# Distribution Determinants of Endangered Iberian Spider *Macrothele calpeiana* (Araneae, Hexathelidae)

A. JIMÉNEZ-VALVERDE[1] AND J. M. LOBO

Departamento Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

**ABSTRACT**   Little is known about the habitat preferences of *Macrothele calpeiana* (Walckenaer, 1805), an endangered endemic Iberian spider. In this paper, we seek to identify its distribution determinants and to disentangle their independent from combined effects. Generalized linear models (GLMs) of species presence–absence in southern Iberia were built from available distribution information and a variety of climate, land-use, and vegetation-vigor explanatory variables. Their independent and combined effects were estimated using variation and hierachical partitioning. On the scale of this work, *M. calpeiana* distribution is determined mainly by climate variables, especially by those related with precipitation; high annual precipitation and high precipitation periodicity favors the spider. Temperature is also important, because the species is not found where temperatures reach extremes. Whereas independent vegetative vigor and land-use effects, not easily separated from climate effects, are negligible, loss of forest to agriculture seems to have a negative effect. The slight positive effect of artificial surfaces cannot be directly attributed to anthropogenized-habitat preferences. Failure of climate model interpolation to predict some core species distribution areas in southern Iberia is discussed. The need for reliable distribution information from which to develop accurate habitat models is highlighted.

**KEY WORDS**   climate, habitat models, Iberian Peninsula, land use, *Macrothele calpeiana*

*Macrothele calpeiana* (Walckenaer, 1805), an endemic Iberian spider included in the Bern and Habitat directives, is distributed solely in southern Spain. Practically all populations have been found in the Guadalquivir river basin (Fig. 1), with the exception of a North African (Ceuta) record, considered to be the result of recent introduction by Spain–Morocco maritime traffic (Ferrández and Fernández de Céspedes 1996).

*Macrothele calpeiana*, a nonvagile, long-lived spider (females can live longer than 5 yr; Perry 2002), spins an aerial sheet-web that continues in underground silk tubes (funnel–webs), usually under stones or roots, but also in holes and crevices in bare ground, and even under tree bark several meters above ground (Gallon 1994, Santos Lobatón 1996). Mating seems to occur mainly in spring (May–March); spiderlings emerge in summer (August), probably remaining in the maternal retreat until October (Snazell and Allison 1989, Perry 2002).

*Macrothele calpeiana* populations, found mainly in cork oak (*Quercus suber*) forests (Snazell 1986, Snazell and Allison 1989, van Helsdinge and Decae 1992), where winters are warm, summer temperatures are high, and rainfall is copious, find a variety of habitats suitable (e.g., scrub land, pine forests, eucalyptus plantations; van Helsdinge and Decae 1992). Supplying some nonsystematic density information, van Helsdinge and Decae (1992) speculated that *M. calpeiana* was favored by a moderate amount of anthropogenic activity and did not consider the species to be an indicator of cork oak forests. However, studies have not been designed to provide conclusive information on *M. calpeiana* habitat preferences, so that its habitat requirements remain practically unknown.

In this paper, we explored, at a resolution of 1 by 1 km, the determinants of *M. calpeiana* distribution in the southern Iberian Peninsula, its main distribution area. Taking the place of experimentation, impossible on large spatial scales, habitat modeling (Guisan and Zimmermann 2000, Guisan and Thuiller 2005), the technique used in this study, is the most effective for testing species habitat preferences while identifying major predictor variables most reliably. Because explanatory variable correlation is an obstacle to the determination of probable causal factors, effects of climate, land-use, and vegetation-vigor groups were studied by variation partitioning (Legendre and Legendre 1998), whereas independent effects of single variables were studied by hierarchical partitioning (Chevan and Sutherland 1991).

[1] Corresponding author, e-mail: mcnaj651@mncn.csic.es.

**Fig. 1.** Records of *M. calpeiana* (Walckenaer, 1805) in the Iberian Peninsula referred to 1 by 1-km UTM squares. Spanish provinces cited in the text: A, Alicante; AL, Almería; BA, Badajoz; C, Cáceres; CA, Cádiz; CO, Córdoba; GR, Granada; HU, Huelva; JA, Jaén; MA, Málaga; SE, Sevilla; VA, Valencia. Window showing the 10 by 10-km presence points available (92).

## Materials and Methods

**Biological Data and Extent of Study.** *Macrothele calpeiana* presence data in 1-km$^2$ universal transversal mercator system (UTM) squares were extracted from the literature (Blasco and Ferrández 1986, Snazell and Allison 1989, van Helsdinge and Decae 1992, Santos Lobatón 1996, Luque 2001); 89 presence points were available for the species in the southern Iberian Peninsula (maximum latitude 39°39′; Fig. 1). It is necessary to note that some populations only had distribution data at a resolution of 10 km$^2$ and no 1-km$^2$ UTM square was available (Fig. 1, inset).

