11 Greenhouses, land use change, and predictive models: MaxEnt and Geomod working together

Benito de Pando B and Peñas de Giles J

Abstract
We have developed a methodology which predicts the expansion of greenhouses and evaluates the results, combining a species distribution model (MaxEnt) and a simulator of land use change (Geomod). In the simulations, we take into account not only the effect of different environmental variables governing greenhouse expansion but also the spatial distribution of the error. The method has been tested on a region of SE Spain to establish future greenhouse-expansion scenarios. The results indicate that the combination of MaxEnt and Geomod improves the predictive capacity, as well as the functional interpretation of the land use change models.

Keywords: Geomod, MaxEnt, land use, distribution model.

11.1 Introduction

In the context of global change, the study of land use change takes on great relevance because small changes at the local scale (plots of a few ha), added together, can exert an impact on the scale of the entire planet. An example is the deforestation of tropical jungles, which diminishes atmospheric carbon fixation, imposing long-term consequences on global climate (Dixon et al. 1994). An analogous problem of emerging importance involves the expansion of greenhouses, a form of industrial agriculture that is developing on a grand scale in certain regions of the planet. In 1999, an estimated 682,000 ha were occupied by greenhouses throughout the world, especially in China (380,000 ha), followed by Mediterranean countries (161,300 ha in France Italy, Spain, Greece, Turkey, Morocco, and Algeria (Takakura and Fang 2002).

The problems arising from the spread of greenhouses are directly related, on the one hand, to the natural resources available in the affected region (biodiversity, natural habitats, water resources, etc.) and, on the other hand, to human resources (nearby populations). The construction of greenhouses covers the soil, depriving it of its ecological functions (evapotranspiration, infiltration of precipitation, supporting habitats, etc.), and it degrades the dynamics of natural habitats by fragmenting and destroying them. Greenhouse crops, though designed to make maximum use of irrigation, nevertheless demand huge quantities of water, altering the regime of aquifers. Other problems associated with greenhouses that can concern human health are plastic waste and organic debris contaminated by pesticides and fertilizers.

In the last two decades, the European food market has generated a high demand for fresh vegetables and fruits, triggering the uncontrolled proliferation of greenhouses in productive regions. The growth rate of greenhouses and the lack of a territorial management policy have wreaked havoc, inflicting grave environmental repercussions. In this context of uncontrolled land use change, management plans are indispensable for balanced regional development in which the economy and natural conservation are in balance.

Some GIS-based methods are useful to design and improve land use management plans, such as the land use and cover change simulations (LUCCs) (e.g., cellular automata, Geomod or Markov chains), which experimentally replicate the transition between land uses (Pontius et al. 2001, Jantz et al. 2003, Pontius and Pacheco 2004, Aguilera 2006). Other applicable methods are the species distribution models (SDMs) (e.g. Bioclim, GARP, MaxEnt), which provide knowledge on the potential distribution of targeted species and are increasingly in use for the design of conservation plans (Guisan and Zimmermann 2000, Posillico et al. 2004, Johnson and Gillingham 2005).

In this paper, we propose a method to predict land use change based on the integration of SDMs and LUCCs. The main idea is to use MaxEnt to compute distribution models, and use them in Geomod as suitability maps to perform better land use change simulations.

The main objectives of this paper are:

- Compare simulations performed by Geomod used in stand alone mode with the combined simulations performed by Geomod and MaxEnt, to test the feasibility of integrating the two methods.
- Introduce Procrustes analysis as a tool to evaluate the spatial agreement between simulations and ground-truth information.
- Introduce an easy method to compute the spatial distribution of certainty in Geomod simulations in order to generate certainty maps for assessing simulation accuracy.
- Test the proposed method in the period 1987-2001 (using 1987 data to calibrate and 2001 data to evaluate) to perform simulations for the period 2001-2010, in order to provide and explore three future scenarios of spreading of greenhouses.

11.2 Test area, data sets and tools

11.2.1 Test area

The test area selected was the province of Almeria (SE Spain, see Fig. 11.1), located between 3.14ºE and 1.62ºE longitude and 36.6ºN and 37.46ºN latitude (Fig. 11.1). The surface area analysed is 7,171 km². The climate is Mediterranean, with rainfall of 200-300 mm and means annual temperatures of 16-17ºC. Geologically, post-orogenic sedimentary materials predominate, and the landscape is dominated by a mosaic of chamaephyte plant communities, xerophytic grasslands and varied communities of annual plants. Greenhouses have been spreading in the area since 1960, occupying around 37,000 ha in 2001.

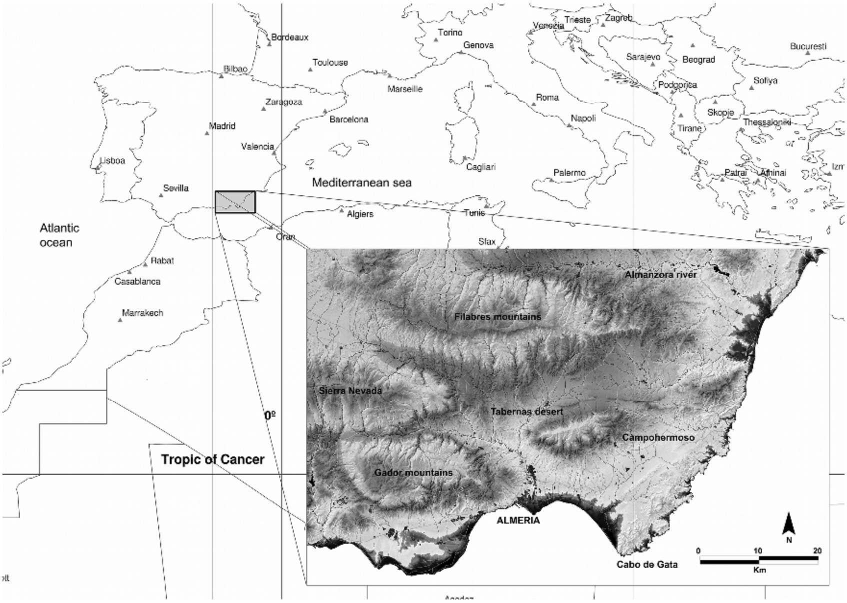

Fig. 11.1 Situation of the test area in the Mediterranean geographical context

11.2.2 Data sets

11.2.2.1 Environmental and geographical variables

From a digital elevation model of 20 m resolution (provided by the Environmental Information Network of the Andalusian Regional Government) a total of 11 topographical variables were derived: elevation, slope, northness, southness, eastness, westness (in gradient from 100 to 0), direct solar radiation (mean, minimum and maximum computed by the Solar Analyst extension for ArcView 3.2), topographical wetness index (TWI) and sediment transport index (STI) (both computed in ILWIS 3.4 Open using the Flow_indices script available at http://spatial-analyst.net).

