

A new bioclimatic model calibrated with vegetation for Mediterranean forest areas

Michel VENNETIER^{1*}, Christian RIPERT¹, Eric MAILLE¹, Laurence BLANC², Franck TORRE³, Philip ROCHE¹, Thierry TATONI³, Jean-Jacques BRUN⁴

¹ Cemagref, UR Écosystèmes Méditerranéens et Risques, Aix en Provence, France

² ONEMA, Délégation Régionale Midi-Pyrénées-Aquitaine, France

³ Institut Méditerranéen d'Ecologie et Paléoécologie, UMR 6116, Université Paul Cézanne, Marseille, France

⁴ Cemagref, UR Écosystèmes Montagnards, Grenoble, France

(Received 4 March 2008; accepted 30 June 2008)

Keywords:
bioclimatic model /
Mediterranean forest /
water availability /
floristic analysis /
PLS regression /
Provence

Abstract

- Water availability is one of the main factors explaining flora composition and growth in Mediterranean regions, where it may decline with climate change.
- Our goal was to develop a model for forest site assessment in Mediterranean environments, focusing on water availability to assess potential vegetation composition and productivity in any places, whatever their level of disturbance.
- We designed a statistical model, using global climatic and geographic variables, as well as detailed local topographic and edaphic variables, to compute a bioclimatic index for Mediterranean forest environments. This model was calibrated in France with a flora index from 325 old forests. The model explained 80.3% of the flora index variance. The method fills a gap in existing models, bridging scales from the region to forest sites.
- Beyond its theoretical aspect, it was designed to allow practical tools to be derived from it for decision-making and management, such as the assessment of climate change impact on vegetation, and of forest productivity. Its development and adaptation is possible in other Mediterranean regions, and in any region where water is one of the main limiting factors.

Mots-clés :
modèle bioclimatique /
forêt méditerranéenne /
disponibilité en eau /
analyse floristique /
régression PLS /
Provence

Résumé – Un nouveau modèle bioclimatique calibré à l'aide de la flore pour les milieux forestiers méditerranéens.

- La disponibilité en eau est un des principaux facteurs contrôlant la composition et la croissance de la flore en région méditerranéenne. Elle devrait décroître avec le changement climatique.
- Notre objectif était de développer un modèle pour l'analyse stationnelle des milieux forestiers méditerranéens. Basé sur la disponibilité de l'eau, il devait permettre l'évaluation de leur flore potentielle et de leur productivité, quel que soit leur état de perturbation.
- Nous avons conçu un modèle statistique, utilisant conjointement des variables globales, climatiques et géographiques, et des variables locales topographiques et édaphiques pour calculer un indice bioclimatique adapté aux milieux forestiers méditerranéens. Ce modèle a été calibré en France sur 325 placettes de forêts âgées à l'aide d'un indice floristique dont il explique 80.3 % de la variance. Avec des échelles de travail allant en continu de la région à la station forestière, la méthode comble un vide dans la gamme des modèles existants.
- Ce modèle ouvre des perspectives pour l'évaluation de l'impact du changement climatique sur la flore et de la productivité forestière. La méthode peut être adaptée à d'autres régions méditerranéennes et à toute région où le bilan hydrique est un des principaux facteurs limitants.

* Corresponding author: michel.vennetier@cemagref.fr

1. INTRODUCTION

1.1. Context

Mediterranean forests of Europe and North Africa are particularly affected by climate change: drought is already one of the main environmental constraints (Daget, 1977) and the predictions of General Circulation Models forecast a faster warming than in most other continental areas, associated with a reduction of rainfall during the growth season (Hesselbjerg-Christiansen et al., 2007). This puts vegetation at risk of not adapting to local changes in its environment (Jump and Penuelas, 2005) nor following the fast shift in the limits of its climatic niche envelope. In order to assess and understand vegetation productivity and organization, and be able to forecast its response to expected global changes, scientists have developed many vegetation and bioclimatic models. However, few of them have been designed to predict vegetation structure or the distribution of dominant species on scales comprised between the region and forest sites. In California, Franklin (1998) and Meetenmeyer et al. (2001) compared various statistical methods to assess landscape-scale patterns of shrub-species abundance. They used climate data and geographic coordinates, adding local topographically mediated variables computed from digital maps such as slope, potential solar irradiance, substrate rockiness and soil moisture. Although the simulated distribution of plants generally fitted vegetation maps, error analysis suggested that important factors were not considered. A multiple scale approach (regional, landscape and site) is found in Ben Wu and Smeins (2000) for the prediction of rare plant occurrence in Texas (USA). However, each scale was tackled separately, local scale models using fine-resolution field data, while large scales models used independently coarse-scale geographic and climate information.

Although partly satisfactory according to their initial goals, most of the other attempts to model vegetation distribution and functions on local or intermediate scales come up against various problems, including human impact, which can partly hide the role of natural factors (Zimmermann and Kienast, 1999), the small size of the possible targeted area (Vogiatzakis and Griffiths, 2006), or difficulties accurately measuring some key variables (Mao et al., 2007). A thorough and realistic description of site characteristics seems necessary to understand and simulate vegetation composition, dynamics or productivity on a local scale (Rathgeber et al., 2003). This has rarely been done up to now, because local data are lacking or not sufficiently included in model parameters.

