Using habitat distribution models to evaluate large-scale landscape priorities for spatially dynamic species

Regan Early*, Barbara Anderson and Chris D. Thomas

UKPopNet, Department of Biology, University of York, York YO10 5YW, UK

Summary

1. Large-scale conservation planning requires the identification of priority areas in which species have a high likelihood of long-term persistence. This typically requires high spatial resolution data on species and their habitat. Such data are rarely available at a large geographical scale, so distribution modelling is often required to identify the locations of priority areas. However, distribution modelling may be difficult when a species is either not recorded, or not present, at many of the locations that are actually suitable for it. This is an inherent problem for species that exhibit metapopulation dynamics.

2. Rather than basing species distribution models on species locations, we investigated the consequences of predicting the distribution of suitable habitat, and thus inferring species presence/absence. We used habitat surveys to define a vegetation category which is suitable for a threatened species that has spatially dynamic populations (the butterfly Euphydryas aurinia), and used this as the response variable in distribution models. Thus, we developed a practical strategy to obtain high resolution (1 ha) large scale conservation solutions for E. aurinia in Wales, UK.

3. Habitat-based distribution models had high discriminatory power. They could generalize over a large spatial extent and on average predicted 86% of the current distribution of E. aurinia in Wales. Models based on species locations had lower discriminatory power and were poorer at generalizing throughout Wales.

4. Surfaces depicting the connectivity of each grid cell were calculated for the predicted distribution of E. aurinia habitat. Connectivity surfaces provided a distance-weighted measure of the concentration of habitat in the surrounding landscape, and helped identify areas where the persistence of E. aurinia populations is expected to be highest. These identified successfully known areas of high conservation priority for E. aurinia. These connectivity surfaces allow conservation planning to take into account long-term spatial population dynamics, which would be impossible without being able to predict the species’ distribution over a large spatial extent.

5. Synthesis and applications. Where species location data are unsuitable for building high resolution predictive habitat distribution models, habitat data of sufficient quality can be easier to collect. We show that they can perform as well as or better than species data as a response variable. When coupled with a technique to translate distribution model predictions into landscape priority (such as connectivity calculations), we believe this approach will be a powerful tool for large-scale conservation planning.

Key-words: Marsh Fritillary, connectivity, reserve selection, response variable, metapopulation dynamics

Introduction

The identification of conservation priority areas across large spatial scales is important in order to develop effective solutions for species protection (Margules & Pressey 2000; Vos et al. 2001; Rouget et al. 2006). There is also a growing recognition that these priority areas should contain areas with high projected species persistence (Araujo & Williams 2000; Possingham, Ball & Andelmann 2000; Cabeza & Moilanen 2001). However, species distribution data are often not complete at the high spatial resolution and large spatial extent required to achieve this. Habitat distribution models are now being used regularly to help identify priority areas in situations where existing species data are insufficient (Anderson & Martinez-Meyer 2004; Gibson et al. 2004; Posillico et al. 2004; Wintle, Elith & Potts 2005). However, constructing high-resolution distribution models using species data can be problematic if the species is not recorded at...
all the sites which are actually suitable for it. These sites are known as ‘false-negatives’ and cause considerable problems with the accuracy of distribution models (Heglund 2001; Tyre et al. 2003; Guisan & Thuiller 2005).

Incomplete distribution data stem commonly from a lack of sufficient recording, but they can also arise from the natural dynamics of a target species. If population densities are low, individuals of mobile species may not be present at a particular site at the time of recording. In addition, the metapopulation assumption that populations persist through colonizations and extinctions of subpopulations (Hanski & Gaggiotti 2004) implies that species distributions are temporally variable even where the distribution of habitat is unchanging. Thus, the recorded locations of a metapopulation species at any one time will include temporary species absences, which are false-negatives in the context of distribution models (Heglund 2001; Tyre et al. 2003). In order to identify all suitable locations, long-term location data sets may be required (Welk 2004; Herben et al. 2006). However, even long-term species data do not solve the problem unless accompanied by long-term data on changes to the environment: historical records may correspond to locations where the species’ habitat has subsequently been destroyed.