**Environmental Data.** At a resolution of 1 by 1 km in southern Spain, eight climate variables were considered: yearly days of frost; insolation (annual hours of sunlight); annual precipitation, precipitation periodicity (the coefficient of variation of monthly scores); mean annual temperature; minimum winter temperature; maximum summer temperature; and annual temperature range. All these variables were courtesy of the Spanish Instituto Nacional de Meteorología (INM, Ministerio de Medio Ambiente, Madrid, Spain), whereas mean altitude, slope, and aspect were obtained from a Digital Elevation model, at a resolution of 3 arc-s (≈90 m), provided by the United States Geological Survey (2005). The mean normalized difference vegetation index (NDVI, the photosynthetically active radiation that plants absorb, is a measure of plant density and vegetation health; Chong et al. 1993, Pettorelli et al. 2005) for 2001 was provided by the Instituto Nacional de Técnica Aeroespacial (CREPAD, Gran Canaria, Spain); to minimize cloud and/or aerosol noise, mean annual NDVI was calculated from maximum monthly values. For each 1 by 1-km UTM square, the percentage of: forest, agricultural land, scrub, and grassland, open places with little or no vegetation, and artificial surfaces were extracted from the Corine Land Cover 2000 (100-m resolution; http://terrestrial.eionet.europa.eu/CLC2000)

**Statistical Analyses.** Tabulated maximum and minimum values of the 17 above-mentioned variables at

presence points defined the multidimensional envelope for *M. calpeiana* (Busby 1991, Lobo et al. 2006). From the area outside the envelope, 801 pseudo-absences were randomly selected (prevalence = 0.1). As absences from a biological atlas are not necessarily true absences (the species may be present in a particular cell but not recorded), their inclusion as erroneous data reduces model prediction power; their substitution by pseudo-absences limits the amount of noise in the data. Moreover, these pseudo-absences may be used with prediction techniques using both presence and absence data to enhance prediction accuracy (Zaniewski et al. 2002, Engler et al. 2004, Lobo et al. 2006).

Generalized additive models (GAMs) with penalized regression splines (Wood 2000, Wood and Augustin 2002) were used to explore spider presence–absence relationships with predictors. We used the mgcv package (Wood 2004) to fit GAMs with four initial degrees of freedom in R (R Development Core Team 2004). To reduce the effects of multi-collinearity, predictors were first classified in intracorrelated groups by means of an $r \geq 0.8$ classifier threshold ($r$, Pearson's correlation coefficient; Silva and Barroso 2004). The members of the group explaining less GAM deviance or with a complex or unrealistic relationship with presence-absence data were dropped.

Occurrence of *M. calpeiana* was finally modeled using logistic regression analysis (generalized linear models [GLMs] with binomial distribution and logit-link function; McCullagh and Nelder 1997). Models were backward-stepwise fitted (Harrell 2001), producing nested models to be Akaike information criterion (AIC)-tested (Buckland et al. 1997), a penalization of the log-likelihood of the model as function of the number of degrees of freedom. GLMs were fitted in R (R Development Core Team 2004). Probabilities produced by logistic regression, unavoidably biased toward the most common event (Cramer 1999), cannot be considered indicative of habitat suitability, and they must be rescaled (Jiménez-Valverde and Lobo 2006). For this purpose, we used the favorability function proposed by Real et al. (2006) to eliminate the random element from the logistic equation:

$$F = 1 - \frac{1}{1 + e^{\left(\ln \frac{P}{1-P} - \ln \frac{n_1}{n_0}\right)}}$$

where F is the favorability value, $P$ is the logistic probability, $n_1$ is the number of presences and $n_0$ is the number of absences. The favorability probabilities so derived were mapped. Residuals of the logistic functions were examined and tested for autocorrelation using the Moran's I spatial autocorrelation statistic (Sawada 1999), selecting a lag distance of 12 km. Moran's I test was checked for significance with the Bonferroni-corrected significance level. Spatial autocorrelation in the residuals usually indicates that the model must be enlarged to incorporate spatially structured variables not otherwise accounted for (Odland 1988); addition of complex spatial terms (the third-

degree polynomial of latitude and longitude) to the model can be expected to account for those ignored variables.

Models were "leave-out-one," jackknife validated; i.e., one observation was excluded, the model parametrized again with the remaining $n − 1$ observations, a predicted probability obtained for the excluded observation, and the procedure repeated $n$ times (Olden et al. 2002). With these new jackknife probabilities, the area under the receiver operating characteristic (ROC) curve (AUC), a measure of overall discriminatory power (Swets 1988; Fielding and Bell 1997), was computed. Also, sensitivity (presences correctly predicted) and specificity (absences correctly predicted) were calculated using the threshold which minimizes their difference (Jiménez-Valverde and Lobo 2006). All validation computations were run in R (R Development Core Team 2004).

Although previously corrected for multi-collinearity, explanatory variables remain unavoidably correlated. In the assessment of the relative influence of each group of climate, land-use, and NDVI explanatory variables on *M. calpeiana* presence, variation partitioning (Legendre and Legendre 1998) determined the independent effects of (1) climate variation alone; (2) land use alone; (3) NDVI alone; and the combined effects of (4) climate and land-use components; (5) climate and NDVI components; (6) land use and NDVI components; and (7) the three components. Explained deviance (D):

$$D = \frac{ND − RS}{ND}$$

where ND is the null deviance and RS is the residual deviance, can be used to compare models from different combinations of factors (Guisan and Zimmermann 2000). Total deviance is obtained by regressing the dependent variable against the three groups of factors. Percentage of explained deviance is also computed for pairs of variables and for each variable alone. The independent effect of each group of variables is obtained by subtracting variation explained by the combination of the other groups, from variation explained by the combination of all explanatory variables together. Variation attributable to the combined effect of pairs of groups may be obtained by simple sums and subtractions (Legendre and Legendre 1998, Muñoz et al. 2005).

