Supplementary data S1

Ground-based map data

Three data sources were used to calculate ground-based map models: (1) a bioclimatic atlas of Salamanca (León Llamazares, 1991); (2) the WorldClim series (Hijmans et al., 2005; http://www.worldclim.org/); and (3) a Digital Elevation Model (DEM) built with the contour lines of 1:50000 topographic maps by the Junta de Castilla y León (Spain). These three data sources were combined in two datasets: GrB-L, with variables from (León Llamazares, 1991) and the DEM; and GrB-W, with variables from Worldclim series and the DEM.

Altitude was obtained from a Digital Elevation Model (DEM) built with the contour lines of 1:50000 topographic maps by the Junta de Castilla y León (Spain). Slope and orientation (in radians) were derived with the Spatial Analyst extension of ArcInfo 9.2. These variables were aggregated from their original pixel size (50 m) to 1 km^2 using the Aggregate function of ArcInfo 9.2. The new aggregated pixel was calculated by mean of the original pixels. Only variables with a correlation lower than ±0.8 were included in the ground-based map models. Therefore, only Slope was included in models performed with Leon Llamazares's variables (that is, GrB-L dataset; Supplementary Table 1; León Llamazares, 1991); Altitude, Slope and Exposition were included in models performed with Worldclim variables (that is, GrB-W dataset; Supplementary Table 1; Hijmans et al., 2005).

A total of 11 climatic variables were collected from (León Llamazares, 1991; Supplementary Table 1). These were scanned into images, georeferenced using six points of known coordinates (two middle points and the four map corners), in the ArcInfo 9.2 software, and automatically vectorised with the ArcScan extension, as well as transformed in a raster file with a pixel size of 1 km^2 (see variable example in Supplementary Figure 1). The original value of the vector line was maintained in the pixels that intersected the vector line. The mean value of two vector lines were assigned to

the pixels located between that two vector lines (Supplementary Figure 1). Finally, the GrB-L dataset included 12 variables.

A total of 4 climatic variables were collected from Worldclim series (Supplementary Table 1). Worldclim variables (monthly precipitation and mean, minimum, and maximum temperature) are interpolated climate surfaces for global land areas (excluding Antarctica) at a spatial resolution of 30 arc s (1 km^2), gathered from a variety of sources and the 1950–2000 period. The monthly variables were aggregated in mean and minimum precipitation; maximum of maximum temperature; and coldest month temperature. Finally, the GrB-W included 7 variables.

Satellite imagery data

Altitude was obtained from the SRTM. Slope and orientation (in radians) were obtained with the same methodology used for the other DEM (Supplementary Table 2 and Supplementary Figure 2). A cloud-free and snow-free Landsat-5 TM scene (path 203, row 32) acquired on Jun 23, 1999 was selected, extracting a subset of 2552 rows and 3981 lines (upper left corner: 146502, 4564957; lower right corner: 243762, 4439317, UTM coordinates). The scene was corrected with the Fast Atmospheric Correction Algorithm (ATCOR) (Richter, 1991); georeferenced to the UTM coordinate system (Datum Europeum 1950 for Spain and Portugal) with the Thin Plate Spline algorithm using 102 control points from 1:50 000 topographic maps, located in recognizable human structures; and resampled to a 30 m spatial resolution with the nearest neighbour method, in order to maintain the original pixel values (Chuvieco, 2000). The root-mean square (RMS) error in the geometric registration was lower to one pixel. These transformations were made using the GCPWorks package of PCI software.

From the Landsat 5 TM image four variables were obtained (Supplementary Table 2, Supplementary Figure 2): Radiance, Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), and a Land cover map.

1) Radiance, was calculated, as a indirect measure of insolation, with the expression (Chuvieco, 2000):

$L_{sensor,k}=a_{0,k}+a_{1,k}\ ND_k$,

where $L_{sensor,k}$ is the radiance of a particular sensor in a k channel; a_0 and a_1 are specific coefficients for each spectral sensor channel (Supplementary Table 3); and ND_k the digital number of the pixel k. The total radiance (L_T) was the sum of all k radiances, except the thermal channel 6, excluded because it has a lower spatial resolution (120 m).