The problems posed by metapopulation dynamics or otherwise incomplete species records for distribution modelling are widely acknowledged (for example Collingham et al. 2000; Luoto et al. 2001; Austin 2002). However, there is little consensus on how to compensate for them. Guisan & Thuiller (2005) recommended that the stochasticity of fluctuations in habitat occupancy be used to set a limit on how well a model can explain a given species’ distribution. Alternatives are to use co-occurring species as surrogates (Munzbergova & Herben 2004), or to remove spatial population dynamic effects by incorporating metapopulation principles into distribution models (Rushton et al. 1996; Wahlberg, Moilanen & Hanski 1996; Heikkinen et al. 2005; Helm, Hanski & Partel 2006). However, given the lack of established statistical techniques available to compensate for the various problems, species presence/absence data may not always be the ideal response variable when attempting to model a habitat distribution. Here we have developed an alternative approach in which we use the location of a species’ habitat, rather than the species itself, as the response variable in distribution models. We used this approach to identify the distribution of potential habitat at a higher resolution (1 ha) and over a larger geographical area (approximately 2 million ha) than it would have been feasible to map the target species directly.

In addition to habitat quality, the quantity and distribution of the habitat fragments that metapopulation species occupy is vital to their long-term persistence (Hanski & Ovaskainen 2000; Hokit, Stith & Branch 2001; Vos et al. 2001). The concept of ‘connectivity’ is used commonly to sum up these attributes of a landscape (Moilanen & Nieminen 2002). The connectivity of a given cell represents the likelihood that an individual will disperse to it, were all the habitat populated by the target species. Thus, connectivity surfaces can identify areas of habitat concentration which contribute to population persistence (Early & Thomas 2007). Here, we use habitat distribution predictions to calculate habitat connectivity surfaces and identify priority landscapes over a large geographical area.

We tested the use of species’ habitat as the response variable in distribution models, using the habitat distribution of the butterfly Euphydryas aurinia (Rottemburg, 1775) in Wales. This species is threatened in Europe (Warren 1994) and displays metapopulation dynamics, such that it exhibits local extinction and colonization dynamics and occupies only a low proportion of its potential habitat at any one time (Lewis & Hurford 1997; Wahlberg et al. 2002; Wang et al. 2004; Schtickzelle et al. 2005; Bulman et al. 2007). We constructed 1-ha resolution distribution models for its habitat in well-mapped focal landscapes using predictor variables for which data are widely available. These were then compared with distribution models constructed using the available records of E. aurinia. We used the habitat-based distribution models to predict habitat connectivity in other well-mapped landscapes to test model accuracy. We then extrapolated to the whole of Wales, and tested whether habitat-based models accurately predicted the known distribution of the butterfly species itself.

Methods

‘HABITAT’ AS A RESPONSE VARIABLE

Presence or absence of suitable habitat was taken from habitat surveys commissioned by the Countryside Council for Wales (CCW). Surveys were conducted by professional fieldworkers employed by CCW, following the expert- and analysis-based guidelines set out by Fowles & Smith (2006). Habitat definitions were derived through consultation with a wide range of butterfly experts and land managers. ‘Absence’ (i.e. unsuitable) locations were those where the butterfly’s host plant [Succisa pratensis (Moench)] was absent or ‘rare’ according to the DAFOR scale (Dominant, Abundant, Frequent, Occasional, Rare), and vegetation characteristics implied that radical changes in agricultural management and hydrology would be needed to make the site suitable. All other locations were classed as potential E. aurinia habitat (‘presences’). These were sites where S. pratensis occurred more than rarely and with the underlying vegetation characteristics to become good condition habitat. Although these sites varied in habitat quality, differentiating within these classes would have required information on management practice which is generally unavailable. The habitat predictions made here indicate the distribution of habitat with the potential to be suitable E. aurinia habitat, under appropriate management regimes. Had more detailed habitat quality data been available, we would expect the models to have fitted even better.