A hierarchical partitioning procedure was also applied to the more relevant variables (Chevan and Sutherland 1991, MacNally 1996, 2000). This method aims to measuring the explanatory capacity of individual variables, considering all possible models ($2^k$, where k is the number of variables considered) in a hierarchy and computing the additional explained deviance by adding any one variable to a simpler model that does not include that variable. Mean additional explained deviance per variable, $I_A$ (a denotes the given variable), is considered the explanatory power of each variable independently. The five climate variables and four land-use variables most significantly

related with *M. calpeiana* presence/absence were hierarchical partitioned in R (R Development Core Team 2004). The hier.part package (Walsh and MacNally 2003), restricted to factors linearly related with the dependent variable, was run for land-use variables, whereas climate factors, some of them related by quadratic functions, were manually hierarchical partitioned. Thus, the statistical significance of only the land-use variable $I_A$ could be calculated with 1,000 randomizations of the matrix, followed by recomputation of the distribution of $I_A$ (MacNally 2002), whereas significance of climate factors was manually impossible to calculate. Z ($I_{observed} − \hat{I}_{randomized}/SD_{Irandomized}$) is calculated and the statistical significance is based on the upper 95% of the standard normal distribution ($Z ≥ 1.65$).

## Results

*Macrothele calpeiana* relationships with all the variables considered were statistically significant, except those with aspect and percentage of places without vegetation cover (Figs. 2 and 3). Annual precipitation, precipitation periodicity, and annual temperature range are most explanatory, explaining 58.9, 58.4, and 49.1% of deviance, respectively; the first two positively, linearly related with spider presence-absence and the third negatively related. NDVI, maximum summer temperature, and minimum winter temperature are the three variables next in importance, with 35.3, 31.7, and 31.5% of explained deviance, respectively. From a midpoint, *M. calpeiana* is positively, linearly related with NDVI. The relation with minimum winter temperature is positively linear until a threshold point (1–2°C), where the slope of the curve approaches zero. The relation with maximum summer temperature is bell-shaped, with the maximum at ≈28°C. Slope and annual days of frost explain 17 and 16.1% of deviance, respectively. *M calpeiana* seems to avoid flat terrain and prefers areas with the smallest number of annual days of frost. Altitude, mean annual temperature, and insolation explain negligible proportions of deviance (<4%); *M. calpeiana* avoids high altitudes (>1,500 m), low mean annual temperatures, and medium insolation values. In general, the effect of land-use variables is low; most important are the percentage of forest and agricultural lands, with 8.56 and 12.8% of explained deviance, respectively. The relationship with percentage of forest is complex and cubic, whereas the relationship with the percentage of agricultural land is negatively linear. Percentage of scrub and grassland and the percentage of artificial surfaces explain <3% of deviance, both positively, linearly related with *M. calpeiana* presence-absence (Fig. 3).

Correlation analysis identifies a group of highly correlated explanatory variables, at the 0.8 threshold, composed of altitude, annual days of frost, mean annual temperature, and minimum winter temperature; the fourth variable, explaining significantly more variation than the other three, was selected as representative. Variables (and their transformations) selected
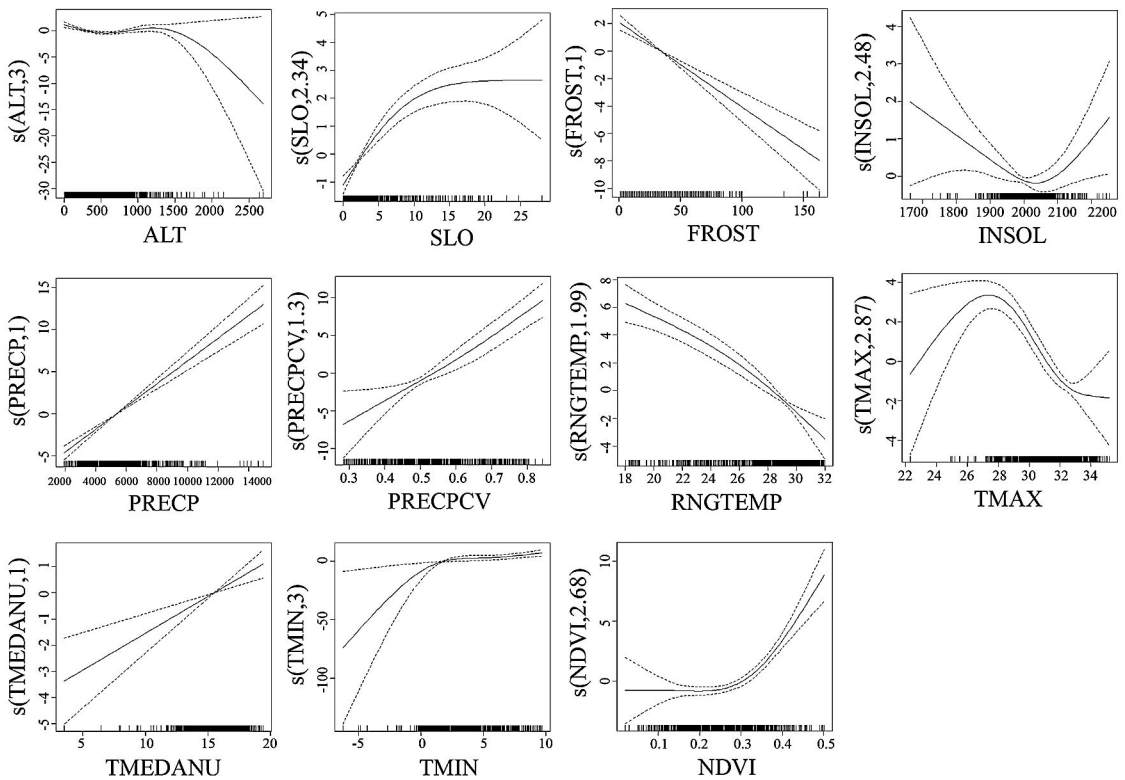
**Fig. 2.** Estimated GAM terms describing the relationships of *M. calpeiana* (Walckenaer, 1805) with the statistically significant climate, topography, and NDVI variables. Estimates are shown as solid lines, 95% confidence intervals as dashed ones, and cases as a rough plot along graph bottom. Explained deviance: mean altitude, ALT = 3.68%; slope, SLO = 17%; annual days of frost, FROST = 16.1%; insolation, INSOL = 1.72%; annual precipitation, PRECP = 58.4%; precipitation periodicity, PRECPCV = 58.9%; annual temperature range, RNGTEMP = 49.1%; maximum annual temperature, TMAX = 31.7%; annual mean temperature, TMEDANU = 3.25%; minimum annual temperature, TMIN = 31.5%; normalized difference vegetation index, NDVI = 35.3%. Predictor $P < 0.01$ ($\chi^2$ test).

