Predicting the richness of aquatic beetles and bugs in a semi-arid mediterranean region

Daniel Bruno¹, David Sánchez-Fernández¹,², José Antonio Carbonell¹, Félix Picazo¹, Josefa Velasco¹ and Andrés Millán¹

¹ (Departamento de Ecología e Hidrología, Universidad de Murcia, 30100 Murcia, Spain).
² (Museo Nacional de Ciencias Naturales, C/José Gutiérrez Abascal, 2, 28006, Madrid, Spain).

Corresponding author: dbrunocollados@um.es

Received: 20/10/2010 Accepted: 8/6/2011

ABSTRACT
Predicting the richness of aquatic beetles and bugs in a semi-arid mediterranean region

The southeastern Iberian Peninsula is a semi-arid region recognised as an area of high aquatic biodiversity. Water beetles (Coleoptera) and water bugs (Hemiptera) are two of the region’s most thoroughly studied insect groups. An exhaustive database of aquatic beetles and bugs from the Region of Murcia (SE Spain) and a set of 33 variables were employed for two related purposes: a) to describe the sampling effort conducted in the study area on these groups in order to identify zones with reliable inventories and b) with this information as a base, to forecast the distribution of species richness for each group. Theoretical estimates were based on the Clench function applied to the accumulated number of samples and were constructed with general linear models (GLM). The following results were obtained: 1) the percentage of grid cells that can be considered to be well sampled in the study area (143 10 x 10 km UTM grid cells) is approximately 14 % for beetles and 22 % for bugs; 2) well-surveyed cells are evenly distributed according to the previously defined physioclimatic subareas, which enable the use of these grid cells for modelling the richness distribution; and 3) a more accurate picture of the actual richness patterns for both groups is shown by the analysis. These results were combined to identify key areas where future sampling efforts must be focused.

Key words: Sampling effort, predictive modelling, aquatic Coleoptera, aquatic Hemiptera, richness patterns.

RESUMEN
Prediciendo la riqueza de coleópteros y hemípteros acuáticos en una región semiárida mediterránea

El sureste de la Península Ibérica es una región semiárida con una alta biodiversidad acuática, siendo los coleópteros y hemípteros acuáticos dos de los grupos de insectos mejor estudiados. Una base de datos que recopila todas las citas existentes de ambos grupos en la Región de Murcia (SE de España) y un conjunto de 33 variables han sido utilizadas con los siguientes objetivos: a) describir el esfuerzo de muestreo llevado a cabo sobre estos grupos en el área de estudio para identificar zonas con inventarios fiables y usando como base esta información b) predecir la distribución de la riqueza para ambos grupos. Basándonos en estimaciones teóricas mediante el uso de la función de Clench sobre el número acumulado de muestras y modelos lineares generalizados (GLM), hemos encontrado que: 1) de las 143 unidades espaciales utilizadas (10 x 10 km UTM) en torno a un 14 % y un 22 % para coleópteros y hemípteros acuáticos, respectivamente, pueden considerarse bien muestreadas; 2) sin embargo, las celdas bien prospectadas se encuentran equitativamente distribuidas por las subregiones fisioclimáticas previamente definidas, lo que permite el uso de estas cuadrículas para modelar la distribución de la riqueza, 3) obteniéndose una imagen más cercana a los patrones de riqueza reales de ambos grupos. Combinando estos resultados, identificamos áreas clave donde deben ser localizados los futuros esfuerzos de muestreo.

Palabras clave: Esfuerzo de muestreo, modelo predictivo, coleópteros acuáticos, hemípteros acuáticos, patrones de riqueza.
INTRODUCTION

The importance of the Iberian Peninsula in terms of biodiversity conservation has been widely recognised (Domínguez-Lozano et al., 1996; Williams et al., 2000; Reyjol et al., 2007). Approximately 80% of its area is included in the ‘Mediterranean hotspot list’ (Myers et al., 2000). The southeastern Iberian Peninsula is a region of particular interest due to its geographical location. It harbours species from Eurosiberian to northern African distributions. Therefore, although it is located in the most arid zone of Europe, it comprises a wide range of aquatic ecosystems, from freshwater streams, ponds and wetlands to hypersaline streams and continental and coastal salt pans (Vidal-Abarca et al., 1992; Millán et al., 1996). Many of these ecosystems are unique due to their ecological characteristics and the presence of rare and endemic species (Moreno et al., 1997; Sánchez-Fernández et al., 2004a; 2004b; Millán et al., 2006). However, these aquatic ecosystems are subject to high rates of human alteration and to significant water shortages. Therefore, they are at risk for biodiversity loss (Allan & Flecker, 1993; Ricciardi & Rasmussen, 1999; Saunders et al., 2002; Darwall & Vié, 2005).