Presence/absence of potential habitat was determined by overlaying the habitat survey map with a 1-ha resolution grid (in line with the UK Ordnance Survey grid system). Any 1-ha grid cell that contained any potential habitat was classed as a presence. All cells in which some habitat had been surveyed but no potentially suitable habitat was found were classed as absences. This was unlikely to lead to cells in which potential habitat was present, but not surveyed, to be classed as ‘unsuitable’ because surveyors were instructed to investigate any habitat adjacent to survey sites if it appeared to have the potential to be suitable habitat.
PREDICTOR VARIABLES

Vegetation, soil and topographic data were used as predictor variables. The first two were available as global information system (GIS) vector shapefiles (ArcGIS version 9.1 was used for data manipulation throughout) and were converted to a proportion of each vegetation/soil type in each 1-ha cell. Topographic data were available as a 5-m resolution raster, and these were used to give a single average value for each 1-ha cell. All variables were thought to provide information on vegetation composition or agricultural use, both of which are related closely to habitat suitability for *E. aurinia* (Bühler & Schmid, 2001; Bulman 2001).

Vegetation characteristics were derived from professionally conducted Phase 1 vegetation surveys, commissioned by CCW between 1989 and 1997 (Blackstock & Howe 2003). Phase 1 surveys map vegetation units > 25 m². Phase 1 vegetation categories that can support *E. aurinia* habitat (Fowlkes & Smith 2006) were included within models, as was ‘improved’ grassland, which is extremely unlikely to contain *E. aurinia*’s host plant but has a very high coverage in Wales (Table 1). The original vegetation subcategories were grouped into broader categories, in order to avoid classification error, difficulties with ‘transitional’ vegetation and temporary management impacts on vegetation type (Table 1 Cherrill & McClean 1999; Stevens et al. 2004).

Soil data at 1 : 250 000 resolution were taken from the national soil map held by Cranfield University (www.landis.org.uk), which uses classifications defined by Clayden & Hollis (1984). Cambic Stagnohumic gley soil (hereafter referred to as CS gley soil) was the only soil type that was more common at potential sites than elsewhere in the regions surveyed (38% of surveyed sites were on CS gley soil; 80% of these were suitable). Therefore, CS gley soil was the only soil type used in modelling. CS gley soil has a peaty topsoil that is associated frequently with marsh, heath, mire and fen vegetation, and is often waterlogged.

Altitude, slope and a ‘wetness’ index (based on topography) were derived from the ‘nextmap’ digital elevation model (DEM) (www.neodc.rl.ac.uk), which was built from satellite imagery at 25 m² resolution. One-ha grid cells were assigned the mean of each topographic variable for all the 25 m² cells which fell within it. ‘Slope’ (in degrees) for any cell was calculated as the greatest difference between the elevation of it and any of its eight neighbouring cells. The ‘wetness index’ was calculated as,

\[ \frac{\alpha}{\tan \beta} \]

eqn 1

\[ \alpha \] is the area draining through a cell, and \[ \beta \] is the angle of the slope of that cell. A high index indicates water accumulation (Bevan 1997). Wetness may be important in making fine-scale distinctions among vegetation types, for example the boundary between bog and flush, and thus may affect the distribution of *S. pratensis* and the probable agricultural uses for land.

HABITAT-BASED MODEL CONSTRUCTION

Models were constructed in \( \Gamma \) (Ihaka & Gentleman 1996) using generalized linear models (GLMs) with binomial error and a logit link function. GLMs were chosen for their ease of biological interpretation and ability to make predictions outside the training data (Elith & Burgman 2000; Guisan & Zimmermann 2000). Generalized additive models, using ‘stepgam’ function of the ‘gam’ package in \( \Gamma \) (Chambers & Hastie 1992), were also tried but had consistently much lower ability to discriminate between presence/absence within the testing data than the GLM approach. Using GLM, linear relationships of habitat suitability with all variables were tested, as well as quadratic relationships with the topographical variables to permit maximum (or minimum) responses at intermediate values. Stepwise regression in both directions, using the Akaike information criterion (AIC) as the acceptance and rejection criterion, was used to select models (see Johnson & Omland 2004).

Significant multicollinearity did occur among some of the predictor variables, potentially causing erroneous estimates of variable significance and over-complex models (Graden & Hails 2002; Graham 2003). However, correlations were small (Spearman’s \( r \leq 0.2 \)). Additionally, the use of AIC in model selection includes variables which increase the fit of a model relative to the increase in the number of variables, regardless of their significance. Furthermore, model predictive accuracy was tested on separate training and testing landscapes (below).