for the subsequent final GLM analyses are precipitation periodicity (linear); annual precipitation (linear); annual temperature range (linear); maximum summer temperature (quadratic); minimum winter temperature (cubic); slope (cubic); annual days of frost (linear); NDVI (cubic); percentage of forest (quadratic); percentage of agricultural land (linear); percentage of scrub and grassland (linear); and percentage of artificial surfaces (linear). Amount of forest was included as a quadratic term instead of cubic, because such a complex relationship with *M. calpeiana* presence would be of difficult biological interpretation. The inconsistent relation with insolation led to the elimination of this variable.

The climate and topographic model retained only precipitation periodicity and annual precipitation as linear terms and maximum summer temperature and minimum winter temperature as quadratic terms, while accounting for 82.15% of deviance and classifying almost perfectly (AUC = 0.99, sensitivity and specificity scores of 97%). This is the maximum predictive power achievable, because it is not increased by the inclusion of any other variable, neither NDVI nor land use. A model based solely on NDVI accounts for

34.96% of deviance, with an AUC of 0.82 and sensitivity and specificity values of 76%. It includes the quadratic term of the NDVI variable. A model based only on land-use variables retains the linear terms of the percentage of forest, agricultural lands, and artificial surfaces, explaining only 14.76% of deviance (AUC = 0.76, sensitivity and specificity values of 70%). Last, a model combining both NDVI and land-use variables retained the three mentioned land-use variables and the quadratic term of NDVI, explaining 40.91% of deviance (AUC = 0.88, sensitivity and specificity scores of 0.79).

After application of the favorability function, final climate model interpolation to southern Spain (Fig. 4) highlighted the considerable favorability of mainly Huelva, Sevilla, Cádiz, and Málaga provinces, corresponding to the principal *M. calpeiana* distribution areas. Suitable habitat extends also through southern Granada; isolated potential areas on the coast of Valencia and Alicante; in eastern Jaén; and in the southeast of Cáceres province. Autocorrelation analysis showed that residuals of this final climate function are positive and significantly autocorrelated until a distance of 72 km (Fig. 5). From the added third-degree
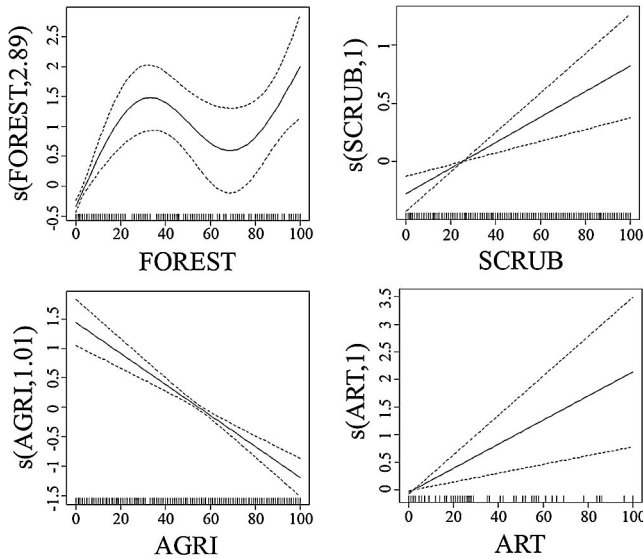
**Fig. 3.** Estimated GAM terms describing the relationships of *M. calpeiana* (Walckenaer, 1805) with the statistically significant land-use variables. Estimates are shown as solid lines, 95% confidence intervals as dashed ones, and cases as a rough plot along graph bottom. Explained deviance: percentage of woodland, FOREST = 8.56%; percentage of agricultural land, AGRI = 12.8%; percentage of scrub- and grassland, SCRUB = 2.24%; percentage of artificial surfaces ART = 1.43%. Predictor $P < 0.01$ ($\chi^2$ test).

polynomial of latitude and longitude (Legendre and Legendre 1998), the former climate model retains only linear terms, raising explained deviance to 89.06% (almost a 7% increase), although the AUC values did not change (0.97), and sensitivity and specificity increased slightly, to 98%. Addition of spatial terms slightly decreases Moran's I autocorrelation scores for the first distance classes, although still positive and significant for the first six (Fig. 5).

