. 108, 681–687.

Kramer, K., Degen, B., Buchsbom, J., Hickler, T., Thuiller, W., Sykes, M.T., de Winter, W., 2010. Modelling exploration of the future of European beech (*Fagus sylvatica* L.) under climate change - Range, abundance, genetic diversity and adaptive response. Forest Ecology and Management 259, 2213–2222.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52, 119–139.

Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J., Scott, J.K., 2012. Climond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. Methods in Ecology and Evolution 3, 53–64.

Landis, J., Koch, G., 1977. Measurement of observer agreement for categorical data. Biometrics 33, 159–174.

MARM, 2006. Tercer inventario forestal nacional.

Mateo, R.G., Croat, T.B., Felicísimo, A.M., Muñoz, J., 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. Diversity and Distributions 16, 84–94.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Journal of Applied Ecology 41, 811–823.

Murray, K., Conner, M., 2009. Methods to quantify variable importance: implications for the analysis of noisy ecological data. Ecology 90, 348–355.

Nakićenović, N., 2000. Greenhouse Gas Emissions Scenarios. Technological Forecasting and Social Change 65, 149–166.

Ninyerola, M., Pons, X., Roure, J.M., 2005. Atlas climático digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica. Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. In Spanish.

Pal, J.S., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Gao, X., Rauscher, S.A., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F.S., Bell, J.L., Diffenbaugh, N.S., Karmacharya, J., Konare, A., Martínez, D., da Rocha, R.P., Sloan, L.C., Steiner, A.L., 2007. Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. Bulletin of the American Meteorological Society 88, 1395–1409.

Peterson, A., Ortega-Huerta, M., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R., Stockwell, D., 2002. Future projections for mexican faunas under global climate change scenarios. Letters to Nature 416, 626–629.

Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. Global Ecology and Biogeography 17, 135–144.

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Radu, R., Déqué, M., Somot, S., 2008. Spectral nudging in a spectral regional climate model. Tellus A 60, 898–910.

Räisänen, J., 2007. How reliable are climate models? Tellus A 59, 2–29.

Randin, C.F., Engler, R., Normand, S., Zappa, M., Zimmermann, N.E., Pearman, P.B., Vittoz, P., Thuiller, W., Guisan, A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. Global Change Biology 15, 1557–1569.

Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K., Böner, J., 2010. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. Ecological Modelling 221, 1221–1229.

Swets, J., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Tadić, M.P., 2010. Gridded croatian climatology for 1961-1990. Theoretical and Applied Climatology 102, 87–103.

Thuiller, W., 2003. BIOMOD - Optimizing predictions of species distributions and projecting potential future shifts under global change. Glob. Change Biol. 9, 1353–1362.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007. A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distributions 13, 397–405.

van der Linden, P., Mitchell, J., 2009. ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Technical Report. Met Office Hadley Centre. Exeter, UK.

van Meijgaard, E., van Ulft, L., van de Berg, W., Bosveld, F., van den Hurk, B., Lenderink, G., Siebesma, A., 2008. The KNMI regional atmospheric climate model RACMO, version 2.1. Tech. Rep. 302. R. Neth. Meteorol. Inst.. De Bilt, Netherlands.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 106, 19729–19736. Arthur M Sackler Colloquium of the National-Academy-of-Sciences on Biogeography, Changing Climates and Niche Evolution, Irvine, CA, DEC 12-13, 2008.

Wilting, A., Cord, A., Hearn, A.J., Hesse, D., Mohamed, A., Traeholdt, C., Cheyne, S.M., Sunarto, S., Jayasilan, M.A., Ross, J., Shapiro, A.C., Sebastian, A., Dech, S., Breitenmoser, C., Sanderson, J., Duckworth, J.W., Hofer, H., 2010. Modelling the Species Distribution of Flat-Headed Cats (Prionailurus planiceps), an Endangered South-East Asian Small Felid. PLOS ONE 5.

Winkler, J.A., Palutikof, J.P., Andresen, J.A., Goodess, C.M., 1997. The Simulation of Daily Temperature Time Series from GCM Output. Part II: Sensitivity Analysis of an Empirical Transfer Function Methodology. Journal of Climate 10, 2514–2532.

Yamamoto, J.K., 2000. An alternative measure of the reliability of ordinary kriging estimates. Mathematical Geology 32, 489–509.