CONSTRUCTING CONNECTIVITY SURFACES

A classification threshold was applied to the probabilistic outputs of the habitat-based models to convert them into presence/absence of potential habitat. Classification thresholds should be chosen according to criteria which are relevant to research aims and the data used (Manel, Williams & Ormerod 2001; Liu et al. 2005). In this case, it was important that models could predict a landscape’s ability to support viable populations, one of the most important characteristics of which is the quantity of habitat (Gaona, Ferreras & Delibes 1998;
For each distribution model we calculated the classification threshold (accurate to 2 decimal places) which minimized the difference between the observed and predicted area of potential habitat in the training data set.

After applying this threshold, predicted presence/absence cell values were converted into connectivity surfaces. A location’s connectivity is a measure of the abundance of habitat in the surrounding area weighted by distance and the dispersal ability of the focal species. Connectivity values for each grid cell were calculated according to equation 2:

\[ S_i = \sum_{j \neq i} \exp(-\alpha \cdot d_{ij}) \]  

where \( i \) is the focal grid cell, \( j \) represents all potentially suitable grid cells (other than the focal cell) and \( d_{ij} \) is the Euclidean distance between the central points of cells; \( \alpha \) determines the dispersal ability of the study species by adjusting the negative exponential relationship between the frequency of dispersal and distance travelled (Hanski, Kuussaari & Nieminen 1994). The \( \alpha \)-value used here was 2. For \( \alpha = 2 \), 10% of dispersing individuals will move 1·2 km from the starting location; 1% will move 2·3 km. This value is supported by information on metapopulation systems of \( E. aurinia \) throughout the United Kingdom (Bulman et al. 2007), and in relation to other butterfly species (Cowley et al. 1999). Altering this value would have expanded or contracted areas of high connectivity, but the rank order of cell connectivity would not have been greatly affected, and hence would have only a limited impact on the locations that are identified as most important for conservation. There is potential for variation in migration parameters between landscapes (Thomas, Hill & Lewis 1998; Heino & Hanski 2001; Schtickzelle, Menechez & Baguette 2006), but integrated planning across large regions is vital in order to set appropriate conservation priorities (Sarkar et al. 2006). We therefore assumed a single \( \alpha \)-value when calculating connectivity for Wales but this may subsequently be refined during closer analysis of key landscapes.

**SPECIES-BASED MODEL CONSTRUCTION**

We constructed 1-ha resolution distribution models based on records of \( E. aurinia \) collected by volunteers between 1990 and 2002 (data supplied by CCW), using the same predictor variable data as for habitat-based models. \( E. aurinia \) records within the surveyed regions (Fig. 1) were used as ‘presences’ (193 records). There were insufficient recorded species absences to use in model construction and so sites that were included in the habitat surveys, but from which there was no \( E. aurinia \) record, were used as absences instead. \( E. aurinia \) is a popular species among volunteer surveyors, and so most of these sites are likely to have been visited by butterfly recorders, even if the ‘absences’ were not documented. These sites therefore constitute potential \( E. aurinia \) habitat that was not recorded as occupied between 1990 and 2002. In most cases these patches would not have been occupied at all, but some of them may have been occupied at some time within this period, given the metapopulation dynamics of this species (Warren 1994; Bulman et al. 2007).

**MODEL TESTING AND EXTRAPOLATION**

We used geographically distinct training and testing data sets in order to test the ability of models to generalize to novel areas (Vaughan & Ormerod 2005; Randin et al. 2006). The survey data were divided into eight regions (Fig. 1). Three of these were geographically discrete units (Gower Commons, Llyn Peninsula, Mynydd Mawr). The rest were part of a single large survey area, which we divided into five regions reflecting clusters of habitat presence data and similar prevalences (Glamorgan 1–5). For habitat- and species-based models, eight models were constructed. Each model was built using seven of the regions (‘training’ data) and tested against the eighth region (‘testing’ data).