Variation partitioning shows the importance of the effect of climate (41.2%) and the virtual lack of relevance of the independent effect of land-use variables and NDVI (0.0%). The most important combined effect is that of climate and NDVI (26.1%), followed by the combined effect of the three groups of factors (8.8%) and the combined effect of climate and land use (5.9%); the combined effect of land-use variables

and NDVI did not explain any significant proportion of deviance (0.0%).

The independent effect of the most relevant climate variables was assessed using hierarchical partitioning (annual precipitation, precipitation periodicity, annual temperature range, maximum summer temperature, and minimum winter temperature; Fig. 6A). The greatest influence is exercised by annual precipitation and precipitation periodicity, with quite similar percentages of independent effects, 32.6% and 30.5%, respectively. Independent effects of the other three factors are smaller and similar (temperature range,
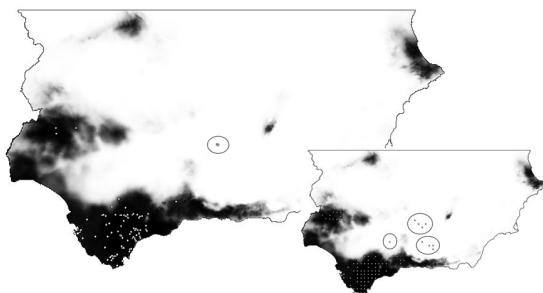


**Fig. 4.** Interpolated environmental favorability surface for *M. calpeiana* in southern Spain, with the 1 by 1-km presence points. In the small window, the 10 by 10-km presence points are overlaid. Circles indicate occurrence areas not predicted by the climate model.
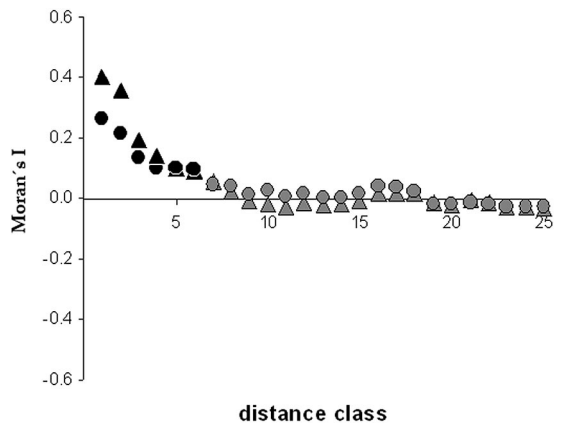


**Fig. 5.** Correlogram for the residuals of the model calculated after rescaling probabilities with the favorability function. Triangles, climate model without spatial terms; circles, climate model with spatial terms. Lag distance, 12 km; black dots, statistically significant Moran's I scores.
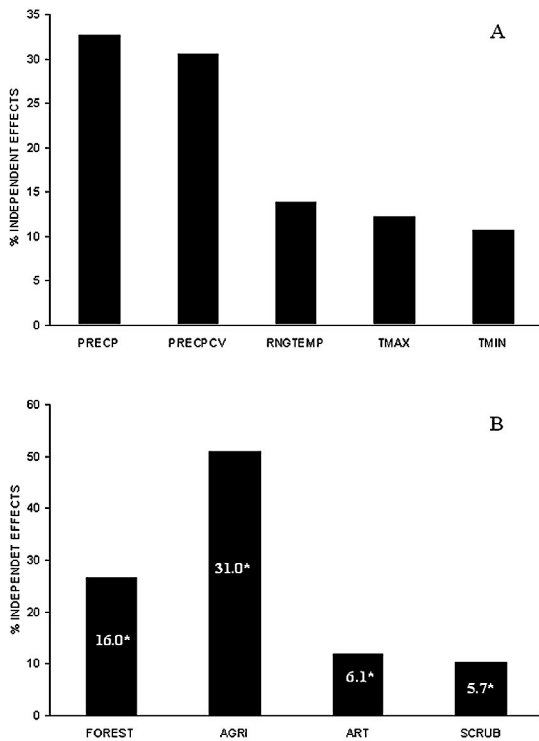
**Fig. 6.** Independent effects of climate (A) and land-use (B) variables calculated by hierarchical partitioning. Annual precipitation, PRECP; precipitation periodicity, PRECPCV; temperature range, RNGTEMP; maximum summer temperature, TMAX; minimum winter temperature, TMIN; percentage of woodland, FOREST; percentage of agricultural land, AGRI; percentage of artificial surfaces, ART; percentage of scrub- and grasslands, SCRUB. Numbers inside columns are Z-scores (*statistically significant).

13.9%; maximum temperature, 12.2%; minimum temperature, 10.7%). Among the land-use variables (Fig. 6B), the greatest influence is exercised by percentage of agricultural land (51%), followed by percentage of forests (26.7%). Percentage of artificial surfaces and of scrub and grassland had smaller percentages of independent effects, 12.0 and 10.3%, respectively. All these effects were statistically significant.

