Ecological Indicators 62 (2016) 271–278

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Assessing species' habitat associations from occurrence records, standardised monitoring data and expert opinion: A test with British butterflies

J.W. Redhead[a,c,*], R. Fox[b], T. Brereton[b], T.H. Oliver[a,c]

[a] NERC Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire OX10 8BB, UK
[b] Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset BH20 5QP, UK
[c] School of Biological Sciences, Harborne Building, University of Reading, Reading, Berkshire RG6 6AS, UK

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 29 October 2015
Accepted 4 November 2015
Available online 2 December 2015

Keywords:
Spatial variation
Recording scheme
Citizen science
Latitudinal gradient
Biological indicators

ABSTRACT

Accurate knowledge of species' habitat associations is important for conservation planning and policy. Assessing habitat associations is a vital precursor to selecting appropriate indicator species for prioritising sites for conservation or assessing trends in habitat quality. However, much existing knowledge is based on qualitative expert opinion or local scale studies, and may not remain accurate across different spatial scales or geographic locations. Data from biological recording schemes have the potential to provide objective measures of habitat association, with the ability to account for spatial variation. We used data on 50 British butterfly species as a test case to investigate the correspondence of data-derived measures of habitat association with expert opinion, from two different butterfly recording schemes. One scheme collected large quantities of occurrence data (c. 3 million records) and the other, lower quantities of standardised monitoring data (c. 1400 sites). We used general linear mixed effects models to derive scores of association with broad-leaf woodland for both datasets and compared them with scores canvassed from experts.

Scores derived from occurrence and abundance data both showed strongly positive correlations with expert opinion. However, only for occurrence data did these fell within the range of correlations between experts. Data-derived scores showed regional spatial variation in the strength of butterfly associations with broad-leaf woodland, with a significant latitudinal trend in 26% of species. Sub-sampling of the data suggested a mean sample size of 5000 occurrence records per species to gain an accurate estimation of habitat association, although habitat specialists are likely to be readily detected using several hundred records. Occurrence data from recording schemes can thus provide easily obtained, objective, quantitative measures of habitat association.

1. Introduction

Associations between species and habitats are one of the basic principles of ecology (Aarts et al., 2013; Yapp, 1922). As habitat loss remains the primary cause of global biodiversity declines (Brooks et al., 2006; Thomas et al., 2004) identifying such associations accurately is important for conservation planning, policy and research. Where species are in decline, accurate information on habitat associations is required so that investigations into likely causes, and subsequent implementation of conservation efforts, can be targeted correctly. Likewise, if a particular habitat is undergoing change, well characterised associations enable predications to be made about which species are most likely to be affected. Accurate knowledge of associations is also vital to selecting appropriate indicator species for use in prioritising sites for conservation, monitoring environmental conditions or assessment of habitat quality (Carignan and Villard, 2002).

Although the habitat associations of some taxa are well characterised, most species are poorly studied. Even for well-studied taxa there may be limitations to our understanding of habitat associations at large spatial scales (Gregory and Baillie, 1998) as many studies are carried out at a local level in response to specific conservation issues (e.g. Knight and Arthington, 2008; Loeb et al., 2000; Rouquette and Thompson, 2005). As a result, information on wider

* Corresponding author at: NERC Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire OX10 8BB, UK. Tel.: +44 1491 692538; fax: +44 1491 692424.
E-mail addresses: johdhe@ceh.ac.uk (J.W. Redhead), rfox@butterfly-conservation.org (R. Fox), tbrereton@butterfly-conservation.org (T. Brereton), t.oliver@reading.ac.uk (T.H. Oliver).

http://dx.doi.org/10.1016/j.ecolind.2015.11.004
1470-160X/

scale habitat associations, including that which forms the foundations of much conservation policy, is often extrapolated from such studies or from qualitative descriptions based on expert opinion (Reif et al., 2010). This is potentially problematic, as both habitat associations and expert perceptions of them have been demonstrated to vary with location (O'Leary et al., 2009; Oliver et al., 2009), spatial scale (Mayor et al., 2009) and environmental change (Pateman et al., 2012). It is thus important to test existing knowledge on habitat associations against quantitative methods. These have the potential to operate at a range of spatial scales, and to take into account spatial or temporal variation. Such methods also have the potential to uncover cryptic requirements or previously unknown plasticities in habitat association.

National or international biological recording and monitoring schemes provide a valuable source of data for analysing large scale patterns in time and space (Bishop et al., 2013; Thomas, 2005). Large sample sizes and extensive spatial coverage make them well suited to use in detecting habitat associations. However, monitoring scheme data vary in quality and quantity, from simple occurrence data (i.e. georeferenced records of species' presence) to detailed demographic data from standardised protocols. Whilst datasets at all points along this spectrum have their value for specific applications, it is important to test which are most suitable for detecting habitat associations, especially as increasing levels of information come at a cost of time and effort in collection, and, consequently, in the number and spatial coverage of records (Bishop et al., 2013).