Vegetation analysis is usually an interesting tool for forest site assessment. Floristic composition and structure integrate all site conditions, including local conditions as well as global climate variables and their variability and extremes in the middle term, and the many interactions between all of them (Berges et al., 2006). Some of the plants are individually considered as good indicators of specific site conditions such as hydromorphy, pH, nutrient status or water availability (Gégout et al., 2003; Rameau et al., 1989). However, Mediterranean vegetation is generally disturbed by fire, grazing or clearing, so that the fast-changing vegetation is not representative of site

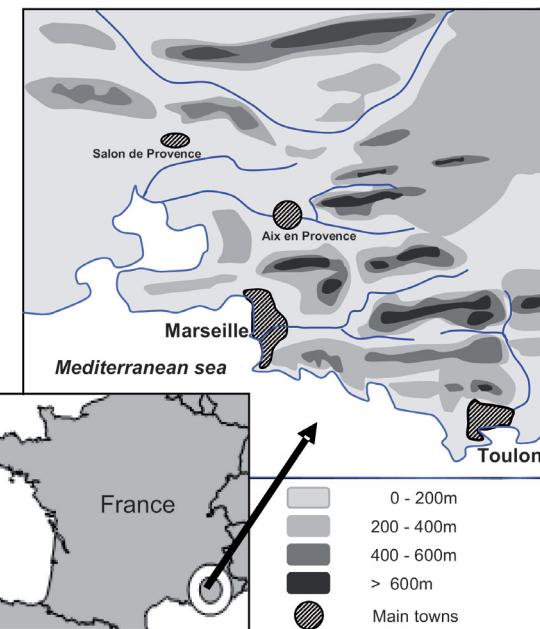


Figure 1. Map of the study area.

potential; in such conditions, site assessment must rely mainly on abiotic variables.

1.2. Basic hypothesis and goals

Our basic hypothesis was that water availability is the main constraint for Mediterranean vegetation (Le Houérou, 2005).

In this study, the main goal was to design a new model based on abiotic variables that could be used in Mediterranean environments for forest site assessment, in any places, whatever their level of disturbance. Focusing on water availability, it should help in assessing potential vegetation composition and productivity. Flora composition of undisturbed forests, as an integrated indicator, was used to calibrate this model, which was designed with three important additional goals to fill a gap in the range of existing models: (i) to integrate scales from very local to regional, (ii) to allow automatic mapping of its outputs on these various scales, so that vegetation dynamics could be simulated without scale constraints, and (iii) to contribute to the assessment of climate change impact on vegetation, as climate variations may directly and rapidly reduce water availability.

2. MATERIALS AND METHODS

2.1. Choice of the study area, stands and plots

The study area is located in Provence, South-Eastern France (Fig. 1). Its Mediterranean climate is characterized by a severe drought in summer (2 to 4 months), mild and humid winters, and a very low cloudiness. The mean annual temperature and rainfall range, respectively, from 15.3 °C/500 mm on the south-western

coast to 9.5 °C/1000 mm on the highest ridges, with an average of 13.2 °C/720 mm.

The limestone-based Provence was chosen for two reasons: (i) because some important soil parameters are homogenous: pH is stable and nutrients are rarely limiting, and (ii) because carbonated soils represent more than 80% of the Provence area. In this typical cuesta-type landscape, soils are generally shallow out of valleys and thalwegs, and ecosystem productivity depends to a large extent on nature and penetrability for the roots of superficial bedrocks.

Forests are dominated by unevenaged stands of *Pinus halepensis* Mill., most of the time mixed with *Quercus ilex* L. in the drier sites and *Quercus pubescens* Will. in more mesic sites and on deeper soils, and to a lesser extent by mixed stands of both oaks. In order to maximize the abiotic signal in the vegetation response, we selected only: (i) sites without any wildfire since the establishment of the dominant trees, (ii) old forests or forests established more than 70 years ago, and (iii) sites with no disturbing activity such as logging, grazing, clearing, trampling or prescribed fire in the last 30 years. In such controlled cases, vegetation should be representative of site conditions.

2.2. Sampling strategy

We used a stratified sampling design based on 4 main variables on a regional scale (for details about variables, see Table I): (1) bioclimatic zone (2 classes : meso-Mediterranean and supra-Mediterranean), (2) altitude in four classes: 200 m each between 0 and 800 m, (3) two main types of superficial soil layers (autochthonous = alterite ; allochthonous = colluvium, alluvium), and (4) potential solar irradiance (Becker, 1984) in 3 classes (hot, neutral and cool). Crossing these 4 main factors produced 48 possible combinations, among which 36 were present in the field. For each of these 36 combinations, a minimum of 7 and a maximum of 15 plots were chosen, balancing as far as possible for each combination and globally for the whole sample the classes of four secondary variables: parent rock, general topography, local site topography and total depth of soil observable layers.

Finally, 325 plots were surveyed, about one-third each year from 1996 to 1998. Their general size and shape is a 400 m² circle, a few plots being rectangular to guarantee homogeneous site conditions in narrow linear environments like thalwegs and ridges.

2.3. Variables

The survey includes:

- *Vegetation composition and structure*: (1) a flora census, using the classical Braun-Blanquet coefficients (Braun-Blanquet, 1932); (2) a description of the structure by the percentage of cover, in six height classes for each of the main tree species and for vegetation as a whole; (3) a wood core sample in five dominant Aleppo pines, the closest to the center of the plot, to assess their age with tree ring counting, and the height and circumference of these trees, and (4) when necessary and possible a count of tree rings on broadleaved species to assess the age of the last fire or logging.

The flora census was performed with a “variable-time strategy with minimum time limit” (Archaux et al., 2006), e.g. for a minimum of 1/2 h with a stop if no new species was found within 5 min.