## Discussion

The distribution of *M. calpeiana* in the Iberian Peninsula, at a resolution of 1 km², is mainly determined by climate factors. Independent effects of precipitation-related variables, annual precipitation and precipitation periodicity, are similarly responsible for a great part (the greatest of the climate factors considered) of the variation in spider presence/absence. *M. calpeiana* lives in wet areas with high annual variation, clearly linearly related with these variables. Temperature-related variables are also important, although to a lesser extent than precipitation variables. While independent effects of temperature range and minimum and maximum temperature are similar, the combina-

tion of the last two variables is more important than temperature range. Minimum and maximum temperature remain in the final climate model, with curvilinear relations with the species, indicating a preference for places with moderate maximum temperatures and avoidance of areas with extreme minimum temperatures. The relevance of annual precipitation, precipitation periodicity, and maximum and minimum temperature in determining the limits of *M. calpeiana* distribution is highlighted by the large single independent effect of climate (41.2%) in variation partitioning analysis and the nearly perfect classification power of the model.

Other climate and topography variables seem to be related with *M. capleiana*, although their low explanatory power would indicate quite little relevance, at least on the scale of this study. In general, restricted by precipitation-related and maximum and minimum temperature variables, *M. calpeiana* shuns flat areas, prefers altitudes <1,500 m, and light, infrequent frosts, small temperature ranges, and high mean annual temperatures. In summary, *M. calpeiana* prefers thermomediterranean areas with maritime influence.

NDVI did not remain in the model when added after climate variables. In fact, as shown by variation partitioning, its independent contribution was virtually null; its contribution to the variation in *M. calpeiana* presence/absence (34.96%) is inseparably correlated with that of climate and land-use variables. Thus, although it seems that *M. calpeiana* is positively related with degree of vegetative vigor, its effect cannot be separated from that of climate variables and, in any case, it is less relevant than climate.

Most land-use variables were weakly related with *M. calpeiana* presence/absence. We believe that possible land-use changes occurred since collection of the older records of the spider (1980s) may not have significantly altered possible relationships. The most important independent effect is that of agricultural land, which seems to impair species presence. The cubic relationship of percentage of woodland seems unrealistic, probably reflecting the effect of other constraining variables. In fact, when all land-use variables are included in a model, the percentage of forest remains as a linear term. So, *M. calpeiana* takes advantage of woodlands, the variable with the second-highest independent effect. A curious, statistically significant, positive relation with percentage of artificial surfaces appears, although its independent effect is quite small. This pattern may reflect either the greater detectability of the spider in such areas, or the bias toward sampling in anthropogenized habitats, or *M. calpeiana* climate preferences for the conditions of the highly urbanized Iberian coast. Percentage of scrub and grassland, although positively related with spider presence, does not remain in the final land-use model, which explains a very low proportion of deviance (14.76%), with an AUC value of 0.76, and percentages of correctly predicted presences and absences of 70%. This AUC value is near the 0.7 value below which models should be regarded with skepticism, as in such cases sensitivity will not be much greater than the false

positives fraction (Swets 1988). Nevertheless, the slight effect of land-use variables cannot be separated from that of climate and NDVI variables, as shown by variation partitioning.

There have been claims, based on nonexperimental observations of local density counts, that *M. calpeiana* is favored by moderate human alteration of landscape (van Helsdinge and Decae 1992). Such favor is probably indirect because of the creation of more potential nesting sites (Ferrández and Fernández de Céspedes 1996). Our study, based on occurrence (not density) data, suggests that, on the study scale, agriculture is much more important than any other land-use type, affecting *M. calpeiana* presence negatively. The positive effect of woodlands is also much more relevant than the positive effect of percentage of artificial surfaces, although the anthropogenic impact advantageous for *M. calpeiana* may not be recognizable in our reclassification of Corine land-use classes into so few, broad categories. Also, our study scale may be too coarse to detect effects of land-use and moderate anthropogenic impact. Nevertheless, the relevance of natural cover in the distribution of *M. calpeiana* is independent of any local spider congregation in physical structures that facilitate nesting. Habitat selection studies on finer scales than this one must be carried out to develop firm conclusions about the impact of land-use on *M. calpeiana.* Additionally, once the species detectability factor has been accounted for (species may be more detectable in anthropogenized habitats), habitat suitability based on density data should be interpreted with caution. Abundance may vary in space because of a number of factors not related with long-term habitat favorability (Van Horne 1983, Nielsen et al. 2005). Reliable causal links can be obtained only from detailed demographic information (Mitchell 2005).

Interpolations to the southern Iberian peninsula of this and a 100-km$^2$ resolution model of the entire Iberian peninsula (unpublished data) are chiefly coincident, but differing mainly in a reduction in potential area around the two core *M. calpeiana* distribution areas along both Guadalquivir river margins. Consequently, unsuitable area surrounding the Guadalquivir river basin is enlarged in the fine-scale model, which fails to predict two 1 by 1-km presence points located outside Jaén city (arrow in Fig. 4). The model also fails to predict three core species areas where there were no fine-resolution occurrences (except the two in Jaén; compare maps of Fig. 4): in the province of Granada; Jaén; the south of Córdoba. Precisely, these distribution–area environmental conditions are the most marginal where *M. calpeiana* can be found; their absence from the model training process may be cause of the general reduction in potential area and the observed underestimation. Representation of the full environmental and spatial gradient in the dependent variable is essential to obtain accurate models (Vaughan and Ormerod 2003, Hortal and Lobo 2005, Jiménez-Valverde and Lobo 2006). Of special relevance, presence points at species environmental gradient limits define a distribution border in perhaps the

most extreme conditions. Occurrence data not recovered from a well-designed sampling scheme, but from heterogeneous sources (bibliography, collections, etc.), may be biased (Dennis and Thomas 2000, Jiménez-Valverde and Ortuño 2006). Despite efforts for avoiding false absences, failure to include them as presences will nevertheless affect model results.