From road maps from 1987, 2000 and 2006 (2006 map includes roads under construction expected for 2010), we mapped the distance to roads of 1st, 2nd, 3rd, and 4th order (motorways and national highways, regional roads, provincial roads and local roads, respectively) for years 1987, 2001 and 2010. For each year, an "accessibility index" coverage was built, computed by a weighted mean of the distance coverages. Weights were: 1 for 1st order roads, 0.75 for 2nd order roads, 0.50 for 3rd order roads and 0.25 for 4th order roads. The weighted sum was scaled into values from 0 to 100 using the module Stretch of the Idrisi Andes software.

Coverages of distances to water resources in years 1987 and 2001 were drawn using the cartography of water infrastructures of the regional government of Andalusia. Distances to water resources in 2010 were computed using the locations of future desalination plants projected by the Water Plan of the Ministry of the Environment of Spain. Areas not suitable for greenhouses (towns, lakes, natural parks) were masked in the datasets so as to be excluded from the analysis.

To avoid high correlation between variables in the dataset, we used Biomapper 3.0 (Hirzel et al. 2006), which computes UPGMA (Unweighted Pair-Group Meted with Arithmetic Mean) trees using Pearson's correlation index as the distance between variables. With 0.75 being selected as the maximum correlation threshold, from each group of highly correlated variables, one was retained. The remaining variables (elevation, slope, topographical-wetness index, mean solar radiation, accessibility index, distance to water resources, and the distances to roads of 1st, 2nd, 3rd and 4th order) were used to compose three data sets corresponding to the years 1987, 2001 and 2010, which had in common topographic variables but differed in the values of the distance variables (see Fig. 11.2).

11.2.2.2 Greenhouse coverages and presence records

For the calibration and evaluation of MaxEnt suitability maps and Geomod simulations, presence records and greenhouse coverages are needed. For calculating greenhouse coverages for 1987 and 2001, digital land use maps from the years 1991 and 1999 (stored as polygon layers) were manually corrected using Landsat images (RGB composites) of years 1987 and 2001 as reference. The resulting polygon layers were rasterized to determine greenhouse (absence/presence) Boolean coverages. From these coverages, 340 and 471 presence records of greenhouses were collected, respectively, by random sampling, establishing a minimum-distance criterion (at least 1000 m between records) in order to avoid spatial autocorrelation effects from using samples too close together (pseudoreplication).

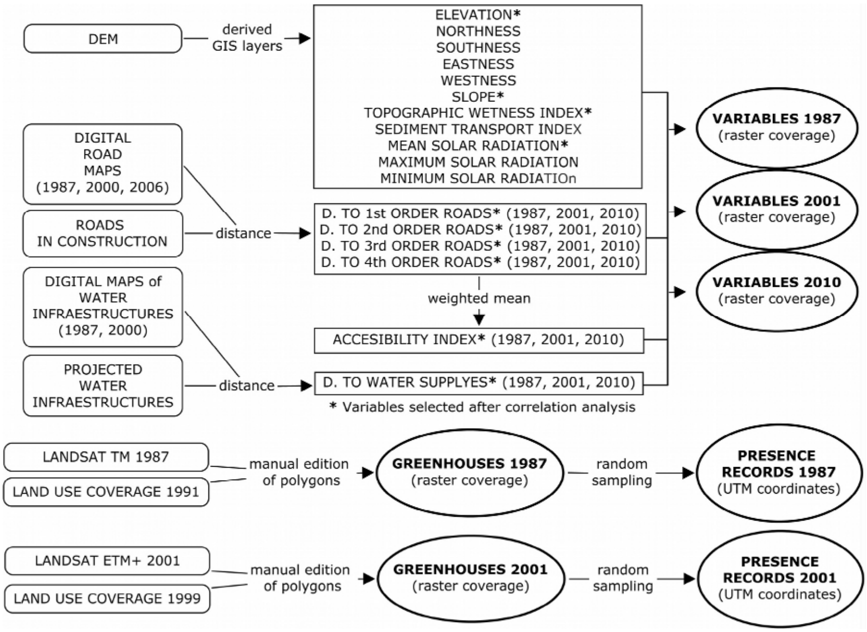

Fig. 11.2 Flowchart I. Making of the environmental datasets and presence samples

Although the initial resolution of both data sets (variables and greenhouse coverage) was 20 m, the process that we wished to model occurred at a lower resolution, related to the dimensions of the greenhouses (Hengl 2006). A study of the mean area of the greenhouses indicated that a pixel size of 80 x 80 m would properly represent the land use change, and therefore all the spatial data were rescaled to this resolution using the module Contract included in Idrisi Andes (pixel aggregation for continuous data and pixel thinning for categorical data).

11.2.3 Modelling and evaluation tools

11.2.3.1 MaxEnt

MaxEnt (Maximum Entropy), a general purpose method for making predictions from incomplete information, has recently been applied to modelling biological species distribution (Phillips et al. 2006). Successful tests have demonstrated that its results are among the best possible within the broad set of algorithms for distribution modelling (Elith et al. 2006).

The algorithm needs a sample of presence records of the organism and a set of environmental variables of the entire study area to compute the distribution model. The environmental variables and functions representing the interactions among them are called "features", from which the ecological niche of the species is defined. Using presences, features, and a background sample (locations taken randomly from the entire study area) MaxEnt searches iteratively for the probability distribution of the maximum entropy (the closest to uniform), but subject to one condition: the expected value for each feature under the estimated distribution matches its empirical average (calculated from the values of the feature in the presence records). The probability is computed in terms of "gain" (log of the number of grid cells minus the average of the negative log probabilities of the sample locations), which starts at zero and increases in each iteration, until differences between iterations fall below a predefined "convergence threshold", or the "maximum iterations" number is reached (Phillips et al. 2006).