The area under the receiver operating characteristic curve (AUC Fielding & Bell 1997) was calculated for the model predictions of the training and testing data sets (using the ‘ROCR’ package of R (Sing et al. 2005). Secondly, after classification thresholds had been applied and connectivity surfaces calculated, the observed and predicted connectivities were compared using (i) the total connectivity of the observed and predicted potential habitat in each region, and (ii) the cell-by-cell Spearman’s rank correlations between observed and predicted connectivity surfaces.

In order to test the ability of distribution models to identify areas of known high conservation value, we identified ‘priority’ landscapes within the testing regions. Priority landscapes consisted of adjacent grid cells with a connectivity value greater than 0·1 (Early & Thomas 2007). Their conservation value was assessed using the sum of their connectivity values, and then ranked. The results for both the observed and the predicted habitat distributions were compared.

Finally, all eight habitat-based models were extrapolated to predict connectivity surfaces for Wales.

**Results**

All eight habitat-based models were very similar (see Appendix SI in Supplementary material). In all but one model, marsh
Table 2. Testing habitat-based models. Area under the curve (AUC) scores are shown for model predictions of the regions on which they were built (‘training’ data set) and the region against which they were tested (‘testing’ data set). Thresholds are those calculated to give the most accurate prediction of the total amount of potential habitat in the training regions. Total connectivity values are the sums of the connectivity values of all grid cells in each area. Correlation coefficients compare observed and predicted connectivity within each grid cell using Spearman’s rank correlation.

<table>
<thead>
<tr>
<th>Testing region (not included in training data)</th>
<th>AUC score</th>
<th>Training region</th>
<th>Testing region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training data set</td>
<td>Testing data set</td>
<td>Threshold</td>
</tr>
<tr>
<td>Glamorgan 1</td>
<td>0.865</td>
<td>0.779</td>
<td>0.33</td>
</tr>
<tr>
<td>Glamorgan 2</td>
<td>0.852</td>
<td>0.898</td>
<td>0.33</td>
</tr>
<tr>
<td>Glamorgan 3</td>
<td>0.852</td>
<td>0.888</td>
<td>0.33</td>
</tr>
<tr>
<td>Glamorgan 4</td>
<td>0.847</td>
<td>0.829</td>
<td>0.37</td>
</tr>
<tr>
<td>Glamorgan 5</td>
<td>0.864</td>
<td>0.812</td>
<td>0.35</td>
</tr>
<tr>
<td>Gower Commons</td>
<td>0.817</td>
<td>0.885</td>
<td>0.39</td>
</tr>
<tr>
<td>Llyn Peninsula</td>
<td>0.847</td>
<td>0.792</td>
<td>0.37</td>
</tr>
<tr>
<td>Mynydd Mawr</td>
<td>0.850</td>
<td>0.859</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 3. Testing species-based resolution models. Area under the curve (AUC) scores are shown for model predictions of the regions on which they were built (‘training data set’) and the region against which they were tested (‘testing data set’). Thresholds are those calculated to give the most accurate prediction of the number of butterfly records in the training regions.

<table>
<thead>
<tr>
<th>Testing region (not included in training data)</th>
<th>AUC score</th>
<th>Training data set</th>
<th>Testing data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glamorgan 1</td>
<td>0.734</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>Glamorgan 2</td>
<td>0.742</td>
<td>0.682</td>
<td></td>
</tr>
<tr>
<td>Glamorgan 3</td>
<td>0.741</td>
<td>0.703</td>
<td></td>
</tr>
<tr>
<td>Glamorgan 4</td>
<td>0.717</td>
<td>0.715</td>
<td></td>
</tr>
<tr>
<td>Glamorgan 5</td>
<td>0.725</td>
<td>0.619</td>
<td></td>
</tr>
<tr>
<td>Gower Commons</td>
<td>0.741</td>
<td>0.656</td>
<td></td>
</tr>
<tr>
<td>Llyn Peninsula</td>
<td>0.742</td>
<td>0.760</td>
<td></td>
</tr>
<tr>
<td>Mynydd Mawr</td>
<td>0.734</td>
<td>0.677</td>
<td></td>
</tr>
</tbody>
</table>
However, the sizes and shapes of the observed high-priority landscapes are predicted sufficiently well by the habitat distribution models that they could be used to develop initial conservation plans and to target further survey work.