Apart from these data-dependent drawbacks, scale may be responsible for model differences. Scale differences (resolution and extent) affect variable relationships, and parameter estimations also (Dungan et al. 2002), highlighting the difficulty encountered in selection of the appropriate scale of analysis, which may not be straightforward. Moreover, most modeling is constrained by data availability, of both dependent and independent variables. Although the effects of scale are widely recognized (Wiens 1989, Bailey et al. 2002, Pearson et al. 2004, Boyce 2006), and multiscale potential habitat studies are recommended (Martínez et al. 2003, Johnson et al. 2004, Luck 2005, Olivier and Wotherspoon 2005, Beever et al. 2006, Seoane et al. 2006), little is known about scale-of-analysis suitability to the structures and processes of study.

As in the case of other funnel-web spiders (Woodman et al. 2006), *M. calpeiana*, a low-vagile species, is probably highly conditioned by local environmental factors. Thus, absence of local variables from the modeling process may be negatively affecting the rate of correct classification, as is corroborated by the slight decrease of autocorrelation in the first distance classes after the inclusion of spatial terms (Diniz-Filho et al. 2003).

We must finally stress the importance to distribution studies of detailed geo-referenced location data, as precise as possible to enable multi-scale approaches to habitat selection, to make reliable inferences about the process under study and to generate fully useful guidance for conservation proposes. Additionally, well-designed field surveys must be carried out to recover all environmental and spatial variation of the target territory and avoid the use of biased data (Jiménez-Valverde and Lobo 2004, Hortal and Lobo 2005), as well as providing a measure of confidence on absence records.

## Acknowledgments

## References Cited

Bailey, S. A., R. H. Haines-Young, and C. Watkins. 2002. Species presence in fragmented landscapes: modeling of species requirements at the national level. Biol. Conserv. 108: 307–316.

Beever, E. A., R. K. Swihart, and B. T. Bestelmeyer. 2006. Linking the concept of scale to studies of biological diversity: evolving approaches and tools. Diversity Distrib. 12: 229–235.

Blasco, A., and M. A. Ferrández. 1986. El género *Macrothele* Ausserer 1871 (Araneae; Dipluridae) en la Peninsula Ibéricaqq. Actas X Congr. Int. Aracnol. Jaca/España I: 311–320.

Boyce, M. S. 2006. Scale for resource selection functions. Diversity Distrib. 12: 269–276.

Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: an integral part of inference. Biometrics 53: 603–618.

Busby, J. R. 1991. BIOCLIM: a bioclimate analysis and prediction system, pp. 64–68. *In* C. R. Margules and M. P. Austin (eds.), Nature conservation: cost effective biological surveys and data analysis, CSIRO, Melbourne, Australia.

Chevan, A., and M. Sutherland. 1991. Hierarchical partitioning. Am. Stat. 45: 90–96.

Chong, D.L.S., E. Mougin, and J. P. Gastellu-Etchegorry. 1993. Relating the global vegetation index to net primary productivity and actual evapotranspiration over Africa. Int. J. Remote Sens. 14: 1517–1546.

Cramer, J. S. 1999. Predictive performance of binary logit model in unbalanced samples. J. R. Statistic. Soc. D. 48: 85–94.

Dennis, R.L.H., and C. D. Thomas. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. J. Insect Conserv. 4: 73–77.

Diniz-Filho, J.A.F., L. M. Bini, and B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. Global Ecol. Biogeogr. 12: 53–64.

Dungan, J. L., J. N. Perry, M.R.T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jakomulska, M. Miriti, and M. S. Rosenberg. 2002. A balanced view of scale in spatial statistical analysis. Ecography 25: 626–640.

Engler, R., A. Guisan, A., and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. 41: 263–274.

Ferrández, M. A., and H. Fernández de Céspedes. 1996. *Macrothele calpeiana* (Walckenaer, 1805), pp. 129–141. *In* M. A. Ramos, D. Bragado, and J. Fernández (eds.), Los invertebrados no insectos de la "Directiva Hábitat" en España, Dirección General de Conservación de la Naturaleza, Madrid, Spain.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24: 38–49.

Gallon, R. C. 1994. Observations on *Macrothele calpeiana* (Walckenaer, 1805) in southern Iberia. J. Br. Tarantula Soc. 1: 1–12.

Guisan, A., and W. Thuiller. 2005. Predicting species distributions: offering more than simple habitat models. Ecol. Lett. 8: 993–1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135: 147–186.

Harrell, F.E.J. 2001. Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, New York.

Hortal, J., and J. M. Lobo. 2005. An ED-based protocol for optimal sampling of biodiversity. Biodiv. Conserv. 14: 2913–2947.

Jiménez-Valverde, A., and J. M. Lobo. 2004. Un método sencillo para seleccionar puntos de muestreo con el objeto de inventariar taxones hiperdiversos: el caso práctico de las familias *Araneidae* y *Thomisidae* (*Araneae*) en la Comunidad de Madrid, España. Ecología 18: 297–308.

Jiménez-Valverde, A., and J. M. Lobo. 2006. The ghost of unbalanced species distribution data in geographic model predictions. Diversity Distrib. 12: 521–524.

Jiménez-Valverde, A., and V. M. Ortuño. 2006. The history of endemic Iberian ground beetle description (Insecta, Coleoptera, Carabidae): which species were described first? Acta Oecol. (in press).