The probability distribution is projected onto the geographical space, resulting in a distribution model with a range of values of between 0 and 100, which expresses in relative terms the suitability of the habitat for the species (suitability map). MaxEnt can also project the model over variables representing a different time, to explore simulated past or future scenarios. In order to provide a better understanding in the relationships between variables and presence records, MaxEnt performs a Jackknife test to measure variable importance, and plots the log response curves for each variable.

Greenhouses require a combination of environmental variables (temperature, solar radiation, etc.) and geographical ones (distance to roads, water resources) that influence its productivity. These requirements determine the greenhouse construction site selection in the same way as a biological species selects an appropriate habitat. This quasi-biological behaviour permits the application of MaxEnt to calculate the potential distribution of greenhouses, using the same method as applied to biological species. According to this idea, high suitability values in a greenhouse MaxEnt distribution model indicate areas adapted for the construction of greenhouses.

11.2.3.2 Geomod

Geomod (Pontius et al. 2001) is a land use change simulator implemented in Idrisi Andes (Clarklabs 2006). It simulates the land use change between two categories (e.g. from unoccupied to occupied by greenhouses) using as start-up information the beginning and ending time of the simulation, a coverage with the initial state of the two categories, the land area changing in use (indicated by the number of cells), a series of environmental variables from which a suitability map is drawn (determining the areas most prone to use change), and a stratification map (enabling the area to be divided into regions that behave differently).

The simulation is based on certain decision rules:

1. Land use change is simulated in only one direction, from occupied to unoccupied or vice versa, but not both simultaneously. If a stratification map is used, Geomod can simulate changes in different directions for different strata.
2. A neighbourhood rule should be defined: in the constrained mode, a radius is established for the edge of the initial use patches within which Geomod will search for the areas prone to change. In the unconstrained mode, it searches for transition areas without restrictions on the radius, throughout the entire territory being analysed.
3. The suitability map for land use change. Geomod computes a suitability map from a set of environmental variables (that influence land use change) and a coverage of the initial state of land use. The computing method reclassifies each variable into categories, assigning to every new category the value of the percentage of cells occupied by the land use towards which the change is going to be simulated. Finally, a weighted sum of the reclassified variables is used to compute the suitability map. The weighting factor may be equal for all the variables or defined for each one by the user. The values of the suitability map are called "lubrication values": larger lubrication values implies high suitability for land use change (for more details, see Pontius and Chen 2006).

11.2.3.3 Procruster analysis

Sensitivity (S) is the conditional probability that a presence cell in the reference image is predicted correctly in a simulation. It can be calculated from the confusion matrix provided by de Crosstab module of Idrisi Andes, dividing the true presences (correctly simulated cells) by the sum of true presences and false presences (incorrectly simulated presence cells). The result (the true positive fraction) is a measure of agreement between a simulation and a reference image in terms of quantity, without bearing spatial differences in mind. We use this additional evaluation measure to support the results of the Procrustes analysis.

11.3 Methodology and practical application to the datasets

Two simulation phases were executed using Geomod: to test the performance of MaxEnt suitability maps and select the decision rules that best represent the spreading of greenhouses, nine simulations using different combinations of decision rules were performed and evaluated for the interval 1987-2001 (1987 data to calibrate and 2001 data to evaluate simulations). Then, using the selected rules, Geomod simulations considering three different land use change scenarios were performed for the interval 2001-2010.

11.3.1 Simulations 1987-2001

The aim is to select the decision rules available to calibrate Geomod simulations that best describe the spreading of greenhouses in the study area. Suitability maps computed by MaxEnt and Geomod, and different neighbourhood rules were combined in nine performed simulations:

- Suitability maps: Three suitability maps were used: *1)* M1, (computed in MaxEnt) model calibrated with the training sample and variables of 1987. *2)* M2, (computed in MaxEnt) model calibrated with the training sample and variables of 1987 and projected over variables of 2001 (using the Projection feature available in the software). Suitability maps computed with MaxEnt were calibrated using the default settings (Phillips et al. 2006). *3)* G1, computed in Geomod with the greenhouse's coverage and variables of 1987 (using the same weighting factor for all variables).
- Neighbourhood: settings used were 80 m (1 cell around), 2,000 m (25 cells around) and unconstrained.

The simulations were calibrated setting the starting time at 1987, initial area of the greenhouses coverage of 1987 (38,743 cells, 24,795 ha.), ending time at 2001 and final area of the greenhouse coverage of 2001 (58,097 cells, 37,182 ha). All simulations were stratified by municipality limits, representing the diversity of land-management policies in different towns. An extra simulation (unstratified, without suitability map and unconstrained neighbourhood) was performed in order to simulate the random spreading of greenhouses, calling this the Random Simulation (hereafter, RS; to clarify this explanation, see Fig. 11.3).

11.3.1.1 Evaluation and spatial certainty of the simulations

It is important to consider that, on comparing a simulation with the reference image, both share the entire area occupied by greenhouses at the starting time (1987). Consequently, any comparison index that we apply will interpret an inflated degree of agreement between the simulation and the reference image. To avoid this inflation, we eliminated (in all the simulations, the RS, and the reference image) the area corresponding to greenhouses in 1987. Therefore, the evaluation took into account only the area of the new greenhouses.

Results were evaluated by Procrustes analysis and sensitivity using the greenhouse coverage of 2001 as the reference image. The simulation with the least m^2 and greatest S with respect to the reference image will determine the decision rules that best describe the spreading of greenhouses. Results were tested separately for Procrustes analysis and sensitivity by factorial ANOVA, establishing a “suitability map” (levels: G1, M1 and M2) and “neighbourhood” (levels: 80 m, 2,000 m, and unconstrained) as categorical predictors. The relationship between m2 and S were assessed by linear regression.