- *Abiotic variables*: (1) climatic, geographic and orientation variables mapped with a GIS over the whole study area with a ground resolution of 50 m, and named globally “CG variables”; (2) variables which can be reliably measured only on a local scale, mainly topographic and edaphic, named “TE variables”.

For climatic variables, we interpolated the data from eighty-one meteorological stations with complete series over at least 38 years (1961–1998), over the whole study area and a peripheral buffer of 50 km to avoid border effects.

2.4. Statistical analyses

2.4.1. Designing the model

In order to identify floristic gradients that could be related to bioclimatic constraints and particularly water availability, we first performed a Correspondence Analysis (CA) with flora. We only took taxa present in at least three plots into account, keeping 192 plants among the 323 found in the plots.

We checked the relevance of the main CA axes towards bioclimatic gradients, introducing as supplementary variables the disjunctive classes of all abiotic variables (Becker, 1979). After testing several axes and gradients, alone or in pairs (Vennetier, 2007), the coordinate of each plot on the first CA axis was finally retained as its flora index (Fi).

In order to model this “Fi” index with relevant abiotic variables as predictors, we used Partial Least Squares (PLS) regression (Ter-Braak and Juggins, 1993). PLS regression was particularly adapted as it has been designed to handle many variables with relatively few observations (Cramer et al., 1988), and to cope with correlated variables (Tenenhaus, 1998). The final choice of variables was performed by an ascending and descending stepwise PLS regression. The predicted value of Fi for each plot, using abiotic variables in the PLS model, was considered as its bioclimatic index (Bi).

2.4.2. Robustness of the model design

We checked the robustness of the CA axes towards potential inaccuracies of floristic censuses due to time or spatial strategies or the observer (Archaux et al., 2006; 2007) and towards analysis options. This verification was performed with a Multiple Factorial Analysis (MFA) (Escofier and Pages, 1994) testing the stability of CA axes and plot coordinates: on one hand, by increasing the number of plots where a plant must be present to be taken into account from 3 to 30, and on the other hand, comparing presence/absence and Braun-Blanquet coefficients in the analysis.

Some variables seemed to have a non-linear relationship with Fi. To optimize the PLS model, we first checked the relation of all relevant variables to Fi using neural networks (Guiot, 1991) and transformed some of them. Before being validated, each transformation was tested in the model, comparing the transformed and the raw variable in a new neural network optimization, and in the PLS model.

Details on the resampling and robustness tests of all our statistical analyses are given in Annex 1.

Table I. Selected variables, their normalized coefficient, transformation and short description.

| | Variable | Transform | Coeff. | Description | Formula and classes |
|--------------|--------------|------------|--------|---|--|
| CG variables | ikr | \wedge^2 | -0.183 | Becker light-climate index | $IKR = (\text{SIN}(H \cdot \text{ATAN}(A * \text{COS}(O))) / \text{SIN}(H))$: with H = constant linked to the annual average angle with vertical of the sun at zenith (changes with latitude). A = Slope angle in radians. O = Slope orientation in radians. |
| Tan | | | -0.153 | Mean annual temperature ($^{\circ}\text{C}$) | 9.6–15.5 $^{\circ}\text{C}$ |
| alti | Sigmoidal | | 0.131 | Altitude (m) | 30–820 m |
| Rsu | log | | 0.115 | Summer rainfall : June to August cumulated rainfall (mm) | 58–175 mm |
| Res | | | 0.082 | Annual rainfall excluding summer rainfall (mm) | 454–913 mm |
| Amax0 | Polynomial 3 | | 0.169 | Mask to the sea ⁽¹⁾ ; maximum altitude between a site and the coastline in a given direction (m) | 30–1010 m. Selected directions are the nearest point of coast line (0) and 247°. |
| Amax24 | | | 0.146 | Distance to the sea ⁽¹⁾ in a given direction (km) | 1–118 km |
| Dsea247 | Sigmoidal | | 0.106 | | |
| TE variables | topG | | -0.136 | General topography on landscape and slope scales | 1 = bottom of valley or thalweg, 2 = lower concave slope, 3 = plain or plateau, 4 = slope, 5 = convex summit or ridge |
| topS | | | -0.107 | Topography on local scale (plot size). Crossing transversal and longitudinal topography in 3 classes each = Concave (+1), Plane (0), Convex (-1) | Five classes : -2 ; -1 ; 0 ; 1 ; 2 |
| Oc% | whc | | -0.083 | Percentage of parent rock outcrops on the plot | 0–95 % |
| Cf% | | | 0.100 | Water-holding capacity of earth (mm/cm) based on soil texture by totE variable | 0.7–2 mm/cm |
| totE | log | | -0.091 | Percentage of coarse fragments in the soil (for the depth defined by totE variable) | 0 ; 1–5 ; 6–10 ; 11–30 ; 31–60 ; 61–90 ; >90 %. The central value of each class is used for computation |
| | | | 0.119 | Total soil depth (cm) including: (i) alteritic and colluvium layers over unaltered hard parent bedrock, and (ii) total depth of alluvial or colluvium layers when they allow root development | 0–4.0 m (limited for computation to 1.6 m) |

¹ For the distance to the sea and the mask to the sea, three directions were initially chosen according to the main winds bringing the rain to the study area and the important buffer effect of the sea on temperature near the coast. Two were finally selected.

3. RESULTS

3.1. Robustness of the basic analysis

In the MFA factorial map, the first two CA axes as well as the coordinates and ranks of plots on these axes were very stable ($R^2 > 0.97$ for the coordinates) when the minimum occurrence for plants to be taken into account was increased from 3 to 25, cutting their number by more than 2. Only from the limit of 30 plots of occurrence, which means the elimination of more than 75% of all the plants and of 60% of those with at least 3 occurrences, a significant rotation of axes occurred and differences were found in the relative rank of plots on the first CA axis.