To contribute to the effective conservation of these interesting and threatened ecosystems, it is desirable to rely on an exhaustive compilation of all faunal records, designed in a way that will allow the examination of the relationship between the distribution areas and their environmental and geographical determinants. Once the information is gathered, the assessment of the quality of the data is essential prior to any further analysis (Lobo & Martín-Piera, 2002; Hortal et al., 2004; 2007).

Despite the general lack of survey-effort assessment (Hortal et al., 2007), the quality of biodiversity databases from a variety of regions and covering a multitude of taxonomic groups is regularly evaluated (e.g., Dennis et al., 1999; 2006; Lobo & Martín-Piera, 2002; Reddy & Dávalos, 2003; Romo et al., 2006). To date, however, only one study of this type has focused on aquatic biodiversity (Sánchez Fernández et al., 2008), probably due to the paucity of inventory data for freshwater systems (Lévêque et al., 2005), especially in Mediterranean countries. Therefore, it is imperative to design sampling schemes that quickly and efficiently fill faunistic gaps in data that encompass the environmental and geographic variability of the study area (Funk et al., 2005; Hortal & Lobo, 2005). These efforts are particularly important to identify poorly-surveyed areas and to study the bias associated with the available information. Complete inventories may be used to study the patterns and processes that have generated biodiversity and to develop reliable predictive distribution models. These models may aid in the selection of protected natural areas (Hortal et al., 2001; Lobo & Martín-Piera, 2002; Hortal et al., 2004; Ferrier & Guisan, 2006; Lobo, 2008) and facilitate the design of more efficient survey strategies (Sánchez-Fernández et al., 2011).

In the study area, several authors studied species distributions, ecological traits and their use as biodiversity surrogates for both Coleoptera (Millán et al., 1996; 2006; Moreno et al., 1997; Sánchez-Fernández et al., 2004a; 2004b; 2006; Abellán et al., 2005; 2007) and Hemiptera (Millán et al., 1988; Moreno et al., 1997; Carbonell, 2008; 2010). The initial attempts to assess the sampling effort in the study area through use of the Clench function have suggested relatively complete inventories of water beetles (74%) (Sánchez-Fernández et al., 2004a) and water bugs (95%) (Carbonell, 2010) on a regional scale. However, data on faunistics and information on occurrence and diversity patterns are probably incomplete and determined by the distribution of previous sampling efforts (Lobo & Martín-Piera, 2002; Hortal et al. 2007). The expansion of the knowledge of water beetle and bug distribution patterns enables their use as conservation tools. This principle is especially true for water beetles, which have been suggested by several authors as indicator taxa for monitoring population trends in other species and for identifying areas of high regional biodiversity (Davis et al., 1987; Foster et al., 1990; Foster & Eyre, 1992; Sánchez-Fernández et al., 2004a; 2004b; Abellán et al., 2005).
Predicting aquatic beetles and bugs in a semiarid region

The main goal of this paper is to make simple predictions of the geographic distribution of the species richness of aquatic Coleoptera and Hemiptera in the southeastern Iberian Peninsula. To achieve this aim, an assessment of sampling effort was carried out in the study area. Only data for areas with reliable information were selected and the values estimated by the asymptotic value of the Clench accumulation curves were used as a dependent variable in the subsequent modelling procedure. Whether or not well-surveyed areas covered the environmental heterogeneity in the study area was also checked. By combining these results, we identified key areas where future sampling efforts must be focused.