2) Land Surface Temperature was derived from the radiance of the thermal channel using the Fast Atmospheric Correction Algorithm (ATCOR; Richter, 1991).

3) Normalized Difference Vegetation Index (NDVI) was calculated from the radiance values of channels 3 and 4 (Tucker, 1979):

NDVI=(IRC-R)/(IRC+R),

where IRC is the near infra-red channel 4 and R is the red channel 3.

4) To make the Land cover map, the study area was visited several times to map the vegetation, drawing the vegetation polygons on two prints of a RGB composition with the 4-3-5 channels of the Landsat 5 TM sensor. An average of three training fields for a minimum of 500 pixel extension were selected for each land cover type: water, pine, holm-oak, ‘dehesa’ (an open forest exploitation with agriculture and cattle use), grown vegetation, shrub, meadow, grown agriculture fields, ploughed agriculture fields, abandoned agriculture fields, and ground. Principal Component Analysis (PCA) was performed in order to reduce the number of variables into few uncorrelated vectors (Sokal and Rohlf, 1995), selecting the first two principal components that explained the 98.9% of variance. A Maximum Likelihood supervised classification was produced with the two PCA components and the NDVI (Chuvieco, 2000). The accuracy assessment gave a value of 70% . Each vegetation type was exported to a new channel, where pixels of a given vegetation type were reclassified as 1 and remaining pixels as 0. Three variables related with agriculture fields were

combined in a unique one, called agriculture fields (Supplementary Table 2).

The 14 satellite imagery variables were aggregated from their original pixel size (30 m) to 1 km^2 using the Aggregate function of ArcInfo 9.2. The new aggregated pixel was calculated by mean of the original pixels. Therefore, each land cover class changed from categorical variable (i.e. the presence/absence of a particular land cover class) to continuous variable (i.e. density of this class in a particular pixel).

Modelling techniques

Bioclim (Nix, 1986) uses mean and standard deviation for each environmental variable separately (assuming normal distribution) to calculate bioclimatic envelope associated to the occurrence points. Each variable has its own envelope represented by the interval [m - c*s, m + c*s], where 'm' is the mean; 'c' is the cut-off input parameter; and 's' is the standard deviation. Besides the envelope, each environmental variable has additional upper and lower limits taken from the maximum and minimum values related to the set of occurrence points. In this model, any point can be classified as: *Suitable*, if all associated environmental values fall within the calculated envelopes; *Marginal*, if one or more associated environmental value falls outside the calculated envelope, but still within the upper and lower limits; and *Unsuitable*, if one or more associated environmental value falls outside the upper and lower limits. Bioclim's categorical output is mapped to probabilities of 1.0, 0.5 and 0.0 respectively.

The Mahalanobis distance (Sokal and Rohlf, 1995) is the maximum distance to the reference in the environmental space, above which the conditions will be considered unsuitable for presence. Since 1 corresponds to the biggest possible distance between any two points in the environment space, setting the maximum distance to this value means that all points in the environmental space will have an associated probability. The probability of presence for points that fall within the range of the maximum distance is inversely proportional to the distance to the reference point (linear decay).

GARP (Stockwell and Noble, 1992) is a genetic algorithm that creates ecological niche models for species. The models describe environmental conditions under which the species should be able to maintain populations. For input, GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive. Like Maxent (see above), GARP extracts randomly background data, i.e. data from all the geographic surface of the study area (Stockwell and Peters, 1999), including pixels with and without species records.