Accordingly, the 4 main CA axes and coordinates of plots were very stable whatever the code used for flora (presence/absence vs. Braun-Blanquet code: $R^2 = 0.98$ on the first CA axis).

Neural Network analyses of 24 CG variables and 27 TE variables confirmed that the relations between abiotic variables and F_i were not always linear and required transformations, particularly for the CG group (Tab. I). The main TE variables did not require transformations, except soil depth.

3.2. Flora analysis and the bioclimatic index

An average of 25 taxa were found in plots (8 to 51, $sd = 7.8$). With an eigenvalue twice that of axis 2, the first CA axis was dominant by far (Fig. 2a). In this study, we did not focus on the ecological interpretation of individual plants and their position in the CA plane, but only on the main gradients resulting from flora composition as a whole.

Figure 2b displays as an example the classes of some of the main CG variables projected as supplementary variables on the first CA plane. All CG variables were well ordered throughout this map and correlated with the first axis. The same result was obtained with all TE variables related to water availability. Variables describing vegetation structure were not correlated with the first two CA axes.

Figure 2c shows a synthetic representation of four main ecological gradients interpreted from the distribution of all variables in the CA map. These gradients were calculated with a linear regression on the coordinates of the classes of variables concerned in the plane:

- (1) climate and geographic gradient: rainfall, temperature, altitude and continentality (resulting from Fig. 2b),
- (2) orientation gradient, taking into account orientation on all scales,
- (3) soil texture gradient: sand vs. clay and silt percentages, and related water-holding capacity (whc),
- (4) topography (from local to landscape scales) and soil quality gradient (local TE variables).

Without any exception, all variables had in the left half of the CA plane their classes related to low water availability (dry and hot climate, rocky or shallow soils, low water-holding capacity and unfavorable topography), and in the right half of the

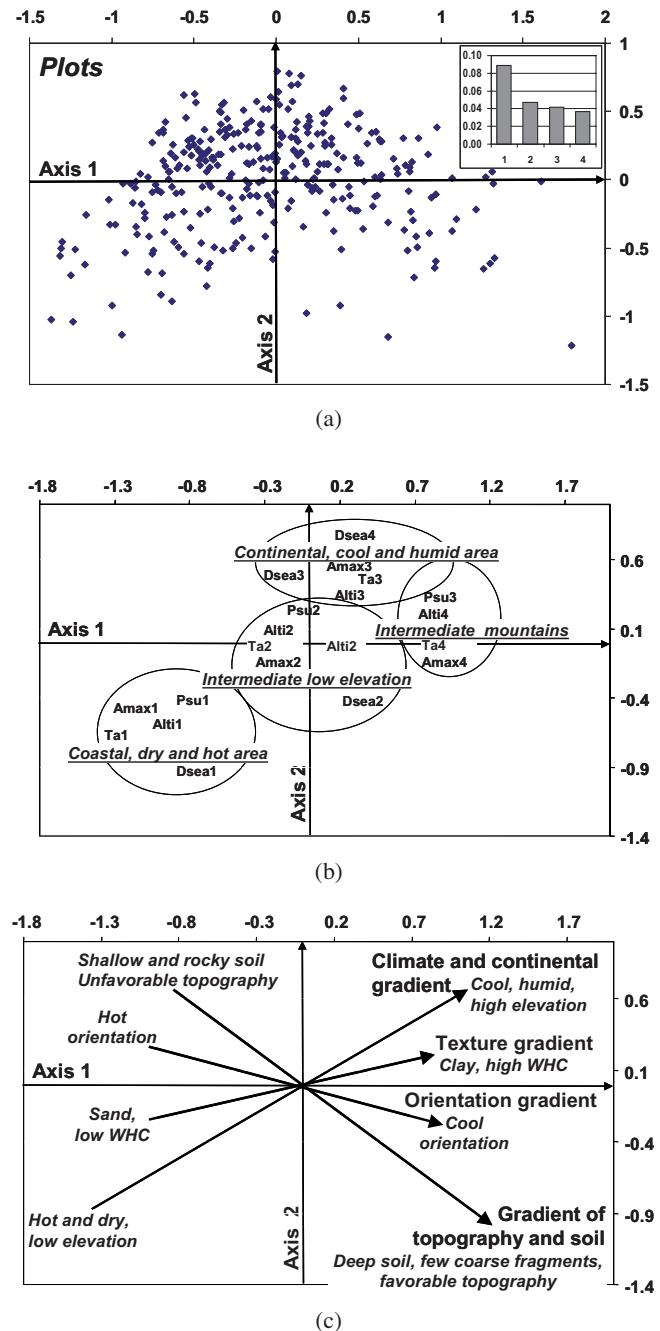


Figure 2. (a) CA eigenvalues and distribution of the 325 plots in the main CA factorial map (axes 1/2). Axis No. 1 is by far dominant (twice the eigenvalue of No. 2). The slight arch effect is partly due to the geographical structure of the study area (see Fig 2b) with intermediate mountains close to the sea. (b) Disjunctive classes of some of the main CG variables (climate, continentality) projected as supplementary variables in the CA map (axes 1/2). A clear ordination appears with all hot and dry situations bottom left and cool and humid ones on the top right. The resulting climate and continental synthetic gradient is shown in fig 2c with three other gradients obtained with the same methods and other groups of variables. (c) Synthetic representation of the 4 ecological gradients combining all relevant variables by groups.