METHODOLOGY

Study area

The study area corresponds to the ‘Región de Murcia’, located in the southeast of the Iberian Peninsula and encompassing an area of 11,137 km$^2$ (Fig. 1). This region has a Mediterranean climate, with an annual average rainfall of approximately 300 mm and an annual average temperature of 16°C. In a European context, it is an area of high environmental heterogeneity and singularity. It includes a high diversity of bioclimatic stages, lithology and geology and consequently supports a wide range of aquatic ecosystems, from headwater streams to coastal wetlands. The various conservation statuses of these areas serve to increase the interest of the present study (Millán et al., 2006). This paper focuses on the 143 UTM 10 × 10 km grid cells that overlap the ‘Región de Murcia’. The defined cells are based on the Mercator Coordinate System.

Source of biological data

The records of aquatic bugs and beetles were compiled from an exhaustive database called BIODIVERSIDAD developed by the University of Murcia ‘Ecología Acuática’ group. This database is the most complete source of information available for aquatic invertebrates (water bugs and beetles at the specie level and the remaining aquatic macroinvertebrates at the family level) in the study area, including all available geographical and biological data from the literature up to 2009 and additional data from museum and private collections, PhD theses and other unpublished sources.

From this database, 3825 records for 162 species of water beetles (see species richness dis-
tribution in figure 2a) and 1463 records for 33 species of water bugs (see species richness distribution in figure 2b) in the study area were collected. The data were referenced at a resolution of 100 km² (UTM 10 × 10 km squares) and these cells were used as geographical units (n = 143).

Assessing sampling effort and identifying well-surveyed cells

To determine the adequacy of sampling (degree of completeness) in each cell, the asymptotic value of the Clench function for the accumulated number of database samples was employed. The Clench function (see Soberón & Llorente, 1993) estimates the cell-specific mean rate of species addition per unit of sampling effort, using here the number of database samples as a sampling effort surrogate. The ratio between the asymptotic value of the Clench function and the observed richness was used as a measure of the completeness of each cell’s inventory. Complementarily, a non-parametric estimator (Jackknife 1) was used to guarantee the accuracy and confidence of the selection procedure. This estimator performs best at low sampling effort (Mao and Colwell, 2005) and fine grain (Hortal et al., 2006).

Previous studies have recommended using a 70 % threshold as the basis for considering an area well sampled (Jiménez-Valverde & Hortal, 2003). However, we used a threshold of 65 %, the maximum value that allows us to conduct further biodiversity studies. We felt that this choice was justified for the following reasons: i) the two complementary richness estimators facilitate the confident selection of well-surveyed cells; and ii) the available information is limited by the evident scarcity and paucity of data on non-charismatic groups in Mediterranean countries (Ramos et al., 2001) and especially in freshwater ecosystems (Lévêque et al., 2005). In general, the most restrictive approach has been the Clench function. Accordingly, the values estimated by the asymptotic value of the Clench accumulation curves for these well-sampled cells were used as dependent variables in the subsequent modelling procedure (see Hortal et al., 2004).

Environmental representation by the well-surveyed cells

Unevenness in sampling effort may result in partial (and biased) descriptions of biodiversity variation (Dennis, 2001). This common drawback limits the usefulness of existing databases and/or atlases to accurately describe biodiversity patterns (Hortal et al., 2007). Although several assessments of environmental bias and environmental completeness have been proposed (e.g., Kadmon et al., 2003; 2004; Hortal et al., 2008), we checked whether or not the previously recognised well-sampled cells represented the envi-
ronmental heterogeneity of the study area on a broad scale. To perform this determination, we defined physioclimate subregions with a set of 28 variables (the topographic and climatic variables) (Table 1). All the data were obtained from the EDIT-Geoplatform (Sastre et al., 2009).