Johnson, C. J., D. R. Seip, and M. S. Boyce. 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. J. Appl. Ecol. 41: 238–251.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Lobo, J. M., J. R. Verdú, and C. Numa. 2006. Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). Diversity Distrib. 12: 179–188.

Luck, G. W. 2005. The habitat requirements of the rufous treecreeper (*Climacteris rufa*). 1. Preferential habitat use demonstrated at multiple spatial scale. Biol. Conserv. 105: 383–394.

Luque, F.J.R. 2001. Nuevos datos de *Macrothele calpeiana* (Walckenaer, 1805) para Jaén (España). Revta. Ibérica Aracnol. 4: 34.

Martínez, J. A., D. Serrano, and I. Zuberogoitia. 2003. Predictive models of habitat preferences for the Eurasian eagle owl *Bubo bubo:* a multiscale approach. Ecography 26: 21–28.

MacNally, R. 1996. Hierarchical partitioning as an interpretative tool in multivariate inference. Austral. Ecol. 21: 224–228.

MacNally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – "predictive" and "explanatory" models. Biodiv. Conserv. 9: 655–671.

MacNally, R. 2002. Multiple regression and inference in ecology and conservation biology: further comments on retention of independent variables. Biodiv. Conserv. 11: 1397–1401.

McCullagh, P., and J. A. Nelder. 1997. Generalized linear models: monographs on statistics and applied probability. Chapman & Hall, London, UK.

Mitchell, S. C. 2005. How useful is the concept of habitat?: a critique. Oikos 110: 634–638.

Muñoz, A. R., R. Real, A. M. Barbosa, and J. M. Vargas. 2005. Modelling the distribution of Bonelli's eagle in Spain: implications for conservation planning. Diversity Distrib. 11: 477–486.

Nielsen, S. E., C. J. Johnson, D. C. Heard, and M. S. Boyce. 2005. Can models of presence-absence be used to scale abundance? Two case studies considering extreme in life history. Ecography 28: 197–208.

Odland, J. 1988. Spatial autocorrelation. Sage Publications, Los Angeles, CA.

Olden, J. D., D. A. Jackson, and P. Peres-Neto. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. T. Am. Fish. Soc. 131: 329–336.

Olivier, F., and S. J. Wotherspoon. 2005. GIS-based application of resource selection functions to the prediction of snow petrel distribution and abundance in East Antarctica: comparing models at multiple scales. Ecol. Model. 189: 105–129.

Pearson, R. G., T. P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27: 285–298.

Perry, L. 2002. Captive breeding of the funnelweb spider *Macrothele calpeiana* (Walckenaer, 1805). J. Br. Tarantula Soc. 17: 113–121.

Pettorelli, N., J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth. 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol. Evol 20: 503–510.

R Development Core Team. 2004. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Real, R., A. M. Barbosa, and J. M. Vargas. 2006. Obtaining environmental favorability functions from logistic regression. Environ. Ecol. Stat. 13: 237–245.

Santos Lobatón, M. C. 1996. Estudio sobre *Macrothele calpeiana* Walckenaer, 1805 (Araneae, Hexathelidae) en dos pinares de la provincia de Cádiz (España). Aracnología 24: 1–10.

Sawada, M. 1999. ROOKCASE: an Excel 97/2000 Visual Basic (VB) Add-in for exploring global and local spatial autocorrelation. Bull. Ecol. Soc. Am. 80: 231–234.

Seoane, J., J. H. Justribó, F. García, J. Retamar, C. Rabadán, and J. C. Atienza. 2006. Habitat-suitability modelling to assess the effects of land-use changes on Dupont's lark *Chersophilus duponti:* a case study in the Layna Important Bird Area. Biol. Conserv. 128: 241–252.

Silva, L. C., and I. M. Barroso. 2004. Regresión logística. La Muralla, Madrid, Spain.

Snazell, R. 1986. The spider genus *Macrothele* Ausserer in Spain (Araneae; Dipluridae). Bull. Br. Ecol. Soc. 17: 80–83.

Snazell, R., and R. Allison. 1989. The genus *Macrothele* Ausserer (Araneae, Hexathelidae) in Europe. Bull. Br. Arachnol. Soc. 8: 65–72.

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

United States Geological Survey. 2005. Digital elevation model (Iberian Peninsula) (http://www.usgs.gov).

van Helsdinge, P. J., and A. E. Decae. 1992. Ecology, distribution and vulnerability of *Macrothele calpeiana* (Walckenaer) (Araneae, Hexathelidae). Tijdschr. Entomol. 135: 169–178.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. J. Wildlife Manag. 47: 893–901.

Vaughan, I. P., and S. J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conserv. Biol. 17: 1601–1611.

Walsh, C., and R. MacNally. 2003. hier.part: hierarchical partitioning. R package version 0.5–1. R Foundation for Statistical Computing, Vienna, Austria.

Wiens, J. A. 1989. Spatial scaling in ecology. Funct. Ecol. 3: 385–397.

Wood, S. N. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. B. 62: 413–428.

Wood, S. N. 2004. mgcv: GAMs with GCV smoothness estimation and GAMMs by REML/PQL. R package version 1.1–8. R Foundation for Statistical Computing, Vienna, Austria.

Wood, S. N., and N. H. Augustin. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecol. Model. 157: 157–177.

Woodman, J. D., J. E. Ash, and D. M. Rowell. 2006. Population structure in a saproxylic funnelweb spider (Hexathelidae: *Hadronyche*) along a forested rainfall gradient. J. Zool. 268: 325–333.

Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157: 261–280.