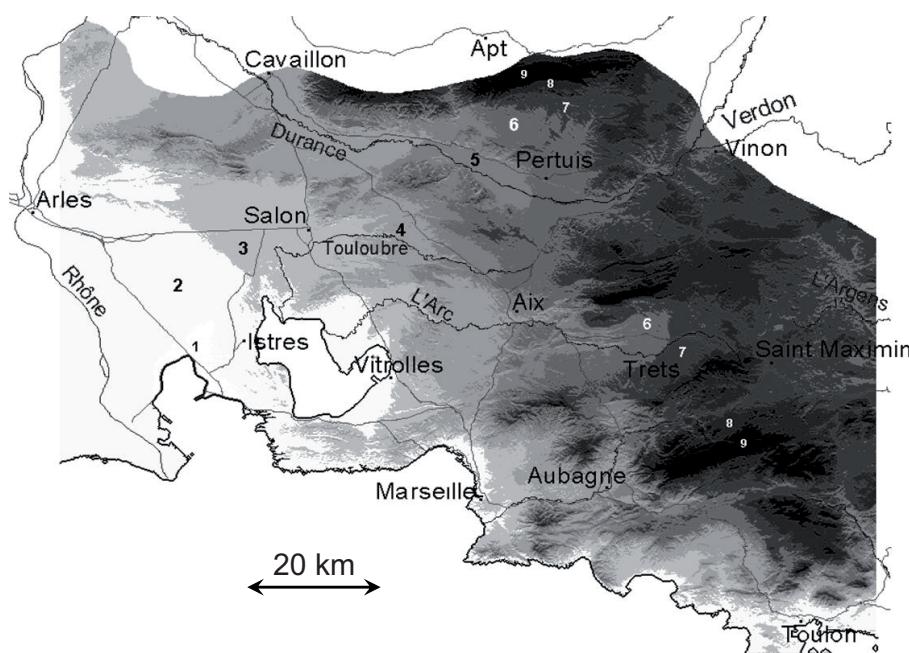


Figure 3. Regional map of the global bioclimatic index (Bi-CG) computed from 8 CG variables, centered and split into 9 classes, from 1 (the driest and hottest sites), to 9 (the more mesic ones). Limits of the classes: -1.101 ; -0.496 ; -0.357 ; -0.218 ; -0.541 ; 0.520 ; 0.216 ; 0.380 ; 0.601 ; 1.71. Maximum and minimum predicted values for plots: -0.832 and 1.03. The extreme positive values come from very steep north slopes with elevations between 800 and 1100 m (less than 1% of the study area not taken into account in the calibration of the model) ; the minimum prediction for plots, just a bit higher than the minimum value for the study area, shows that our sample correctly represented very dry and hot sites.

plane classes describing opposite situations. This first axis, resulting only from flora distribution, could be clearly related to water availability, which confirms our basic hypothesis.

Considering its dominance, this first axis was retained alone for modeling. The coordinate of each plot on this axis was considered as its flora index (Fi). This index was modeled, keeping finally 14 raw or transformed CG and TE variables (Tab. I) to obtain its abiotic estimate (Bi). The model explained 80.3% of Fi variance, among which CG and TE groups of variables explained, respectively, 47.7% and 32.6%.

We computed the bioclimatic index (Bi) for each plot, splitting it into two components based on the two groups of variables: a global index (Bi-CG), and a local index (Bi-TE). Although linked in the model by the common computation and probable interactions, these two indexes can be calculated separately.

3.3. Mapping the Bi-CG bioclimatic index

Figure 3 shows the map of the Bi-CG index split into 9 classes on a regional scale. Each class includes 1/9 of the total variation interval after exclusion of the 5% extreme values (2.5% at each end). These extreme values were merged with the first and last classes, respectively.

4. DISCUSSION

4.1. Robustness of the modeling approach based on flora

Although the flora is rather rich, the MFA proved that 30% of the plants, the most frequent, all perennial or easy to recognize in any season, were enough for expressing the water availability bioclimatic gradients; so that overlooking a few species during the flora census could not have modified the bases of our model built on the whole flora. As also demonstrated by the MFA, a possible slight variability in the notation of Braun-Blanquet codes between the experienced botanists which participated in the study would have had no influence on the result. Our model is thus very robust towards the variations and mistakes of random or human origin which usually spoil vegetation surveys. This can be explained by the overwhelming place held here by water availability in interaction with temperature, deeply structuring plant communities. A slight arch effect appeared in the factorial map (Fig. 2a) which could partly be explained by the geographic structure of the study area. Whatever its causes, we only used axis 1 for modeling, so that detrending the CA was not useful.

As we wanted the flora index to be used for the calibration of the abiotic model, and thus to be independent of environmental variables, whose list was long and not fixed a priori, Canonical Correspondence Analysis (CCA), although a standard method to link biotic gradients and environmental variables, did not fit our requirements.