The results of extrapolating habitat-based models throughout Wales are shown in Fig. 4. The areas shown are the predicted high-priority areas, i.e. those with connectivity greater than 0.1. There are three areas where the eight habitat models show a high level of disagreement with one another (identified by arrows in Fig. 4c), none of which intersect with known E. aurinia populations (Fig. 4d).

To test the ability of habitat-based models to predict the distribution of E. aurinia itself, we compared the habitat predictions throughout Wales to the observed species locations between 1990 and 2002 (Fig. 4d). There were 458 cells (1 ha) containing E. aurinia records throughout Wales. These cells were predicted to contain potentially suitable habitat, as predicted by the habitat-based models, more often than in any of 1000 random picks of 458 cells (P < 0.0001 for all models).

We then compared the ability of habitat- and species-based models to predict the distribution of E. aurinia across Wales. E. aurinia locations were estimated by volunteers to the nearest 100 m, and even if accurate could have an error of up to 141 m attached to them. Additionally, butterflies may move short distances to find adult nectar sources, and thus a butterfly sighting within 200 m of suitable habitat is likely to be associated with that site. We therefore counted any 1-ha grid cell containing one or more E. aurinia records within 200 m of a predicted habitat grid cell as being predicted successfully by a model. For the habitat-based models, potential habitat grid cells across Wales were defined using their classification thresholds. For the species-based models, we calculated a classification threshold which allowed them to predict the same percentages of E. aurinia grid cells as the predictions of the habitat-based models, and again used a 200-m buffer zone when assessing the coincidence of predicted and observed presences. On average, over twice as many 1-ha cells were required using the species-based models than the habitat-based models (Table 4). This indicates that the habitat-based models identified the location of E. aurinia records considerably more accurately than did the species-based model.
Discussion

Identification of high-priority landscapes for *E. aurinia* presented a quite ‘typical’ conservation problem. Distribution data on the butterfly’s habitat distribution were potentially incomplete, requiring the application of distribution models. However, constructing distribution models was problematic because the metapopulation dynamics of the species mean that, at any one time, it is absent from much of the habitat that is important for its long-term persistence. These absences act as ‘false negatives’, and may reduce model discrimination power and ability to be extrapolated to new areas (Tyre et al. 2003). The alternative we adopted was to identify habitat requirements independently, and model the distribution of the habitat itself. Connectivity surfaces were then calculated using the predictions of these models, in order to measure landscape priority.

Use of distribution models in large-scale landscape prioritization

The eight models were similar in model composition and predictions. The three areas of high disagreement in the connectivity surfaces shown in Fig. 4c are caused by differences between models in the coefficients for neutral grassland. Because the butterfly itself has not been recorded from any of these three areas (Fig. 4d), they are unlikely to have high conservation priority. Other areas of intermediate disagreement are caused mainly by small differences between models in the predicted sizes of habitat patches and in the occurrence of very small habitat patches. The similarity in the response–environment relationships identified by each model is not surprising, given that seven of the eight landscapes are included within the training data for each model. It is reassuring, however, that the predictor coefficients (see
Table 4. Comparison of model ability to predict distribution of *Euphydryas aurinia* records

<table>
<thead>
<tr>
<th>Testing region (not included in training data)</th>
<th>Area &gt; threshold (ha)</th>
<th>% <em>E. aurinia</em> records within 200 m of predicted habitat</th>
<th>Area containing same % of <em>E. aurinia</em> records (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glamorgan 1</td>
<td>28 320</td>
<td>84</td>
<td>38 849</td>
</tr>
<tr>
<td>Glamorgan 2</td>
<td>39 705</td>
<td>90</td>
<td>109 516</td>
</tr>
<tr>
<td>Glamorgan 3</td>
<td>26 184</td>
<td>82</td>
<td>31 896</td>
</tr>
<tr>
<td>Glamorgan 4</td>
<td>25 293</td>
<td>84</td>
<td>42 974</td>
</tr>
<tr>
<td>Glamorgan 5</td>
<td>25 038</td>
<td>85</td>
<td>39 688</td>
</tr>
<tr>
<td>Gower Commons</td>
<td>30 996</td>
<td>88</td>
<td>39 998</td>
</tr>
<tr>
<td>Llyn Peninsula</td>
<td>39 732</td>
<td>86</td>
<td>44 477</td>
</tr>
<tr>
<td>Mynydd Mawr</td>
<td>34 765</td>
<td>88</td>
<td>219 684</td>
</tr>
<tr>
<td>Mean</td>
<td>31 254</td>
<td>86</td>
<td>70 885</td>
</tr>
</tbody>
</table>