Usually, when evaluating a simulation by calculating its sensitivity, a homogeneous spatial distribution of certainty must be considered, assuming that all the simulated cells have the same likelihood of being correctly classified. In the real world, if greenhouses are constructed preferably in areas of high suitability (according to the suitability maps) because it favours greenhouse productivity, and Geomod selects as a priority these areas to simulate land use change to greenhouses construction, we can assume that the certainty of the simulation will vary according to the values of the suitability map. Following this reasoning, in the areas of greatest suitability, the probability of finding cells where the presence of new greenhouses has been correctly simulated is higher than in the areas of lower suitability. To test this idea, a joint analysis was made of the best simulation, its suitability map and the reference image (coverage of greenhouses in 2001), in order to: *1)* describe graphically the relationship between the suitability map and the total amount of hits (correctly simulated cells) and errors (incorrect simulated cells) in the simulation, plotting the number of hits and errors against suitability values; *2)* find a suitability-certainty function that relates each suitability value to a given probability for a cell to be correctly simulated, which is useful to compute a simulation certainty map. For this, the percentage of correctly simulated cells was plotted against suitability. The plot represents the specific behaviour of the best simulation, but we are looking for a more general function, capable of predicting approximately the behaviour of different simulations. With this aim, the data was smoothed by means of a moving average (using 25 as span size), and analysed by a polynomial-curve fitting using Octave.

11.3.1.2 Simulations 2001-2010

The spreading of greenhouses in the study area has been continuous from 1954, and today the construction of greenhouses is booming, due to the construction of new infrastructures oriented to increase water supply. But the greenhouses are involved in a dynamic market, and the profitability of the crops depends on multiple economic and social factors difficult to predict. Another emerging factor adding uncertainty in the last years is the competition with other Mediterranean countries with cheaper production. To confront this uncertainty we propose three simple scenarios of spreading of greenhouses for the period 2001-2010:

a) Linear greenhouse area growth with the trend identified for 1987-2001.
b) Accelerated growth (20%) over the linear trend due to increased demand.
c) Slowed growth (20%) under the trend due to increased competition from countries with cheaper production (e.g. Morocco and Algeria).

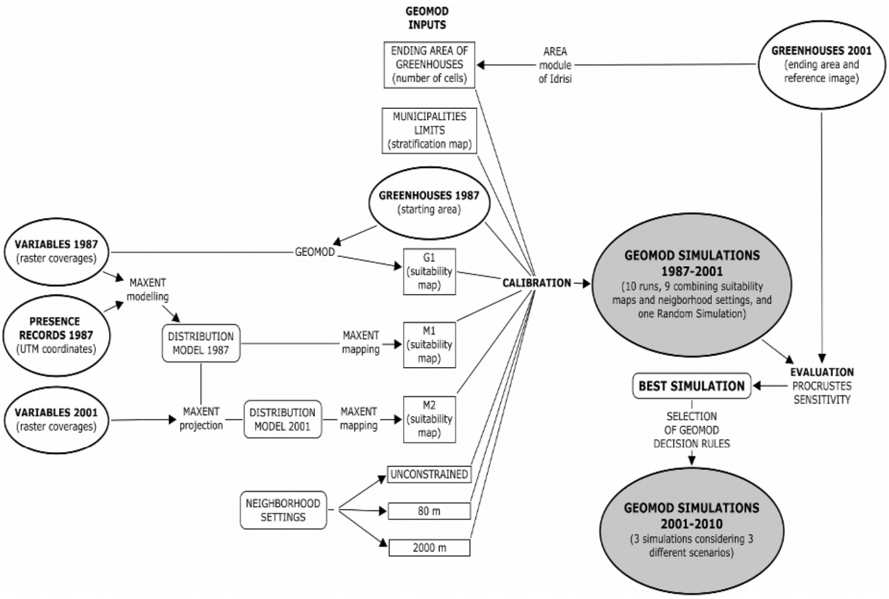

Fig. 11.3 Flowchart II. Steps followed to select the suitability map and the neighbourhood setting most appropriate to simulate the spreading of greenhouses

To be used a suitability map in the projections, a MaxEnt (M3) distribution model was calibrated using the training sample and variables of 2001, and projected over the variables of 2010. Differences in suitability between M2 and M3 were computed by map algebra. Using the 2001 greenhouse coverage as the starting image, the suitability map M3, and the projected areas in the different scenarios, three Geomod simulations for the period 2001-2010 were performed. To assess the spatial certainty of the simulations, the computed suitability-certainty function (see Sect. 11.3.1.1) was applied to the M3 suitability map (replacing M2 values by M3 values and dividing the result by 100 to translate percent values into probabilities).

11.4 Results

11.4.1 Simulations 1987-2001

Fig. 11.4 shows the coverage of greenhouses (1987) and the suitability maps used as inputs to calibrate Geomod simulations. Fig. 11.5 summarizes the influence of the modelling variables in the MaxEnt distribution model (suitability maps M1 and M2) and shows the log-response curves of the most relevant variables.

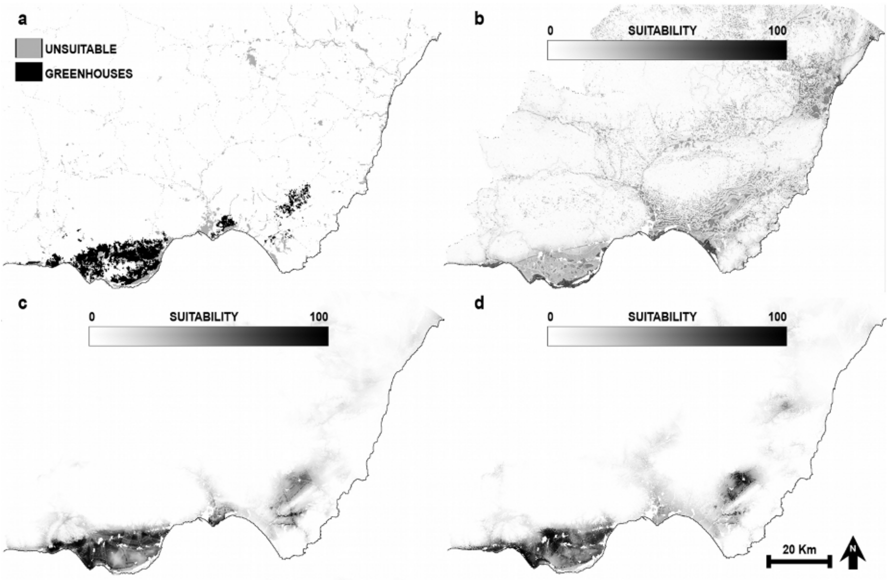

Fig. 11.4 Greenhouses in 1987 and suitability maps. a) Coverage of greenhouses in 1987 (starting time) used to calibrate simulations and non-suitable areas excluded from the analysis; b) G1 Suitability map, computed by Geomod; c) and d) M1 and M2 suitability maps calibrated in MaxEnt using presence records and variables of 1987 (M1), and projecting the model over variables of 2001 (M2). Dark colours indicate high suitability for the construction of greenhouses