Maxent (Phillips et al., 2004, 2006) or Maximum Entropy model is a general-purpose machine learning method, which is particularly well suited to noisy or sparse information and capable of dealing with continuous and categorical variables at the same time. Essentially, Maxent chooses the model with the maximum entropy, i.e. the one that produces the most uniform distribution but still infers as accurately as possible the observed data (e.g. maximize entropy for a given chi-squared value). Maxent estimates the range of a species with the constraint that the expected value of each variables (or its transform and/or interactions) should match its empirical average, i.e. the average value for a set of sample points taken from the species-target distribution. Maxent randomly selects uniformly distributed data from the background squares, i.e. including either pixels with or without species presence (in fact, all presences are included in the data extracted from the background). It uses until 10 000 background points in an iterative way. In each iteration, it learns and improves the model. Maxent stops normally after 500 iterations, when the maximum entropy distribution is reached. Therefore, Maxent cannot be considered a pseudoabsence method, but a background method (Phillips et al., 2009). Maxent was run randomly selecting 75% of the presence records as training data and 25% as test data. Hence, the arithmetic average and the standard deviation of a set of 10 models was calculated through an iterative process (Araújo and New, 2007; Martínez-Freiría et al., 2008; Phillips and Dudik, 2008).

Bioclim, Mahalanobis distance and GARP were performed using openModeller software (Sutton et

al., 2007; www.openmodeller.org); Maxent models were developed with Maxent 3.2.1 software (http://www.cs.princeton.edu/~schapire/maxent). All models calculated the species' realized niche (sensu Sillero, 2011).

References

Araujo, M.B., New, M. (2007): Ensemble forecasting of species distributions. TREE **22**: 42-47.

Chuvieco, E. (2000): Fundamentos de Teledetección Espacial, 3ª ed. revisada. Ed. Rialp, Madrid.

Hijmans, R.J., Cameron, E., Parra, J.L., Jones, P.G., Jarvis, A. (2005): Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**: 1965-1978.

León Llamazares, A. (1991): Caracterización agrícola de la provincia de Salamanca, 2ª Ed. Ministerio de Agricultura, Pesca y Alimentación, Madrid.

Martinez-Freiria, F., Sillero, N., Lizana, M., Brito, J.C. (2008): GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. Diver. Distr. **14**: 452-461.

Nix, H.A. (1986): A biogeographic analysis of Australian Elapid Snakes. In:R. Longmore (ed.), Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series Number 7. Australian Government Publishing Service, Canberra, pp. 4-15.

Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006): Maximum entropy modeling of species geographic distributions. Ecol. Mod. **190**: 231-259.

Phillips, S.J., Dudik, M. (2008): Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography **31**: 161-175 .

Phillips, S.J., Dudík, M., Schapire, R.E. (2004): A maximum entropy approach to species distribution modeling. Proc. 21 Int. Conf. Mach. Learning, 655-662.

Phillips, S. J.; Dudík, M.; Elith, J.; Graham, C. H.; Lehmann, A.; Leathwick, J., and Ferrier, S. (2009): Sample selection bias and presence-only distribution models: implications for

background and pseudo-absence data. Ecol. Appl. **19**: 181-197.

Richter, R. (1991): A fast atmospheric correction algorithm applied to Landsat TM images. IJRS **11**: 159-166.

Sillero, N. (2011): What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. Ecol. Mod. **222**: 1343-1346.

Sokal, R.R., Rohlf, F.J. (1995): Biometry, 3rd Ed. Freeman, New York.

Stockwell, D.R.B., Noble, I.R. (1992): Induction of sets of rules from animal distribution data: A robust and informative method of data analysis. Math. Comp. Simul. **33**: 385-390.

Stockwell, D. and Peters, D. (1999): The GARP modelling system: problems and solutions to automated spatial prediction. IJGIS **13**: 143-158.

Sutton, T., de Giovanni, R., Ferreira de Siqueira, M. (2007): Introducing openModeller - A fundamental niche modelling framework. OSGeo Journal **1**.

Tucker, C.J. (1979): Red and photographic infrared linear combinations for monitoring vegetation. RSE **8**: 127-150.

Supplementary Table 1: Description, origin, datasets, and units of the ecogeographical variables from ground-based maps. Three data sources were combined in two datasets: GrB-L, with variables from León-Llamanzares (1991) and the DEM; and GrB-W, with variables from Worldclim series and the DEM. All variables were aggregated to a spatial resolution of 1 km^2. Ad.= Adimensional.