Zahn, M., von Storch, H., 2010. Decreased frequency of North Atlantic polar lows associated with future climate warming. Nature 467, 309–312.

Code	Variable definition	units	Mean	UAB error			WC error		
				RMSE	*rho*	*Bias*	*RMSE*	*rho*	*Bias*
BIO1	Mean annual temp.	$^{\circ}C$	10.64	0.49	0.97	0.05	0.47	0.97	0.12
BIO2	Mean diurnal temp. range	$^{\circ}C$	11.5	1.22	0.76	0.60	3.13	0.82	2.99
BIO3	Isothermality (BIO2/BIO7) $\times$ 100	%	44.53	2.24	0.43	-0.85	6.37	-0.20	-6.01
BIO4	Temp. seasonality ($\sigma \times 100$)	%	521.51	28.54	0.96	-11.87	44.7	0.95	29.05
BIO5	Max. temp. of warmest month	$^{\circ}C$	25.31	1.07	0.85	-0.4	2.6	0.77	-2.34
BIO6	Min. temp. of coldest month	$^{\circ}C$	-0.53	0.97	0.96	0.42	1.61	0.97	1.47
BIO7	Annual temp. range	$^{\circ}C$	25.84	1.74	0.94	-0.82	4.09	0.94	-3.81
BIO8	Mean temp. of wettest quarter	$^{\circ}C$	5.66	1.36	0.90	0.28	2.72	0.83	2.3
BIO9	Mean temp. of driest quarter	$^{\circ}C$	17.15	0.54	0.91	-0.05	0.8	0.88	-0.50
BIO10	Mean temp. of warmest quarter	$^{\circ}C$	17.26	0.53	0.92	-0.07	0.72	0.91	-0.41
BIO11	Mean temp. of coldest quarter	$^{\circ}C$	4.59	0.64	0.97	0.24	0.64	0.96	0.22
BIO12	Annual precip.	mm	1015.8	173.12	0.91	44.28	339.12	0.73	-130.69
BIO13	Precip. of wettest month	mm	128.38	22.71	0.92	-3.13	44.81	0.81	-22.72
BIO14	Precip. of driest month	mm	38.28	9.37	0.91	6.6	13.65	0.64	4.13
BIO15	Seasonality of precip.($cv \times 100$)	%	33.97	4.84	0.66	-3.42	10.79	0.15	-9.45
BIO16	Precip. of wettest quarter	mm	353.33	61.81	0.92	0.82	124.51	0.79	-63.51
BIO17	Precip. of driest quarter	mm	136.3	28.73	0.91	20.08	44.19	0.69	20.43
BIO18	Precip. of warmest quarter	mm	144.12	26.84	0.91	15.67	47.21	0.79	22.44
BIO19	Precip. of coldest quarter	mm	317.43	58.69	0.91	4.87	123.96	0.65	-68.2

Table 1: Summary of explanatory bioclimatic variables used for climate envelope models. The spatial mean values computed with the reference climatology (UC) are indicated in the fourth column. Errors of the other two climate datasets (UAB and WC) w.r.t. UC data are indicated in terms of their root mean square error (RMSE) Spearman's rho correlation (rho) and bias. σ = standard deviation, cv = coefficient of variation.

Institution	Model	boundary GCM	Reference
Centre National de Recherches Météorol.	RM4.5	CNRM-CM3	Radu et al. (2008)
Danish Meteorol. Inst.	HIRHAM5	CNRM-CM3	Christensen et al. (2006)
Koninklijk Nederlands Meteorol. Inst.	RACMO2	MPI-ECHAM5-r3	van Meijgaard et al. (2008)
Hadley Center/UK Met Office	HadRM3	HadCM3-Q0	Collins et al. (2006)
Abdus Salam Int. Centre for Theor. Phys.	RegCM3	HadCM3-Q0	Pal et al. (2007)
Max Planck Inst. for Meteorol.	REMO	MPI-ECHAM5-r3	Jacob et al. (2001)
Swedish Meteorol. and Hydrol. Inst.	RCA3.0	BCCR-BCM2	Kjellström et al. (2005)

Table 2: Summary of the ENSEMBLES regional climate models used in this study. The driving GCMs and related references are also indicated.

Dataset	BIO Variable
UC	9, 16*, 3, 18*, 14*
UAB	5, 2, 14*, 18*, 16*, 15*
WC	9, 5, 2, 14*, 19*

Table 3: Variable subsets resulting after the application of the variable selection procedure (Section 3.3) to each of the climate datasets. Variables are displayed in their order of inclusion in the models. Precipitation-related variables are marked with an asterisk.