The results of the Procrustes analysis and sensitivity for the nine performed simulations are shown in Fig. 11.6. All the simulations performed better than the RS (m^2=0.04; S=0.02), the results of which are not shown in Fig. 11.5 due to problems of scale. Two simulations calibrated with the suitability map M2 worked better than the remaining ones: the unconstrained neighbourhood simulation (m^2=0.0062; S=0.4060), selected as the best simulation and the 2,000 m neighbourhood simulation (m^2=0.0063; S=0.4050). Factorial ANOVA test found significant differences in simulations performance between suitability maps, but not between neighbourhood rules (see Table 11.1 for a summary of factorial ANOVA results). Procrustes and sensitivity values were closely correlated (adjusted $R^2=0.947$; p-level=0.000006). Unconstrained neighbourhood and MaxEnt suitability map (but replacing M2 with M3) were the settings selected to calibrate and simulate the three future scenarios of land use change.

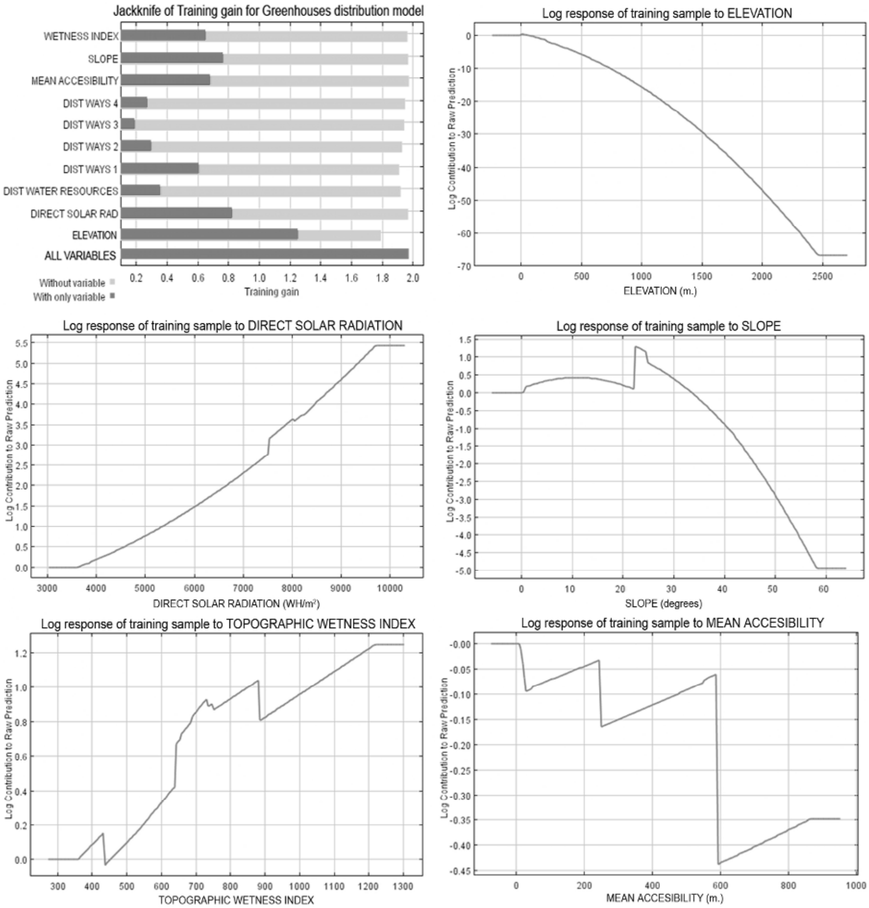

Fig. 11.5 Jackknife test and response curves of MaxEnt distribution model. Bars plot: dark grey bars indicate model gain when computed with only the variable, and light grey is the model gain when computed with the other variables. Minor differences between the two bars indicate major importance of the variable in the model. Log-response curves: the five most important variables are shown. Values over 0 indicate suitable conditions for greenhouses, whereas the values below zero indicate unfavourable conditions

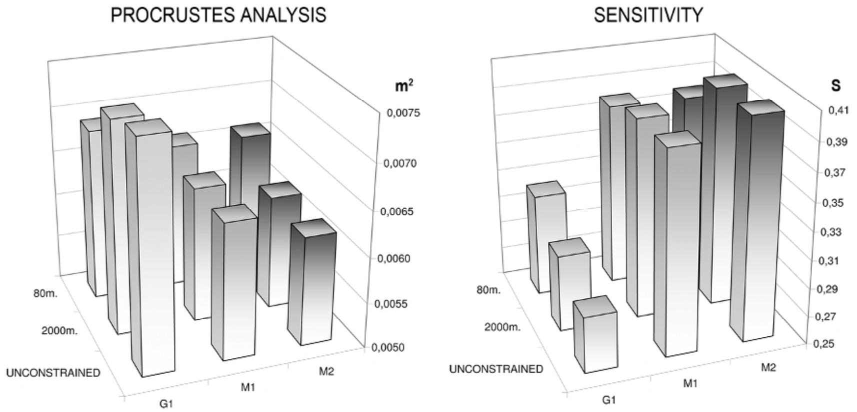

Fig. 11.6 Evaluation results. Evaluation of 9 simulations performed in Geomod combining three neighbourhood rules (80 m, 2,000 m and unconstrained) and three suitability maps, one computed in Geomod (G1), and two computed in MaxEnt (M1 and M2). Each bar corresponds to a performed simulation. The reference image is the real coverage of greenhouses in 2001 (ending time in the simulations). Low values of m^2 are indicative of a good spatial agreement between a simulation and the reference image. Higher values in S indicate a good agreement between a simulation and the reference image in quantity of correctly predicted cells. The best simulation was performed with an unconstrained neighbourhood and the suitability map M2.