†Area of Wales ~ 2 077 900 ha. †Including records of *E. aurinia* within this area plus 200 m buffer zone to ensure comparability with habitat model.

Supplementary material, Appendix SI) are realistic representations of the habitat requirements of this target species. Presumably this is because the variables selected for modelling are ‘direct predictors’ (i.e. are thought to correlate mechanistically with the response variables (Guisan & Zimmermann 2000; Augustin et al. 2001; Austin 2002; Maes et al. 2003; Randin et al. 2006).

Habitat-based models were able to predict patterns of habitat availability in both training and testing landscapes. Calculating connectivity surfaces from predicted habitat distributions allowed us to rank landscape conservation priority based on population dynamic principles. When used to predict the testing data, models appeared to capture these rankings reasonably well (Fig. 3). While candidate priority landscapes will typically require further study before implementing practical conservation, this habitat-based modelling approach represents an objective means of targeting future landscapes.

By building multiple models, the predicted priority landscapes for Wales need not rely on the single model which is apparently ‘best’ (Whittingham et al. 2006). Rather, the consensus and range of predictions can be taken into account when planning conservation action. We believe that the consensus connectivity surface shown in Fig. 4c represents the areas with the greatest capacity to support persistent *E. aurinia* populations, and thus the core areas for the conservation of this species. However, the minimum (Fig. 4a) and maximum (Fig. 4b) connectivity surfaces demonstrate the range of potential connectivity, depending on which part of the response–environment relationship was included in the training data. The highly connected areas in Fig. 4a are those in which potential for habitat management is in least dispute. The areas which are added or expanded upon in Fig. 4b could be interpreted more cautiously, and perhaps be targeted for future surveys to determine their conservation potential. Some of the areas in Fig. 4b which are expansions of those in Fig. 4a, could be managed as ‘buffer zones’ or ‘linkage areas’ between higher priority areas. Discrete areas which are added in Fig. 4b were identified as important by only a few models and may be genuine mistakes (as is likely for the three areas identified in Fig. 4c). Alternatively, they may be areas with lower conservation potential than the core conservation areas, but which could be targeted for local conservation planning.

A caveat to our results is that the habitat predictions produced here are the ‘potential’ distribution of the resources available, rather than maps of current habitat suitability. The realization of this potential will depend upon appropriate habitat management within the priority landscapes.

**HABITAT-BASED MODELLING AS AN ALTERNATIVE TO SPECIES-BASED MODELLING**

Species-based models did have reasonable discriminatory power when tested within the training regions, but their fit was consistently lower than for the habitat models. When extrapolated throughout Wales, species-based model predictions appeared to be considerably less accurate than habitat-based models. The habitat-based models identified the majority of known *E. aurinia* locations across Wales as being within 200 m of potential habitat (after a classification threshold was imposed on model predictions, Fig. 4d). For the species-based models to classify the same number of species records as linked to predicted habitat, the area which had to be classed as ‘suitable’ was more than twice as large as that classed as ‘potential’ habitat by habitat-based models (Table 4). It appears that, for *E. aurinia*, species-based models are less able to distinguish between suitable and unsuitable habitat.