This study used two different butterfly recording scheme datasets – one comprising large quantities of occurrence data and the other, lower quantities of abundance data from a standardised monitoring scheme – alongside data on the extent of British broad-leaf woodland. Butterflies are a useful test case for determining habitat associations. They are frequently used as indicator species (Thomas, 2005) as their host plant specificity and temperature-dependent development and behaviour make them sensitive to environmental changes, whilst their short life cycles ensure that they respond quickly (Oliver et al., 2009; Pateman et al., 2012; Warren et al., 2001). In Britain, they are well recorded, giving sufficient data for analyses, and well-studied, such that expert opinions are likely to be well-founded and consistent and thus a good yardstick by which to measure the performance of data-derived measures of habitat association. We compared data-derived methods for calculating metrics of habitat association from the two butterfly datasets with expert opinion, including their ability to account for spatial variation in association, and assessed the applicability of these methods to other taxa for which data-derived methods might form the only means by which to assess species' habitat associations.

2. Methods

2.1. Species data

We obtained data on 50 butterfly species in Great Britain (GB) from two monitoring schemes – Butterflies for the New Millennium (BNM) and the UK Butterfly Monitoring Scheme (UKBMS). Species nomenclature follows Agassiz et al. (2013).

BNM is a national scheme which collates butterfly records (i.e. species occurrence at a location), with the aim of maintaining an up-to-date database of butterfly distributions (Asher et al., 2001). This study included only BNM records with spatial resolution of 1 km × 1 km Ordnance Survey grid cell or finer. Duplicate records of the same species in the same cell were removed, resulting in a dataset of approximately 3 million butterfly occurrence records. The study used records from 1990 to 2010, to decrease the likely effect of changes over time in woodland extent or habitat association on the results.

The UKBMS differs from BNM in aiming to monitor population trends through a standardised survey method involving weekly visits between April and September (Pollard and Yates, 1993). Although this allows calculation of abundance throughout each survey year and thus analysis of population trends and phenology, it is relatively labour intensive and there are records from far fewer sites than in BNM (data from 1433 sites were included in our analysis).

Although the spatial scale of GB reflects an artificial imposition onto an ecologically meaningful hierarchy of scales, being neither the full range of a species nor of an individual butterfly, it reflects the scale at which national policy for particular species and habitats tends to be formulated (Roy et al., 2007) and at which biological recording schemes tend to be coordinated.

2.2. Habitat data

Broad-leaf woodland data were obtained from the Land Cover Map 2007 (LCM2007, Morton et al., 2011). We chose this habitat because it is well characterised in LCM2007 and includes various habitats which are prominent in UK planning and policy (e.g. ancient broad-leaf woodland, DEFRA, 2011). The proportion of broad-leaf woodland was calculated for every 1 km grid cell in mainland GB and for a 500 m radius around each UKBMS site centroid, giving a consistent scale of analysis between datasets. This scale also reflects the relatively coarse resolution at which much large scale habitat data is readily available. These analyses were performed in ArcGIS (v 9.3.1 © 2010 ESRI, Redlands, California).

2.3. Scoring habitat association from biological recording data

Analyses were performed independently. To distinguish 'genuine' absences for each species from a 1 km cell in the BNM data, as opposed to pseudoabsence generated by lack of recorders or non-detection (Prendergast et al., 1993), we applied a threshold of species detection. Cells in which more than five butterfly species were recorded (i.e. c. 10% of the total UK species pool, following Hickling et al. (2006)) but which lacked a record of the species in question were assumed to be genuine absences, whilst others were removed from all further analyses. We did not use more analytically complex methods of accounting for recorder effort (e.g. Hill, 2012; Isaac et al., 2014; Mason et al., 2015) because UK butterflies are generally well recorded, not particularly speciose, and have several ubiquitous species which are well recorded across the entire of the country. Therefore, although there is a latitudinal gradient in butterfly species richness in the UK, the 5 species threshold is met by a relatively consistent proportion of cells per region supplementary material, Table S2). Whilst butterfly species have been shown to vary in detectability (Isaac et al., 2011) there is little evidence for a systematic bias whereby the detectability of individuals varies with woodland area and where this relationship varies between species, which would be the only situation in which detectability would automatically influence relative habitat association scores. To account for potential variation in species' habitat associations across GB, data were analysed on a regional basis, splitting the dataset into 100 km by 100 km cells (from here on referred to as a 100 km region). Regions where a species had less than 30 of each of presence and 'genuine' absence records were unlikely to provide robust estimates and were excluded. We also limited analyses to species that were recorded on a minimum of ten UKBMS sites.

General linear mixed effects models (GLMM) were used to model the relationships between habitat and butterflies, using the *lme4* (Bates et al., 2013) package in R (R Core Team, 2013). For the BNM data, we fitted species presence/absence to proportion of broad-leaf woodland cover in the 1 km grid cell, with a