Table 11.1 Summary of results of factorial ANOVA. Significant values in bold

Dependent variable	m^2		S	
Categorical predictor	suitability map	neighbourhood	suitability map	neighbourhood
F	**9.132**	0.009	**29.158**	0.108
P	**0.032**	0.990	**0.004**	0.900

Fig. 11.7 (left) shows the graphical analysis of correctly and incorrectly simulated cells of the best simulation. Fig. 11.7 (right) shows the polynomial relationship between the suitability values (of the M2 suitability map) and the percentage of cells correctly predicted for that suitability value, calculated from the best simulation. Eq. 11.1 expresses the suitability-certainty function (R^2=0.99; RMSE=1.24).

$$\% \text{ HITS} = 1.186\cdot10^{-13}\ M2^9 + -4.729\cdot10^{-11}\ M2^8 + 7.417\cdot10^{-9}\ M2^7 + -5.681\cdot10^{-7}\ M2^6 + 2.08\cdot10^{-5}\ M2^5 + -0.0002511\ M2^4 + -0.003183\ M2^3 + 0.1024\ M2^2 + -0.3118\ M2 + 0.3413 \quad (11.1)$$

M2: values for every pixel of the M2 suitability map.

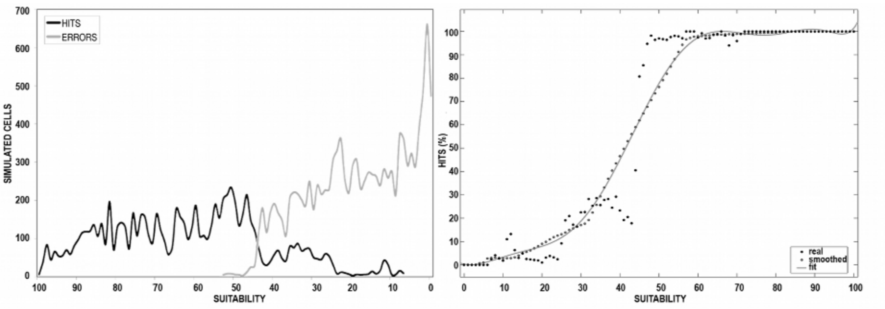

Fig. 11.7 Certainty against suitability in the best simulation. Left: the plots describe the behaviour of the best simulation in terms of total amount of correctly (hits) and incorrectly (errors) predicted cells against suitability. Right: 9th order polynomial relationship between the percentage of correctly predicted cells and suitability (M2 suitability map). The black plots represent the real data, and the grey dots the smoothed data. The curve represents the curve fitted to the smoothed data

11.4.2 Simulations 2001-2010

Fig. 11.8 shows the differences in suitability between M2 and M3. The construction of new infrastructures (roads and desalination plants) boosts the suitability for greenhouses in areas that already were fulfilling suitable topographic conditions. The relative importance of the variables and the response curves in the M3 distribution model were similar to those of the M2 distribution model, confirming that both models had close similarities.

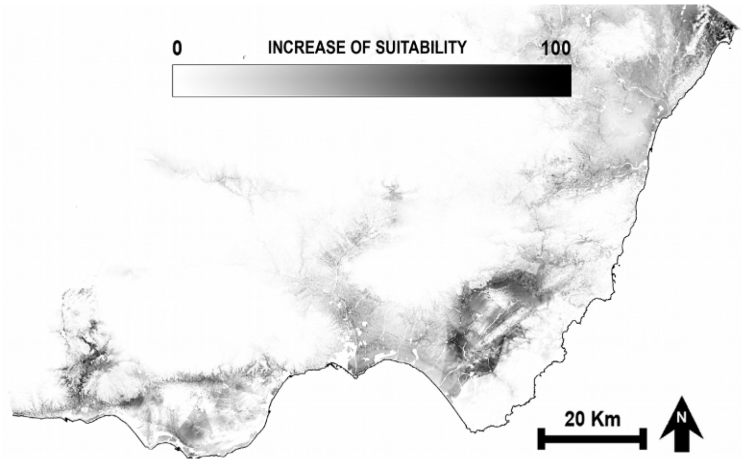

Fig. 11.8 Differences in suitability between M2 and M3. Differences in suitability between M2 and M3 were computed as M3-M2 in the raster calculator of Idrisi Andes. Higher values are indicative of new suitable areas for the spreading of greenhouses

Fig. 11.9 shows the simulations corresponding to the proposed scenarios A, B, and C, compared to the real greenhouse-occupied area in 2001.

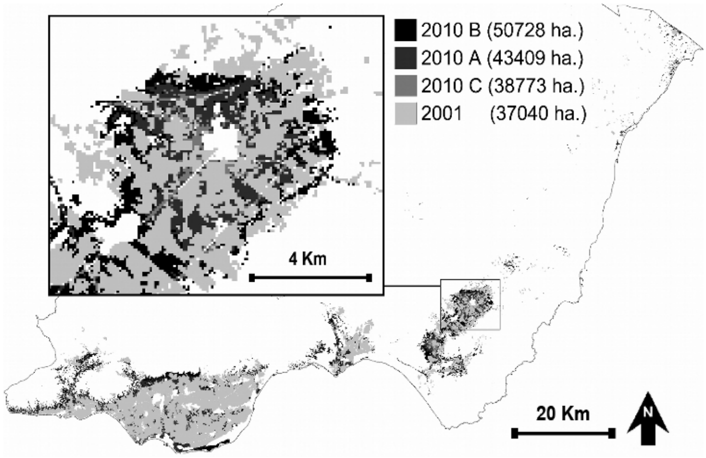

Fig. 11.9 Simulations of scenarios A, B, and C. Scenario B is the sum of 2010 B, 2010 A, 2010 C, and 2001 occupied areas. Scenario A is the sum of 2010 A, 2010 C, and 2001 occupied areas, etc. The zoomed area is a detail of Campohermoso (see Fig. 11.1), a locality with an intense growth of the area occupied by greenhouses in recent years

Fig. 11.10 shows the certainty map of the simulations of scenarios A, B, and C for the year 2010.