Variables	Origin	Datasets	Units
Variability of deficit lower than 50 mm in Jun	León Llamazares, 1991	GrB-L	%
Variability of deficit lower than 50 mm in October	León Llamazares, 1991	GrB-L	%
Annual duration of the hot period	León Llamazares, 1991	GrB-L	nº days
Annual duration of the cold period	León Llamazares, 1991	GrB-L	nº days
Annual Duration of the dry period	León Llamazares, 1991	GrB-L	nº days
Annual potential evapotranspiration	León Llamazares, 1991	GrB-L	mm
Annual mean temperature	León Llamazares, 1991	GrB-L	ºC
Winter mean precipitation	León Llamazares, 1991	GrB-L	mm
Autumn mean precipitation	León Llamazares, 1991	GrB-L	mm
Summer mean precipitation	León Llamazares, 1991	GrB-L	mm
Nº days of October included in the cold period	León Llamazares, 1991	GrB-L	nº days
Annual mean precipitation	Worldclim	GrB-W	mm
Annual minimum precipitation	Worldclim	GrB-W	mm
Annual maximum of maximum temperature	Worldclim	GrB-W	ºC

Coldest month temperature	Worldclim	GrB-W	ºC
Altitude a.s.l.	DEM Topographic maps	GrB-W	m
Orientation	DEM Topographic maps	GrB-L/GrB-W	Radians
Slope	DEM Topographic maps	GrB-W	ºC

Supplementary Table 2: Description, origin, units and spatial resolution of the ecogeographical variables collected from satellite imagery.

Variables	Origin	Units	Spatial resolution
Radiance	Landsat 5 TM	Wm-2 rad-1	30 m
Land surface temperature from Thermal channel 6	Landsat 5 TM	ºC	30 m
NDVI	Landsat 5 TM	Ad.	30 m
Water from supervised classification	Landsat 5 TM	Ad.	30 m
Agriculture fields from supervised classification	Landsat 5 TM	Ad.	30 m
Dehesa from supervised classification	Landsat 5 TM	Ad.	30 m
Holm-oak from supervised classification	Landsat 5 TM	Ad.	30 m
Shrub from supervised classification	Landsat 5 TM	Ad.	30 m
Pines from supervised classification	Landsat 5 TM	Ad.	30 m
Grassland from supervised classification	Landsat 5 TM	Ad.	30 m
Grown vegetation from supervised classification	Landsat 5 TM	Ad.	30 m
Altitude a.s.l.	SRTM Radar DEM	m	100 m
Orientation	SRTM Radar DEM	Radians	100 m
Slope	SRTM Radar DEM	%	100 m

All variables were aggregated to a pixel size of 1 km^2. Ad.= Adimensional.

Supplementary Table 3: Values of the coefficient a0 and a1 for the calculation of the each radiance channel (Lk).

Channel	a0,k	a1,k
1	-1.5	0.602
2	-2.8	1.17
3	-1.2	0.806
4	-1.5	0.815
5	-0.37	0.108
6	0.124	0.00563
7	-0.15	0.057

Supplementary Figure 1: Examples of two climatic variables obtained from ground-based maps. Annual mean temperature, in degrees Celsius (A and B), and annual mean precipitation, in mm (C and D). Variables were initially vectorized from the ground-based map scanned images (a and c) and afterwards the corresponding vector lines were rasterized in the Landsat scene for the study area (B and D). The pixels that intersected the vector lines maintained the original value, and the pixels located between two vector lines had its mean value assigned. See the Supplementary Table 1 for details about the variables.

Supplementary Figure 2: Six examples of variables obtained from satellite imagery. (A) NDVI; (B) radiance; (C) land surface temperature; (D) holm-oak map; (E) grown vegetation map, from Landsat 5 Thematic Mapper (TM); and (F) altitude, from the Shuttle Radar Topographic Mission (SRTM). See Supplementary Table 2 for details about the variables.