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topographic variables</strong></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>Altitudinal range (m) (A3-A2)</td>
</tr>
<tr>
<td>A2</td>
<td>Minimum altitude (m)</td>
</tr>
<tr>
<td>A3</td>
<td>Maximum altitude (m)</td>
</tr>
<tr>
<td>A4</td>
<td>Mean altitude (m)</td>
</tr>
<tr>
<td><strong>Climatic variables</strong></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Annual mean temperature (°C)</td>
</tr>
<tr>
<td>B2</td>
<td>Max temperature of warmest month (°C)</td>
</tr>
<tr>
<td>B3</td>
<td>Min temperature of coldest month (°C)</td>
</tr>
<tr>
<td>B4</td>
<td>Mean temperature of wettest quarter (°C)</td>
</tr>
<tr>
<td>B5</td>
<td>Mean temperature of driest quarter (°C)</td>
</tr>
<tr>
<td>B6</td>
<td>Mean temperature of warmest quarter (°C)</td>
</tr>
<tr>
<td>B7</td>
<td>Mean temperature of coldest quarter (°C)</td>
</tr>
<tr>
<td>B8</td>
<td>Temperature mean diurnal range (Mean of monthly (max temp-min temp)) (°C)</td>
</tr>
<tr>
<td>B9</td>
<td>Temperature annual range (B2-B3) (°C)</td>
</tr>
<tr>
<td>B10</td>
<td>Temperature seasonality (standard deviation *10)</td>
</tr>
<tr>
<td>B11</td>
<td>Isothermality (B8/B9)(*10)</td>
</tr>
<tr>
<td>B12</td>
<td>Average annual precipitation (mm)</td>
</tr>
<tr>
<td>B13</td>
<td>Precipitation of wettest month (mm)</td>
</tr>
<tr>
<td>B14</td>
<td>Precipitation of driest month (mm)</td>
</tr>
<tr>
<td>B15</td>
<td>Precipitation seasonality (Coefficient of Variation)</td>
</tr>
<tr>
<td>B16</td>
<td>Precipitation of driest quarter (mm)</td>
</tr>
<tr>
<td>B17</td>
<td>Precipitation of wettest quarter (mm)</td>
</tr>
<tr>
<td>B18</td>
<td>Precipitation of warmest quarter (mm)</td>
</tr>
<tr>
<td>B19</td>
<td>Precipitation of coldest quarter(mm)</td>
</tr>
<tr>
<td>B20</td>
<td>Monthly mean solar radiation (Mj/m²/day)</td>
</tr>
<tr>
<td>B21</td>
<td>Aridity (B12/B23)</td>
</tr>
<tr>
<td>B22</td>
<td>Real evapotranspiration (mm)</td>
</tr>
<tr>
<td>B23</td>
<td>Potential evapotranspiration (mm)</td>
</tr>
<tr>
<td>B24</td>
<td>Hydrologic balance (mm)</td>
</tr>
<tr>
<td><strong>Spatial variables</strong></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Longitude (UTM X)</td>
</tr>
<tr>
<td>C2</td>
<td>Latitude (UTM Y)</td>
</tr>
<tr>
<td><strong>Lithologic variables</strong></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>Lithological diversity (Shannon Índex)</td>
</tr>
<tr>
<td><strong>Water bodies availability and diversity</strong></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>Total water surface (m²)</td>
</tr>
<tr>
<td>E2</td>
<td>Water environment diversity (Shannon Índex)</td>
</tr>
</tbody>
</table>
which is freely accessible at http://edit.csic.es/GISdownloads.html. A Principal Component Analysis (PCA) (with the scores of these variables at each $10 \times 10$ UTM square standardised to a mean of 0 and a standard deviation of 1) was performed to obtain uncorrelated factors (Varimax rotation) that compress the environmental variability into new factors that show eigenvalues above 1. Subsequently, CLUSTER and SIMPROF analyses (using Primer 6) (Clarke, 1993) were applied to the new variables or axes obtained from the PCA to elaborate a classification of the main physiologic subregions. Finally, a Chi-square test was performed to determine whether the well-surveyed cells presented a balanced distribution in the different defined subregions.

Modelling species richness

Spearman’s correlations were used to identify highly correlated variables within the complete set of potential predictors (Table 1) to avoid redundant information in the analysis. From among each group of highly correlated variables ($R > 0.95$), those that simultaneously showed a clear biological meaning were selected. As a result of this screening procedure, 18 environmental variables were removed due to high correlations with other variables. The longitude, latitude, mean altitude, altitudinal range, temperature annual range, temperature mean diurnal range, maximum temperature in the warmest month, mean temperature in the wettest quarter, isothermality, precipitation in the wettest month, precipitation in the coldest quarter, monthly mean solar radiation, total water surface, water environment diversity and lithological diversity were selected as the variables suitable for use as potential predictors of the target-group distribution patterns.