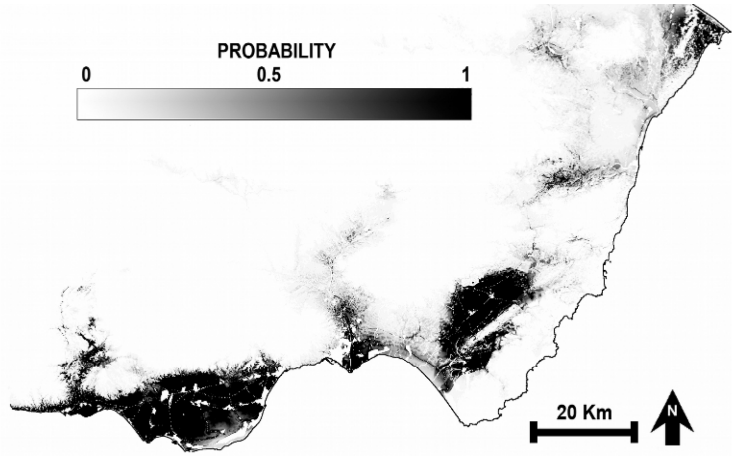

Fig. 11.10 Certainty map for simulations of scenarios A, B, and C. The certainty map computed from Eq. 11.1 applied over the M3 suitability map. The values indicate the probability that a cell with presence of greenhouses simulated by Geomod will really be occupied in 2010

11.5 Validation of the results and discussion

11.5.1 Simulations 1987-2001

The ANOVA analysis of the evaluation values indicate that the simulations performed with MaxEnt suitability maps work better than those performed with Geomod ones (in terms of S and m²). While MaxEnt uses scattered presence points as input, Geomod uses the complete land use coverage. If, for reasons other than the suitability of the territory, there is a great concentration of greenhouses in a given area (e.g. historical causes), the combination of variables present at this site takes on relatively high importance with respect to the other combinations of variables in the rest of the territory, resulting in a misleading suitability map. The same can happen in the MaxEnt models when the presence records provided as input were very close together, but the initial treatment that we applied (minimum distance between presence points greater than 1,000 m) diminish any possibility that aggregation effects could affect the quality of the models. Another difference between the two methods to calculate suitability maps is based on the relative weight given to the variables. Geomod does not use any algorithm to compute weights, and they have to be established by the user (by subjective criteria, or criteria based on previous statistical analysis). MaxEnt includes a Jackknife test, which automatically computes the relative contribution of every variable to the model. Another advantage of the MaxEnt algorithm successfully explored in this paper is the “Projection” feature, capable of projecting a distribution model over variables with values expected for the future. It is a useful tool to explore alternative land use change scenarios bearing in mind expected changes in the values of the variables (accordingly to known information, such as projected roads). Our results support the idea that the MaxEnt algorithm can generate useful suitability maps applicable to Geomod simulations, outperforming the results given by a stand-alone use of Geomod.

The results for the importance of the variables in the suitability maps computed by MaxEnt (Fig. 11.4) and of their response curves (Fig. 11.5) indicate that the fundamental factors influencing greenhouse distribution in the study area are related to topography and distance to roads. The open plains (which coincide with the areas having high indices of topographical moisture) at low altitudes have the temperature, slope, and solar radiation appropriate for greenhouse operation. The factors related to the distances to roads do not appear to be limiting, although the longest distance to first-order roads (motorways) are related to a lower presence of greenhouses. The variable “accessibility index”, the fourth in importance, accurately summarizes the distances to different types of roads, and it is useful to predict the spreading of greenhouses. Interpretation problems arise with the variable Distance to Water Resources, because the great majority of greenhouses do not depend on centralized resources such as reservoirs or desalination plants, but rather use their own wells, which pump water from aquifers. This variable is the only one that has lost relevance over time among the 1987-2001 and 2001-2010 models, since desalination of sea water has proliferated on the Almería coast. Even so, the low gain shown by all the models indicates that the contribution of desalination does not significantly affect the results.

In the evaluation of the simulations, we considered two sides: agreement in number of predicted cells, expressed in terms of S, and spatial agreement, tested by Procrustes analysis and expressed in terms of m^2. Both measures were correlated, but not perfectly because, for two simulations with the same sensitivity (compared with a reference image), there may be differences in the location of the errors detected by the Procrustes analysis. Procrustes analysis is a quick and simple way to assess spatial agreement between simulations and real land use coverages.

The analysis of hits and errors of the best simulation (Fig. 11.7, left) shows that it works better in the section of higher values of the M2 suitability map (especially in the range 100-50), and the errors increases when suitability trends toward zero. When the percentage of hits against suitability values is smoothed by a moving average (Fig. 11.7, right), the pattern remains quite clear, making it possible to find, by means of curve fitting, a function (9th-order polynomial, see Eq. 11.1) describing the behaviour of the simulation.

11.5.2 Simulations 2001-2010

The construction of new roads and desalination plants can increase the area suitable for greenhouses, as shown in the map of differences between M2 and M3 suitability maps (Fig. 11.8). Apart from the increase of suitable area, both models show identical behaviour regarding the relative contribution and the response curve that every variable presents. During the studied periods 1987-2001 and 2001-2010, the relationships between the presence of greenhouses and the variables that influence their distribution did not significantly change.

Geomod is designed to predict the locations of land use change, not the quantity of area that changed. Therefore, the validity of the simulations is based on a solid interpretation of the data for surface-area growth of greenhouses. Using only the two available sets of temporal data (1987 and 2001), we used a linear estimation to calculate the amount of occupied area in the scenarios A, B and C, but it would be more appropriate to use data from temporal series with a greater number of control points. The problem arises when inflexion points are foreseen in the growth curves of the occupied area, a possibility in the study areas because there is a high degree of saturation (a large area of land that can be occupied by the greenhouses is already occupied) and the resources supporting the greenhouse industry (mainly soil and water) are being depleted. Thus, to our knowledge, scenarios A and C are probably the closest to reality (see Fig. 11.9).