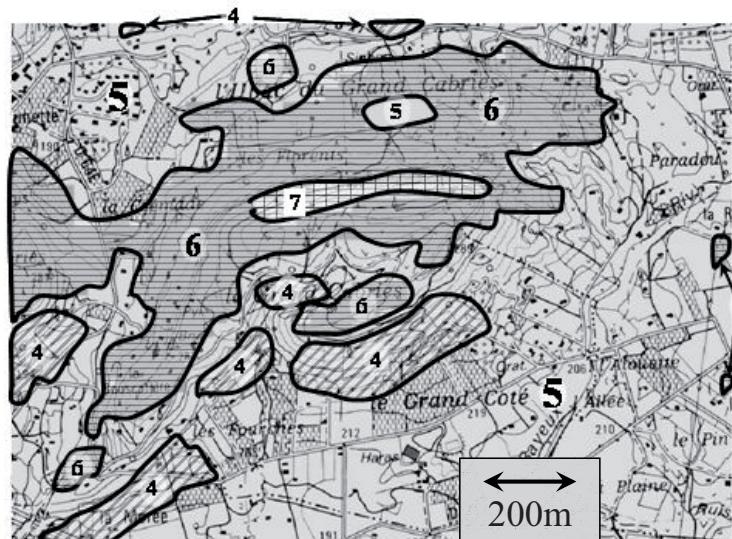


Figure 4. Zoom in Figure 3 on the scale of a small local hill, with the same classes of Igb(CG) index as Figure 3, showing the high fragmentation of vegetation potential niche envelopes.

4.2. Weight of variables and their interactions

Vegetation of relatively undisturbed Mediterranean forests appeared to be a reliable indicator of site conditions, in accordance with what is generally observed in other regions (Gégoût et al., 2005). Reciprocally, a restricted number of environmental variables related to water availability allowed the reliable assessment of the Flora index. This should help in assessing the potential flora of any site of the study area, whatever its present state of evolution or disturbance. Among these variables, the set of global ones, easy to compute over large areas, allowed the calculation and mapping of a geographically and biologically wise bioclimatic index.

The Becker light-climate index (ikr) had the highest weight among individual variables. It expresses the contrast between cool and warm hillsides, particularly marked in the study area because of the east-west dominant orientation of mountain ranges, which creates a majority of large north- or south-facing slopes. Ikr is also the only global variable which remains a key variable on a local scale.

The specific position of the main mountain ranges perpendicular to humid winds coming from the sea creates a positive correlation between rainfall and altitude ($r^2 = 0.54$), and negative between rainfall and temperature ($r^2 = 0.44$). Furthermore, any rise in temperature increases evaporation and evapotranspiration by plants, and thus influences negatively the global and local water balance. It is thus impossible in the model to accurately separate the respective roles of rainfall, temperature and their interaction. Functionally, the contrast between south- and north-facing slopes, as well as the contrast between low and high altitudes are as much a matter of aridity as a matter of thermal contrast.

Geographic variables, defining together a continentality index, carry weight in the model. Although expressing mainly humidity and temperature gradients, they could not be re-

moved without a loss of more than 10% of the total explained variance. They integrate local effects of mountain ranges (cloudiness, cold air flows, temperature inversions, wind circulations and accelerations), which cannot be assessed with the interpolation of basic climatic variables.

All selected TE variables could be directly linked to the water gradient on a local scale. A soil useful water reserve is of paramount importance. This reserve combines several of our TE variables: soil depth (totE), water-holding capacity (whc) and the percentage of coarse fragments (Cf%).

Topography proved to be the most important TE factor. Topographic variables have a direct role in water availability, as local water circulation, concentration or drainage depends a lot on the shape of landscape on local and slope scales: concave, plane or convex. Soil quality and depth are also correlated with topography, although our sampling design limited this interaction.

4.3. The operational scales of the model and its potential for assessing climate change impact

Even when mapped with only the Bi-CG index, the landscapes of the study area appeared as a highly fragmented patchwork on a local scale (Fig. 4). The contrast between both sides of a small hill may display locally up to 4 classes among the 9 found on a regional scale.

As TE variables explain one-third of Fi variance, adding the TE component increases the variability of this mosaic: every Bi-CG class may include several contrasted classes of Bi-TE on a local scale, particularly due to topography and soil depth, but also bedrock outcrops. When both components of the model are added, differences equivalent to 5 or 6 classes

among the 9 of Bi-CG can be found within a few hectares of hilly landscape.

The local limits of bioclimatic envelopes are thus very irregular and fragmented: many enclaves of meso-Mediterranean microclimate exist far inside the global envelope of the supra-Mediterranean area and vice versa. Thus, taxa adapted to warm and dry environments, liable to replace less tolerant ones in the case of climate warming and drying, are already scattered in the landscapes of the hinterland and ready to spread from these advanced bases. In contrast, numerous cooler or wetter microclimatic and edaphic poles are disseminated in the landscape at low altitude. They should allow the water-demanding species to survive beyond the average limit of their distribution area, at least for some time.

The notion of niches as redefined by Chase and Leibold (2003) allows the integration of evolutionary phenomena across levels and scales of organization. Although some models tried to predict the assemblage of species on the basis of both global and some local variables (Franklin, 2002), very few were made to downscale these modeled niches from global to very local scales. In forest environments, precise soil maps are rarely available. However, simulating the variability of edaphic variables is possible by combining existing models of soil type distribution in landscapes (Scull et al., 2003) and local validation and calibration tests (Lagacherie et al., 1995). Our model is the only one in which the topographic structure of landscape on different scales is integrated directly as variables, allowing either a precise site assessment on a local scale on observed data, or an automatic downscaling on simulated data. Mapping precisely both our indexes and derived ecological niches could be more fully automated in the future.

The classes chosen to map the Bi-CG index in this paper are arbitrary and could be adapted, split or grouped together to fit the specific conditions of particular sites, the limits of the observed distribution area of a species or of a nature reserve, for example. Characterizing the ecological niche of the main species or groups of species with a combination of the two components of our model can contribute to the assessment and mapping of their current and future potential niche distribution. The model may also help in evaluating realistically the potential dispersal of species throughout the landscape with various scenarios of climate change, as its ground definition is adapted to the average seed dispersal distance of the majority of plant species (Clark et al., 1999). This downscaling capacity is among its most original contributions. Furthermore, being able to integrate several scales allows one to approach the impact of climate change for species as well as for metacommunities (Leibold et al., 2004).