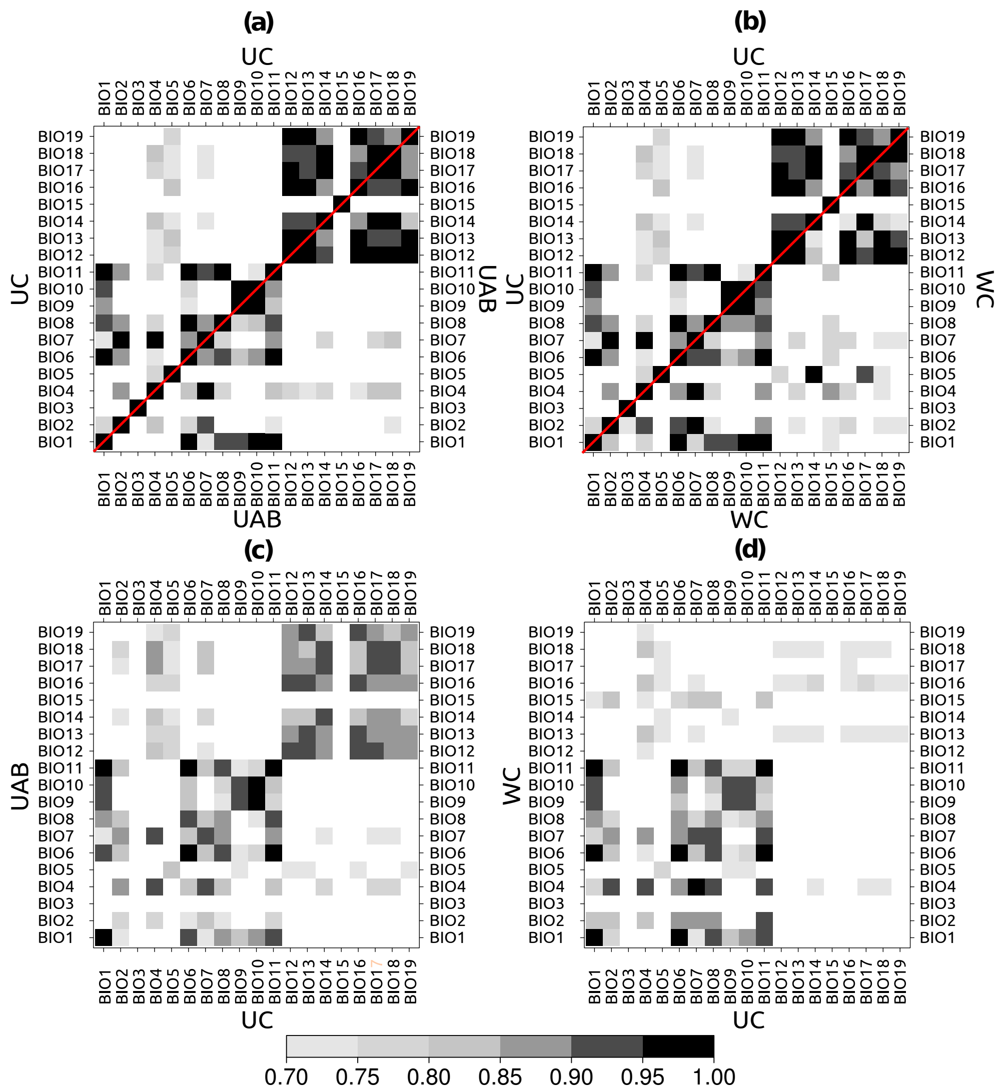

Figure 2: Pairwise cross–correlation matrices of the bioclimatic variables (Spearman's rho correlation coefficients. Values below 0.7 not shown). Intra-dataset correlation matrices (truncated) are displayed in the upper panels for UAB (a) and WC (b). Note that the benchmark UC dataset is represented in both panels (a and b) for better comparability. Inter-dataset correlation matrices are displayed in the lower panels: (c) UAB *vs.* UC and (d) WC *vs.* UC. Note that variables from BIO1 to BIO11 are related with temperature, and from BIO12 to BIO19 with precipitation (see Table 1 for details).