The certainty map (Fig. 11.10) can be useful to assess the expected accuracy of the simulations when it is not possible to validate them with ground-truth information. Nevertheless, the function used to compute the certainty map of the simulations 2001-2010 has been calculated for a simulation performed for the period 1987-2001 and the suitability map M2, there are at least two ideas that may justify its application:

- M2 and M3 distribution models are quite similar, and therefore a significant behavioural change in the suitability-certainty function between models is not expected.
- The smoothing of the data by moving the average prior to the curve fitting removes bias and generalizes the function, allowing its application to other simulations performed under the same conditions.

However, it is important to bear in mind the limitations of this application: the function does not take into account the effect of the area that will predictably undergo land use change. This effect is important because it tends to increase the percentage of correctly predicted cells of the simulation and can influence the relationship between certainty and suitability, altering the shape of the curve and changing the coefficients of the function. This can lead to an erroneous interpretation of the probability values of the certainty map. It would be useful to make an in-depth study of the relationship between suitability and certainty for different simulations to find an equation that can function in a general way in order to associate each simulating cell with a particular certainty value.

11.6 Conclusion and perspectives

11.6.1 Conclusions

The combined use of MaxEnt and Geomod provide a series of significant advantages with respect to the stand-alone use of Geomod in land use change simulations:

- Geomod simulations using MaxEnt distribution models as suitability maps significantly outperform simulations calibrated with suitability maps computed by Geomod. In addition, the "projection" feature of MaxEnt makes it possible to explore alternative scenarios by changing the values of the variables used to calibrate the model.

- It is possible to predict accurately the spreading of greenhouses using only topographic variables and distances to roads. In this sense, the proposed "accessibility index" is a useful variable that summarizes distances to different types of roads.
- Relationships between greenhouses and variables are stable in time for the periods studied, allowing the exploration of future scenarios.
- Procrustes analysis is a powerful tool to assess spatial similarity between simulations and ground-truth information, and provides a simple and easily interpretable measure of agreement (m^2).
- The certainty of Geomod simulations is not spatially uniform. There is a strong relationship between the amount of correctly simulated cells and suitability. This relationship is useful to compute certainty maps to assess the spatial accuracy of simulations.
- The proposed methodology can be applied to territorial management of areas in which greenhouse expansion can represent an environmental problem, as in the Mediterranean countries mentioned in the present work. From the simulations, it is possible to identify the hotspots on which to focus environmental management and conservation efforts.

11.6.2 Perspectives

In the context of global change, the studies on land use change are becoming as relevant as those related to climatic change. Though we have specifically oriented tools, the complexity of the web of factors affecting land use is such that it is difficult to develop highly accurate techniques. It is necessary to continue to delve into the possibilities offered by geographic information technology to formulate predictions that enable us to face coming changes.

The present study seeks to combine two different perspectives: ecological species-distribution models used in biology, and land use change models used in geography. Both approaches combined can generate powerful tools to analyse our changing world and to explore alternative scenarios.

Although the analysis proposed for land use change is applied only to greenhouses, it has other potential applications (perhaps irrigation, urbanization, tourist facilities, etc.). This implies another alternative to the different use of change models currently being used.

Acknowledgements

This research is supported by the Spanish Education and Science Ministry (REN2003-09427-C02 project) and the Andalusian Innovation, Science and Industry Regional Ministry (Consejería de Innovación, Ciencia y Tecnología de la Junta de Andalucía) (RNM 1067 project). We are indebted to David Nesbitt for the English translation. We are also grateful to Aitor Gastón and two anonymous referees, whose valuable commentaries improved this paper.

References

Aguilera F (2006) Predicción del crecimiento urbano mediante sistemas de información geográfica y modelos basados en Autómatas Celulares. Geofocus 6, pp 81-112

Artnzen JW (2006) From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles. Frontiers in Zoology 3, 8

Cobos JJ, López JC (1998) Filmes plásticos como material de cubierta de invernadero. Tecnología de invernaderos II, pp 143-160

Colasanti RL, Hunt R, Watrud L (2007) A simple cellular automaton model for high-level vegetation dynamics. Ecological Modelling 203, pp 363-374

Dixon RK, Brown S, Houghton RA, Solomon AM, Trexler MC, Wisniewski J (1994) Carbon pools and flux of global forest ecosystems. Science 263, pp 185-190

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J McC, Peterson AT, Phillips SJ, Richardson KS, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species distributions from occurrence data. Ecography 29, pp 129-151

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, pp 38-49

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecological Modelling 135, pp 147-186

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, pp 29-36

Hengl T (2006) Finding the right pixel size. Computer & Geosciences 32, pp 1283-1298

Hirzel AH, Hausser J, Perrin N (2006) Biomapper 3.2. Laboratory for Conservation Biology Department of Ecology and Evolution University of Lausanne, Switzerland http://www.unil.ch/biomapper

Jackson DA (1995) PROTEST: a procrustean randomization test of community environment concordance. Ecoscience 2, pp 297-303

Jantz CA, Goetz SJ, Shelley MK (2004) Using the SLEUTH urban growth model to simulate the impacts of future policy scenarios on urban land use in the Balitmore-Washington metropolitan area. Environmental Planning B: Planning and Design 31(2), pp 251-271

Johnson CJ, Gillingham MP (2005) An evaluation of mapped species distribution models used for conservation planning. Environmental Conservation 32 (2), pp 1-12

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, pp 231-259

Pontius Jr RG, Chen H (2006) GEOMOD Modeling. Idrisi Andes Help Contents, Massachusetts Clark University

Pontius Jr RG, Pacheco P (2004) Calibration and validation of a model of forest disturbance in the Western Ghats, India 1920-1990. Geojournal 61, pp 325-334

Pontius Jr RG, Cornell JD, Hall CAS (2001) Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. Agriculture Ecosystems & Environment 85, pp 191-203

Posillico M, Meriggi A, Pagnin E, Lovari S, Russo L (2004) A habitat model for brown bear conservation and land use planning in the central Apennines. Biological Conservation 118, pp 141-150

Takakura T, Fang W (2002) Climate Under Cover. Springer

Zaniewski AE, Lehmann A, Overton J McC (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157, pp 261-28