The estimated species richness values calculated with the accumulation curves (only including well-surveyed cells) were regressed against the final selection of more representative environmental variables with general linear models (GLM, McCullagh & Nelder, 1989). A Poisson error distribution was assumed for the dependent variable (estimated number of species). The dependent variable was related to the set of predictor variables via a logarithmic link function. To allow curvilinear relationships, a quadratic function of each factor was included in the regression model. A stepwise procedure was used to enter the factors into the model (Nicholls, 1989; Lobo & Martín-Piera, 2002). First, the linear or quadratic function of the environmental variable that accounted for the most important change in deviance was entered (a cubic term was included only for spatial variables). The remaining variables were added sequentially to the model according to their estimated explanatory capacity. The procedure was repeated until no more statistically significant explanatory variables remained ($p \leq 0.05$). The statistical significance of the terms that had already been selected was tested at each step by subjecting the new model to a backward selection procedure. The terms that became non-significant in this phase were then excluded. Akaike’s Information Criterion (AIC) was also used to compare the model selected by this stepwise procedure with other models composed of all possible predictor subsets.

The final model was evaluated by detecting outliers (cells with residual absolute values higher than mean ± standard deviation) and calculating the potential leverage (a measure of the distance of each observation from the centroid of the multi-dimensional space defined by the variables included in the model; Nicholls 1989). Thus, the residuals were examined to determine whether they were due to erroneous data or to the environmental uniqueness of the cells. Once the outliers had been removed and after having checked the balanced distribution of the final dataset, the model was rebuilt and applied. In addition, we identified the squares that remained with high residual values (values higher than mean ± standard deviation).

The predictive power of the final model was estimated by a Jackknife procedure. The model parameters were estimated as many times as the number of cells considered to be adequately surveyed (n) by first deleting each cell once and then comparing the predicted values against the values of the dependent variables. The percentage of error for each cell value was subsequently
calculated to obtain the mean error percentage, which was used as a measure of model reliability (Lobo & Martín-Piera, 2002; Hortal et al., 2004, Sánchez-Fernández et al., 2011). The STATISTICA 6.1 package (StatSoft Inc. 2004) was employed for all statistical computations.
RESULTS

Well-surveyed cells

According to the Clench function, only 18 grid cells (10 × 10 km UTM) for aquatic beetles (Fig. 3a) can be considered to be well surveyed (23% of the cells with data and 13% of the whole study area). For bugs, 31 grid cells can be considered to be well surveyed (44% of the cells with data and 22% of the whole study area) (Fig. 3b).

Physioclimatic subregions

The PCA compressed the variables that defined the established physioclimatic subregions. The compressed description showed that three factors accounted for 89.8% of the total variability. Factor 1 was positively related to the annual mean temperature and negatively related to the annual precipitation. Factor 2 was positively related to altitudinal range and Factor 3 was negatively related to the temperature mean diurnal range. The CLUSTER and SIMPROF analysis revealed three main physioclimatic subregions (Fig. 4a and 4b). These three areas represent the climatic and topographic gradients (NW-SE) present in the ‘Región de Murcia’, from the high, wet and cold northwest to the low, dry and hot southeast. The distribution of the final set of cells used to construct the model after the detection of outliers was balanced across the three obtained physioclimatic subregions for both the beetle data ($p \leq 0.01$ and 999 permutations) (Chi-square test $\chi^2 = 2.262; p = 0.323; d.f. = 2$) and the bug data ($p \leq 0.01$ and 999 permutations) (Chi-square test $\chi^2 = 2.108; p = 0.348; d.f. = 2$).