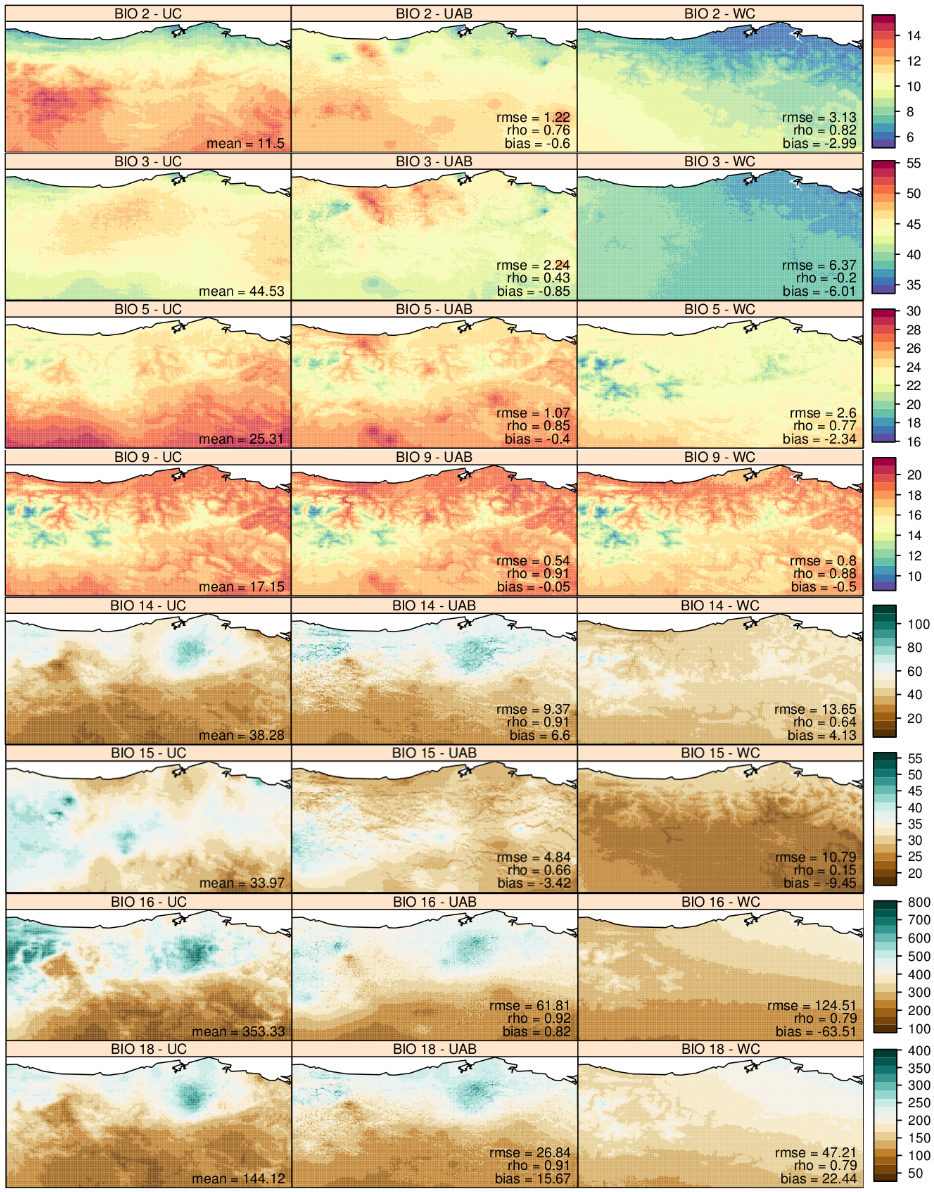

Figure 3: . Bioclimatic variables included in the UC, UAB and WC subsets after the variable selection procedure (Table 3). Mean UC values are indicated in the lower right hand side of the corresponding panels. For UAB and WC, the root mean square error (rmse), Spearman's rho correlation coefficient (rho) and bias with regard to UC are indicated. For details on variable definition and units see Table 1.