5. SYNTHESIS AND PROSPECTS

5.1. Other potential scientific and operational uses of the model

The relation established between flora composition and environmental abiotic variables comes out in many potential scientific and operational outputs, particularly the assessment of

the impact of climate change, the follow-up of vegetation dynamics and the assessment of forest productivity.

The Bi indexes are by themselves good indicators which can be used in forest site assessment and mapping, and to compare various sites at different locations for experimental designs requiring homogeneous constraints. A link can be established between forest productivity and both components of the model, as a single homogeneous framework on large areas (Vennetier, 2007).

It is possible for each plot to compare on one hand, Bi variations with various expected changes in climate, and on the other hand, Fi variations with a simulated vegetation shift. Fitting Fi and Bi variations may help in assessing climate change impact on vegetation. Resampling our plots in time will allow the same comparison with real climate and flora data.

The difference between the Fi index of burned or disturbed sites, or abandoned land, with their Bi index and the follow up of this difference in time can show how far they are from recovery and how fast they are recovering.

With fast climate warming, wildfire frequency and intensity should increase. This uncontrolled hazard may be the major factor reshaping the vegetation, faster and deeper than the climate change itself (Cary et al., 2006). The risk is all the more important as Bi indexes are low. The simulation and mapping of Bi-CG spatial evolution with climate change, combined with Bi-TE maps, could be used as inputs in fire risk modeling at landscape level.

5.2. Conclusion

We designed a new bioclimatic model adapted to the Mediterranean context. Computing a bioclimatic index which explains a large percentage of the observed variations of flora, it confirmed that water availability was one of the main factors shaping Mediterranean vegetation.

It is related to statistical models (Guisan et al., 2002) whose goals are to propose mathematical bases for the interpretation of ecological relationships observed or suggested by general models, and particularly to highlight the decomposition of the role and the respective weight of the explanatory variables.

Calibrated with an analysis of the global flora of studied sites, and thanks to the very detailed description of key local variables including topography, this model has no current equivalent in terms of design and of span of operational scales. It allows one to analyze the bioclimatic structures of landscapes and may help in assessing accurately the impact of climate change on vegetation on very local to regional scales.

The development of the method is possible in other regions: (1) in all the French Mediterranean area; (2) in any area where water is the main limiting factor and where a sufficient number of sites remain in a good state of conservation for a balanced flora sampling, and (3) in other regions where any single limiting factor for vegetation is dominant.

Acknowledgements: This research was funded by the French Ministry for Agriculture, Fishing and Foods, the Provence-Alpes-Côte d'Azur Région, the Ministry for the Environment and Sustainable

Development (GICC program, REFORME project), and the Cemagref. The authors are indebted to the technicians and students who contributed together to more than 200 days of field work over several years in sometimes difficult weather conditions, particularly Olivier Chandioux, Roland Estève, Fabien Brochiéro, Yaacoub Nassif and Cyrille Rathgeber. A detailed review of the manuscript by Bernard Prevosto helped improve the final version of this paper.