The low discriminatory power of species-based models is likely to be due to the low-level of occupancy of suitable habitat, caused presumably by *E. aurinia*’s metapopulation dynamics. Furthermore, species location data alone provide little information on how wide an area surrounding the record is actually suitable habitat. Therefore, at high resolutions only a part of suitable sites will be included in the distribution training data.
In the habitat-based modelling approach, we do not imply that a species’ habitat is homogeneous (Ehrlich & Murphy 1987; Dennis, Shreeve & Van Dyck 2006), but rather that important characteristics of it can be modelled as a discrete entity (Debinski et al. 2002; Moisen & Frescino 2002). We believe that the use of distribution modelling techniques in this way will often be feasible using existing data. Complete data on the locations which a metapopulation species may occupy over time may require extremely long or repeated observation periods (Tyre et al. 2003; Welk 2004; Herben et al. 2006). In comparison, there is potentially a wealth of data available which could be used to implement our approach for other species. For example, data which could be used to define a suitable habitat classification are gathered frequently through studies of survivorship (Webb & Pullin 1996), physiology (Weiss, Murphy & White 1988), population dynamics (Thomas et al. 2001 and references therein) and interviews with experts (Petit et al. 2003). The locations of habitat with which to build distribution models are also collected regularly, for example through habitat surveys for regional conservation management or environmental impact assessments (Gipilin 1995).

When data are limited, a further alternative to our approach is to model at a coarse resolution, for which data are more likely to be generally available (e.g. Eyre et al. 2004). However, for E. aurinia in Wales, coarse-resolution approaches (25 ha and 100 ha) produced far less consistent models than the 1-ha approach (Early 2006). High-resolution models may often be required to identify the very specific and localized environmental requirements of rare species (Meggs et al. 2004). In addition, only a high spatial resolution approach can provide the precision needed to map conservation priority based on population dynamic principles; in this case achieved by the use of connectivity surfaces. Given the high level of ecological specificity of many animals and plants (often making habitat areas very small, even in undisturbed environments), and the limited data on their distributions in most parts of the world, we suggest that variants of the sort of approach we have used, only a high spatial resolution approach can provide additional data on the locations which a metapopulation species may occupy over time may require extremely long or repeated observation periods (Tyre et al. 2003; Welk 2004; Herben et al. 2006). In comparison, there is potentially a wealth of data available which could be used to implement our approach for other species. For example, data which could be used to define a suitable habitat classification are gathered frequently through studies of survivorship (Webb & Pullin 1996), physiology (Weiss, Murphy & White 1988), population dynamics (Thomas et al. 2001 and references therein) and interviews with experts (Petit et al. 2003). The locations of habitat with which to build distribution models are also collected regularly, for example through habitat surveys for regional conservation management or environmental impact assessments (Gipilin 1995).

When data are limited, a further alternative to our approach is to model at a coarse resolution, for which data are more likely to be generally available (e.g. Eyre et al. 2004). However, for E. aurinia in Wales, coarse-resolution approaches (25 ha and 100 ha) produced far less consistent models than the 1-ha approach (Early 2006). High-resolution models may often be required to identify the very specific and localized environmental requirements of rare species (Meggs et al. 2004). In addition, only a high spatial resolution approach can provide the precision needed to map conservation priority based on population dynamic principles; in this case achieved by the use of connectivity surfaces. Given the high level of ecological specificity of many animals and plants (often making habitat areas very small, even in undisturbed environments), and the limited data on their distributions in most parts of the world, we suggest that variants of the sort of approach we have adopted may be useful in many other situations.

In conclusion, distribution models can be used to inform large-scale conservation decisions when united with a metric by which to translate model predictions into conservation priority (e.g. connectivity). Our habitat-based models appeared to capture effectively the relationship between habitat suitability and the environment, thus generating robust predictions. Habitat-based models were considerably more effective at predicting both the habitat distribution of a novel area, and the locations where the butterfly is present throughout Wales, than models built with species data collected over a reasonably long time-period. Thus, at least in some systems, habitat can be a useful surrogate response variable for species location data.

Acknowledgements

Many thanks to C. McClean, A. Dumbrell, A. Fowles, B. Ledger and J. Chipperfield. Data were provided by UKPopNet and the Countryside Council for Wales. The work was funded by the Countryside Council for Wales and UKPopNet, including NERC grant R8-H12-01 and support from Natural England.

References


Received 12 April 2007; accepted 3 September 2007

Handling Editor: Jack Lennon
Supplementary material

The following supplementary material is available for this article.

Appendix S1. Habitat-based habitat distribution models.

This material is available as part of the online article from:
(This link will take you to the article abstract.)

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary material supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.