Modelling species richness

The model explaining the highest percentage of deviance and with the lowest AIC value for water beetles was: $S = \exp(0.16 \text{Altitudinal range}^2 + 0.34 \text{Altitudinal range} + 0.14 \text{Precipitation of the coldest quarter} + 0.12 \text{Water environments diversity} + 3.74)$. This model could explain 68.1% of the total variability of species richness, with a mean error percentage of 32.4%. The model for bugs was as follows: $S = \exp(-0.44 \text{Mean Temperature of wettest quarter}^2 - 0.29 \text{Mean Temperature of wettest quarter} + 0.22 \text{Water environment diversity} - 0.37 \text{Total water surface} +$

![Figure 5. Predicted species richness and squares with high residual values (marked cells with white dots) for a) water beetles and b) water bugs. Riqueza predicha de especies y cuadrículas con residuos altos (celdas marcadas con puntos blancos) para a) coleópteros acuáticos y b) hemípteros acuáticos.](image-url)
This model could explain 80.6% of the total variability of species richness, with a mean error percentage of 38.2%.

These simple final models were then applied to the entire study area. The simulated geographic distribution pattern showed a clear regional richness gradient, increasing from the southeast to the northwest for water beetles (Fig. 5a). The interior of the region was the most diverse area for water bugs (Fig. 5b). The squares that remained with high residual values are also shown in the figures.

DISCUSSION

Although both target groups are macroinvertebrates and are traditionally less well known, both groups showed reliable inventories on a regional scale. Nevertheless, these results suggest that less than a quarter (13% for beetles and 23% for water bugs) of the study area can presently be considered to be well surveyed (completeness values > 65%). For similar levels of survey effort, the aquatic bug inventories are more complete due to their lower diversity. This lower diversity facilitates the construction of a good regional inventory based on fewer records. In this context, a general lack of complete and extensive inventory data for aquatic taxa (Lévêque et al., 2005) at a fine scale (10 × 10 km squares) is apparent because it is probable that water beetles and bugs could be considered the best-studied groups of freshwater invertebrates in the Spanish Mediterranean region.

Based on the self-evident geographic pattern in the survey, it appears that additional effort is required in certain areas. The northeastern and northwestern parts of the region show an evident lack of surveys. Nevertheless, because only well-sampled cells were used to construct the model, this situation does not imply a lack of data reliability. These cells show a balanced distribution in relation to the environmental heterogeneity of the study area.

The predictive model obtained in this study identified the northwest zone as the most diverse area for aquatic beetles and showed an increasing richness gradient from the coast to this zone. Although the most diverse areas for aquatic bugs are located in the interior of the region, the obtained potential distribution map does not show a clear richness distribution pattern. These predicted high-richness areas differ noticeably in their environmental features. The northwest of Murcia shows a high altitude, a steep orography and a supra-Mediterranean climate with more frequent rainfall and lower temperatures than the interior of the region, which has a meso-Mediterranean climate with a high thermic contrast and a medium altitude. The conservation status of the two areas differs markedly. The interior of the region (especially the northeast) has experienced a profound land use transformation that has negatively affected the quality of aquatic ecosystems. In contrast, the northwest remains the most natural area and boasts well-preserved ecosystems. Such ecosystems are generally infrequent in Mediterranean zones (Gasith & Resh, 1999).

The observed differences in the distribution patterns of both groups can be explained by an ecological difference. The ecological requirements of the majority of aquatic beetle species are much narrower than those of generalist aquatic bugs (Polhemus & Polhemus, 2008; Carbonell, 2010). In fact, water beetles have been identified as excellent surrogates for inland water biodiversity due to their highly specific ecological requirements and the significant correlations between their species richness patterns and those of other groups (Bilton et al., 2006; Sánchez-Fernández et al., 2006). Thus, the use of a species richness prediction map as a surrogate for the biodiversity relevance of each cell could be justified by the frequent positive correlation between water beetle species richness and rarity (Sánchez-Fernández et al., 2004), which makes this group an interesting conservation tool. Although aquatic bugs have more complete inventories, their distribution pattern is much more difficult to interpret. It is probable that this difficulty results from their broad ecological requirements and high dispersal ability. Nevertheless, aquatic bugs seem to prefer the central zones of the region because these areas offer a combination of lotic and standing waters with some degree of human
impact. Such areas constitute preferred habitats for this group of insects (Millán et al., 1988; 2002).