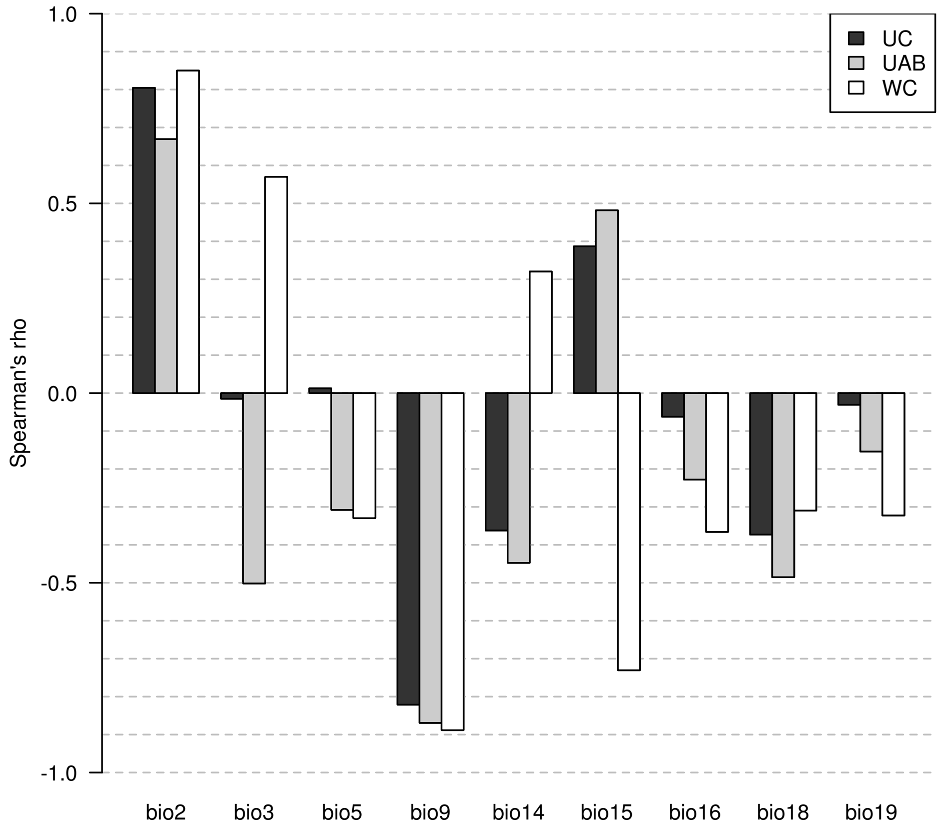

Figure 4: Correlation coefficients of the bioclimatic variables used in the different models with the terrain elevation, according to the three datasets tested.

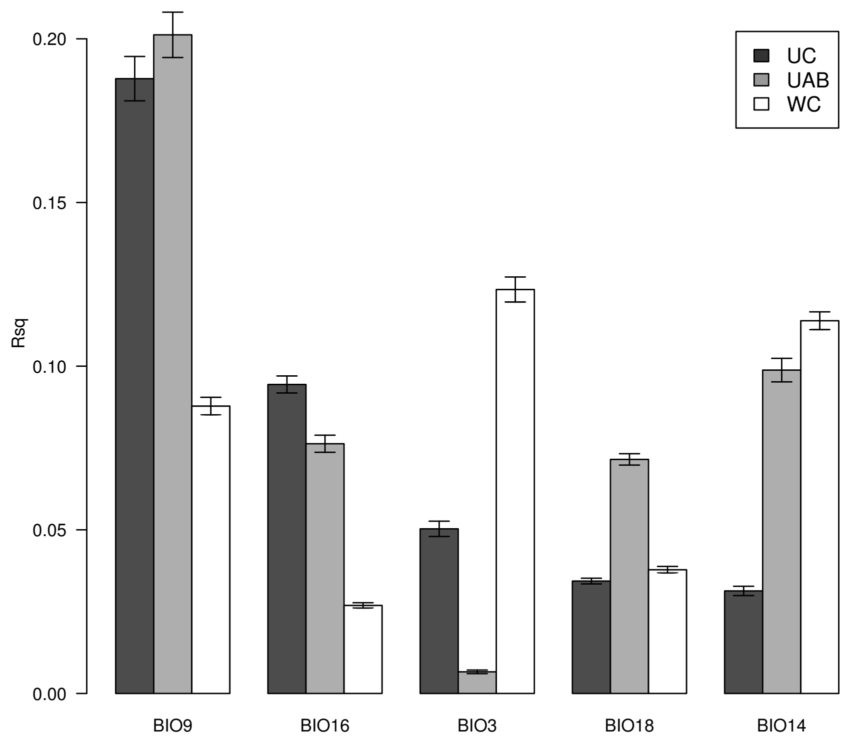

Figure 5: Variable importance (R^2) estimated as the independent effect of each variable following the hierarchical partitioning approach (Section 3.4). Variables selected correspond to the UC model selection. Values represented correspond to the mean ± standard deviation of the k=10 models of the cross validation test.