REFERENCES

- Archaux F., Berges L., and Chevalier R., 2007. Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecol.* 188: 179–190.
- Archaux F., Gosselin F., Berges L., and Chevalier R., 2006. Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *J. Veg. Sci.* 17: 299–306.
- Becker M., 1979. Une étude phyto-écologique sur les plateaux calcaires du Nord-Est (Massif de Haye, 54). Utilisation de l’AFC dans la typologie des stations. Relations avec la productivité et la qualité du hêtre et du chêne. *Ann. Sci. For.* 36: 93–124.
- Becker M., 1984. Indices de climat lumineux selon la pente et l’exposition pour les latitudes de 40 à 50°. *Bull. Ecol.* 15: 239–252.
- Ben Wu X. and Smeins F.E., 2000. Multiple-scale habitat modeling approach for rare plant conservation. *Landscape Urban Plan.* 51: 11–28.
- Berges L., Gégout J.C., and Franc A., 2006. Can understory vegetation accurately predict site index? A comparative study using floristic and abiotic indices in sessile oak (*Quercus petraea* Liebl.) stands in northern France. *Ann. For. Sci.* 63: 31–42.
- Braun-Blanquet J., 1932. Plant sociology. The study of plant communities. McGraw-Hill, New-York.
- Cary G.J., Keane R.E., Gardner R.H., Lavorel S., Flannigan M.D., Davies I.D., Li C., Lenihan J.M., Rupp T.S., and Mouillot F., 2006. Comparison of the sensitivity of landscape-fire-succession models to variation in terrain, fuel pattern, climate and weather. *Landscape Ecol.* 21: 121–137.
- Chase J.M. and Leibold M.A., 2003. Ecological niches. University of Chicago Press, Chicago.
- Clark J.S., Silman M., Kern R., Macklin E., and Hillerislambers J., 1999. Seed dispersal near and far: Patterns across temperate and tropical forests. *Ecology* 80: 1475–1494.
- Cramer R.D.I., Bunce J.D., Paterson D.E., and Frank I.E., 1988. Crossvalidation, bootstrapping, and partial least square compared with multiple regression in conventional QSAR studies. *Quantit Struct.-Act. Relatsh.* 7: 18–25.
- Daget P., 1977. Le bioclimat méditerranéen : caractères généraux, modes de caractérisation. *Vegetatio*, 34: 1–20.
- Escofier B. and Pages J., 1994. Multiple Factor analysis. *Comput. Stat. Data Anal.* 18: 121–140.
- Franklin J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9: 733–748.
- Franklin J., 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. *Appl. Veg. Sci.* 5: 135–146.
- Gégout J.-C., Coudun C., Bailly G., and Jabiol B., 2005. EcoPlant: a forest sites database to link floristic data with soil resources and climatic conditions. *J. Veg. Sci.* 16: 257–260.
- Gégout J.C., Hervé J.C., Houllier F., and Pierrat J.C., 2003. Prediction of forest soil nutrient status using vegetation. *J. Veg. Sci.* 14: 55–62.
- Guiot J., 1991. Methods and programs of statistics for paleoclimatology and paleoecology. J. Guiot & L. Labeyrie Ed., Marseille.
- Guisan A., Edwards T.C., and Hastie T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157: 89–100.
- Hesselbjerg-Christiansen J. and Hewitson B., 2007. Regional Climate Projection. In: IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt K.B., Tignor M., and Miller H.L., (Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 p.
- Jump A.S. and Penuelas J., 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* 8: 1010–1020.
- Lagacherie P., Legros J.P., and Burrough P.A., 1995. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. *Geoderma* 65: 283–301.
- Le Houérou H.N., 2005. The iso-climatic mediterranean biomes: bioclimatology, diversity and phytogeography, Vol. 1 and 2. Le Houérou ed., Montpellier.
- Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F., Holt R.D., Shurin J.B., Law R., Tilman D., Loreau M., and Gonzalez A., 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7: 601–613.
- Mao J.F., Wang B., Dai Y.J., Woodward F.I., Hanson P.J., and Lomas M.R., 2007. Improvements of a dynamic global vegetation model and simulations of carbon and water at an upland-oak forest. *Adv. Atmos. Sci.* 24: 311–322.
- Meetenmeyer R.K., Moody A., and Franklin J., 2001. Landscape-scale patterns of shrub-species abundance in California chaparral - The role of topographically mediated resource gradients. *Plant Ecol.* 156: 19–41.
- R_Development_Core_Team, 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, R Foundation for Statistical Computing.
- Rameau J.-C., Mansion D., Dumé G., Timbal J., Lecomte A., Dupont P., and Keller R., 1989. Flore forestière française. Guide écologique illustré, 1, Plaines et Collines. Institut pour le Développement Forestier, Paris.
- Rathgeber C., Nicault A., Kaplan J.O., and Guiot J., 2003. Using a biogeochemistry model in simulating forests productivity responses to climatic change and [CO₂] increase: example of *Pinus halepensis* in Provence (south-east France). *Ecol. Model.* 166: 239–255.
- Scull P., Franklin J., Chadwick O.A., and McArthur D., 2003. Predictive soil mapping: a review. *Prog. Phys. Geogr.* (27) 2: 171–197.
- Tenenhaus M., 1998. La régression PLS. Théorie et pratiques. Editions Technip, Paris.
- Ter-Braak C.J.T. and Juggins S., 1993. Weighted averaging partial least square regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269/270: 485–502.
- Thioulouse J., Chessel D., Doledec S., and Olivier J.M., 1997. ADE-4: a multivariate analysis and graphical display software. *Stat. Comput.* 7: 75–83.
- Vennetier M., 2007. Un nouveau modèle bioclimatique pour la forêt méditerranéenne. Application à l’étude de l’impact du changement climatique sur la végétation et à l’évaluation de la productivité forestière. Thèse de doctorat, UR Écosystèmes méditerranéens et risques, Cemagref / Université Paul Cézanne (Aix-Marseille III) Aix en Provence.
- Vogiatzakis I.N., Griffiths G.H., 2006. A GIS-based empirical model for vegetation prediction in Lefka Ori, Crete. *Plant Ecol.* 184: 311–323.
- Zimmermann N.E., Kienast F., 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *J. Veg. Sci.* (10) 4: 469–482.

Annex 1: Validation and resampling methods in the statistical analyses.

The number of significant components for PLS regression was chosen with a 10,000-replication permutation test on observations, keeping components whose percentage of explained variance was not passed by more than 5 % of the permutations. With significant components, variables were sorted through a 1,000-resampling cross-validation test ; only variables whose confidence interval (95%) for the partial correlation coefficient excluded 0 were used. For the choice of relevant variables in the model, each step of the ascending and descending stepwise PLS regression was validated by a permutation test on PLS components and a cross-validation for variables concerned. All variables in the final version of the model were highly significant on the first two PLS components ($p < 0.001$).

Neural networks were used combining variables 6 by 6. Each neural network was optimized with a 200-replication

bootstrap, each replication including 10^4 calibration steps. After the optimization of the neural network, the response of each variable was plotted on its whole variation interval, the other variables being maintained at their mean value if they were not correlated with the tested one, or maintained successively at their first, second and third quartile for those which were correlated. In the last case, the 3 responses were combined in a sliding weighted mean to obtain the global response. We only transformed a variable according to a non-linear relation shown by neural networks when this relation was stable throughout these many tests and enhanced the total explained variance or its partial correlation coefficient.

We used ADE4 software (Thioulouse et al., 1997) for CA, MFA and PLS permutation tests, Statgraphics® software for stepwise PLS regression, R software (R_Development_Core_Team, 2004) for the cross-validation of PLS variables, and PPPhalos software (Guiot, 1991) for neural networks.