Although it is probable that the forecast obtained from this species richness model is far from accurate (it has a mean prediction error of approximately 30 %), the resulting potential distribution map shows a picture that could closely approximate reality. Furthermore, the results can be used to determine whether more intense sampling effort has been focused on the most diverse areas. In this case, Spearman’s correlations between the observed richness and the number of records are highly positive for beetles ($R = 0.96$, $p < 0.01$) and bugs ($R = 0.88$, $p < 0.01$). However, the predicted richness patterns of water beetles and bugs differed greatly from the actual sampling effort ($R = 0.1$, $p = 0.37$) for beetles and ($R = 0.1$, $p = 0.41$ for bugs). This result suggests that entomologists’ intuitions could have failed to select the best focus for their sampling effort in this case.

**Identifying future sampling areas**

Because no differences in the distribution of well-surveyed cells were detected among the different physioclimate subregions, we recommend that further sampling effort be located in the most probable species-rich regions that are also insufficiently surveyed. This method seeks a cost-effective selection of further survey areas that will increase the information to a considerable degree (because new records mean new species in the grid cell). In contrast, a complementary strategy involves increasing the number of well-surveyed areas to improve the richness distribution obtained by predictive modelling. To implement this second strategy, two different types of squares need further surveys. These surveys should include moderately surveyed squares with completeness values near the established threshold. The new information from these cells should be included in the construction of the model. The squares with high residual values should also be included because they represent areas where the model cannot be used as a good proxy of real richness and where further survey effort is therefore required. At the same time, a balanced distribution of well-surveyed cells in the defined physioclimate subregions should be maintained.

The areas selected by the first strategy considered (the cost-effective method) are shown in

![Figure 6. Sampling priority areas for a) water beetles and b) water bugs. Áreas de muestreo prioritarias para a) coleópteros acuáticos y b) hemípteros acuáticos.](image-url)
Predicting aquatic beetles and bugs in a semiarid region

According to these results and based on the self-evident geographic pattern in the survey, the priority sites to be sampled for aquatic beetles are located in several northwestern and central areas. For aquatic bugs, the most advantageous areas would be the northeastern parts of the region. Additional isolated areas distributed throughout the region should also be sampled for this insect group.

Implications for conservation

In view of the high human pressure that threatens the ecological status of aquatic ecosystems, it is important to verify the actual distribution of the groups considered by this study and then to increase sampling effort in the areas previously mentioned to improve and validate the models. It is possible that any intensification of survey efforts in these areas would produce lower-than-expected species richness values because of the relatively recent impoverishment of the areas’ natural assemblages.

If the ultimate goal is to develop more accurate conservation strategies by incorporating invertebrate data, there is a substantial need for reliable information on the distributional patterns of these groups. Although it should be noted that the database used here is the most complete source of information available for aquatic invertebrates in the SE of the Iberian Peninsula, additional sampling effort is undoubtedly needed. Predictive modelling techniques could be an effective and useful tool for designing the required sampling protocols and for contributing to a more realistic picture of the distribution of these groups.

ACKNOWLEDGEMENTS

We thank Cayetano Gutiérrez, Pedro Abellán, Oscar Belmar, Paula Arribas and the remainder of the ‘Ecología Acuática’ research group (Universidad de Murcia) for their wholehearted support. This work was supported by predoctoral grants (FPU) from MICINN to D.B. and J.A.C., predoctoral and postdoctoral grants from the Fundación Séneca to F.P. and D.S.-F, respectively, and projects CGL2006-04159, 023/2007 (A.M.).

We also thank the ‘AIL’ committee for recognising the poster based on this paper with an award as the best poster at the XV Congress of the ‘AIL’.

REFERENCES

DENNIS, R. L. H. 2001. Progressive bias in species status is symptomatic of fine-grained mapping...


MILLÁN, A., J. VELASCO & L. RAMÍREZ-DÍAZ. 1996. Distribución espacial de los Adephaga acuá-
Predicting aquatic beetles and bugs in a semiarid region


STATSOFT INC. 2004. STATISTICA (data analysis software system). StatSoft Inc. Tulsa


WILLIAMS, P. H., C. HUMPHRIES, M. B. ARÁUJO, R. LAMPINEN, W. HAGEMEIJER, J. P. GASC & T. MITCHELL-JONES. 2000. En-
Bruno et al.