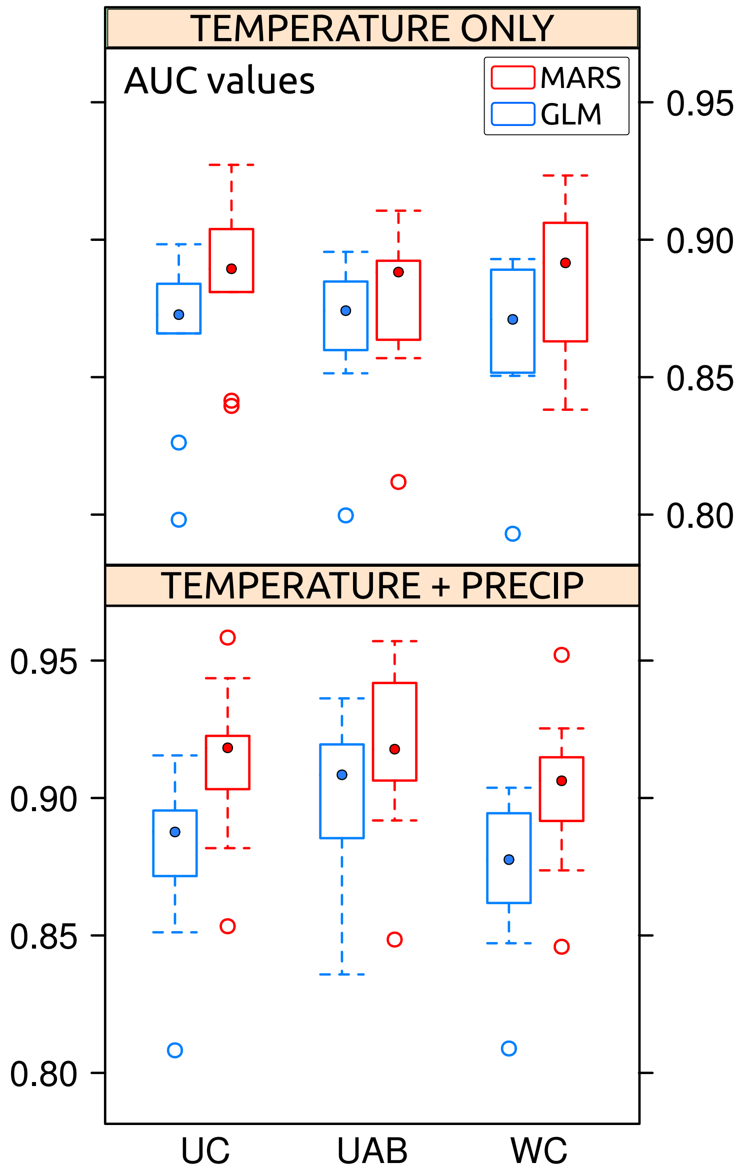

Figure 6: Area under the ROC curve (AUC) attained by the different CEMs in the 10–fold cross validation. The results are shown for both the temperature-only models, and for the temperature and precipitation models (using the variable subsets indicated in Table 3). The results are presented for both the GLM and the MARS algorithms.