Figure S1

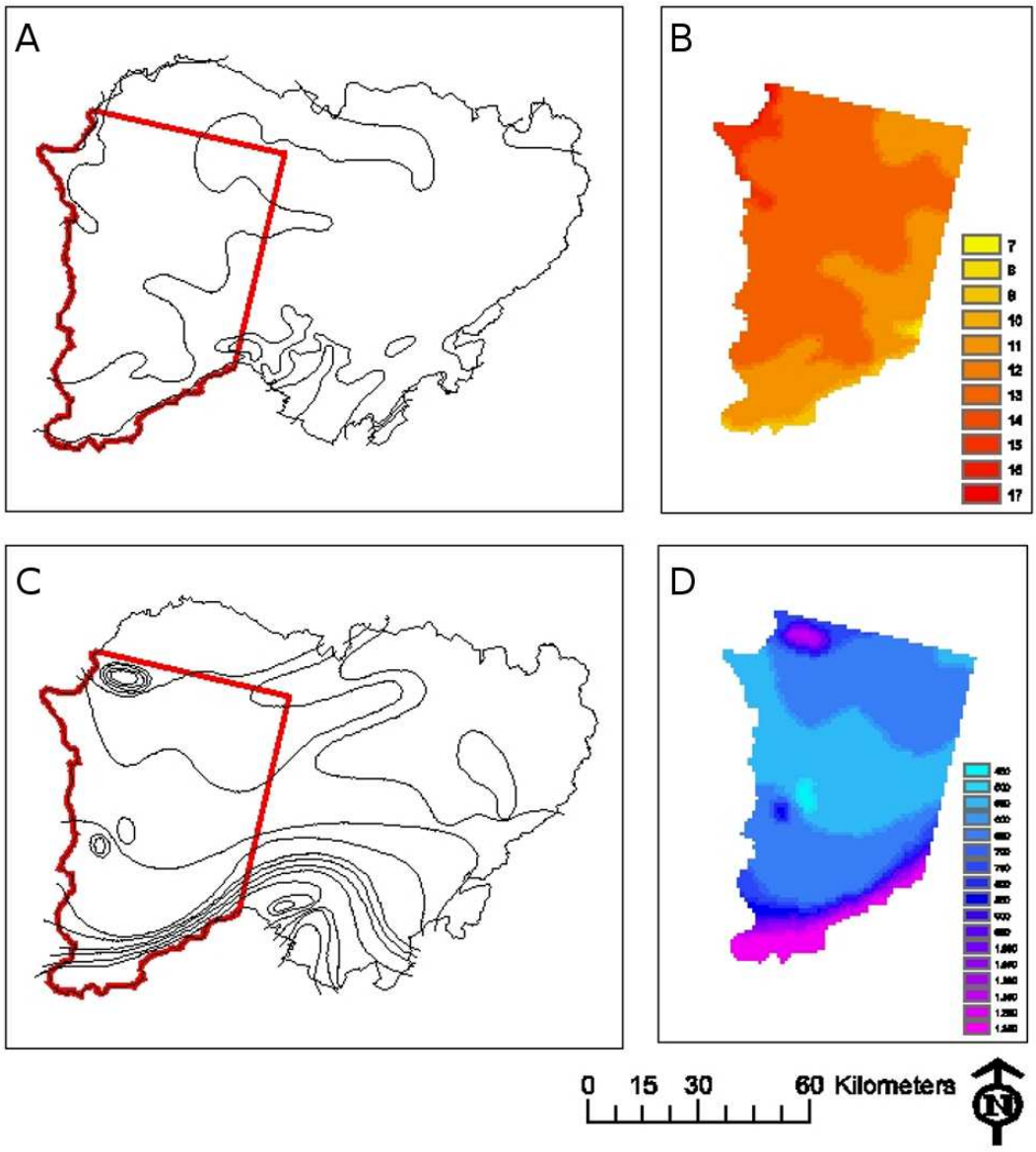
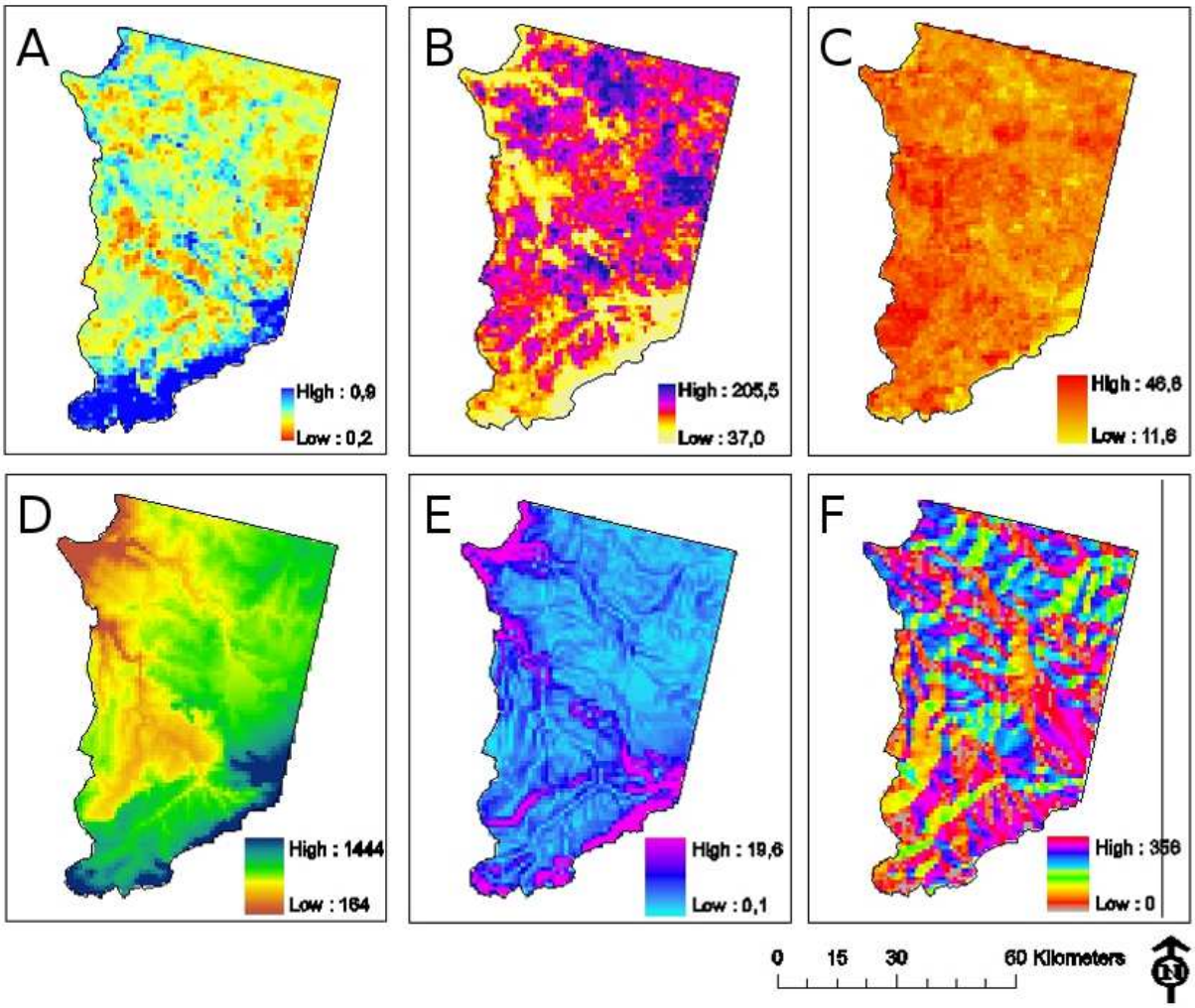


Figure S2


A
High : 0,9
Low : 0,2
B
High : 205,5
Low : 37,0
C
High : 46,6
Low : 11,6
D
High : 1444
Low : 164
E
High : 19,6
Low : 0,1
F
High : 356
Low : 0
0 15 30 60 Kilometers
N