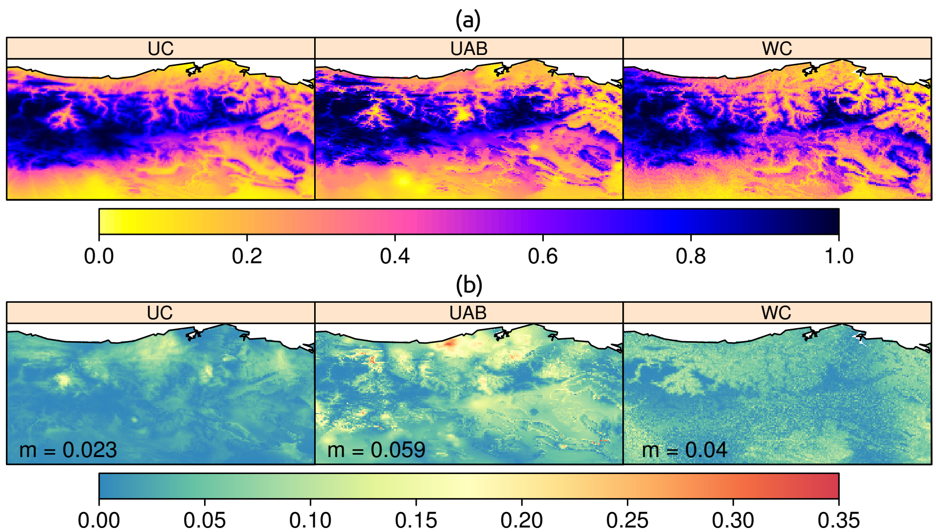

Figure 7: (a): Distribution maps obtained for *Fagus* according to the three datasets tested, using each one its corresponding subset of predictor variables (Table 3). (b): Multi predictor dataset uncertainty (standard deviation units) of the above models (spatial mean (m) is indicated for each panel).

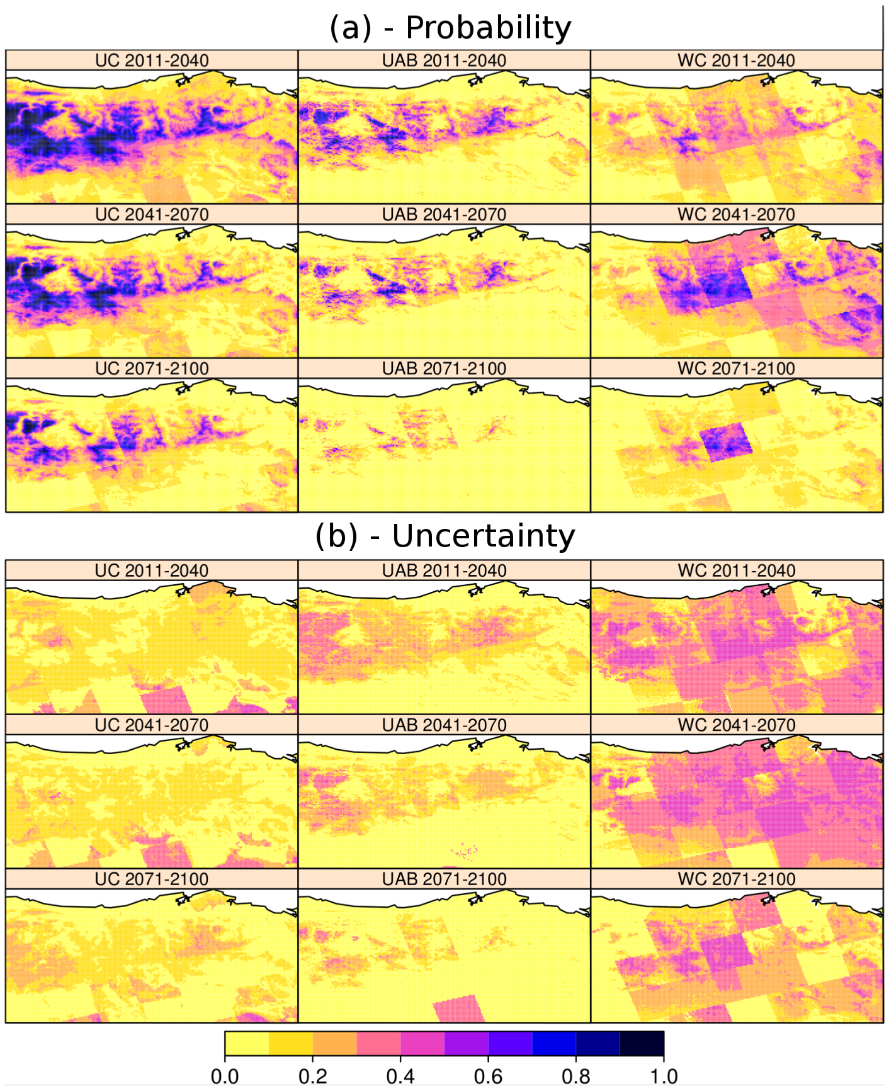

Figure 8: CEM future projections calculated according to the UC, UAB and WC climate datasets, using their respective subsets of predictors (Table 3). Maps in (a) represent the multi-RCM ensemble projections (Table 2) for the three future transient periods considered. Maps in (b) represent the standard deviation of the multi-model ensemble means.