Biological Conservation 192 (2015) 65–73

Contents lists available at ScienceDirect

Biological Conservation

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

Using citizen science data for conservation planning: Methods for quality control and downscaling for use in stochastic patch occupancy modelling

Laura J. Graham [a,b,*], Roy H. Haines-Young [a], Richard Field [a]

[a] School of Geography, University of Nottingham, University Park, Nottingham, UK
[b] Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form 7 August 2015
Accepted 4 September 2015
Available online 22 September 2015

Keywords:
Citizen science
Data quality
Downscaling
Spatial bias
Urban conservation

ABSTRACT

The Incidence Function Model (IFM) has been put forward as a tool for assessing conservation plans. A key benefit of the IFM is low data requirements: widely available species occurrence data and information about land cover. Citizen science is a promising source of such data; however, to use these data in the IFM there are typically two problems. First, the spatial resolution is too coarse, but existing approaches to downscaling species data tend not to extend to patch level (as required by the IFM). Second, widely available citizen science data typically report species' presences only. We devise ten different downscaling methods based on theoretical ecological relationships (the species–area relationship and the distance decay of similarity), and test them against each other. The better performing downscaling methods were based on patch area, rather than distance from other occupied patches. These methods allow data at a coarse resolution to be used in the IFM for comparing conservation management and development plans. Further field testing is required to establish the degree to which results of these new methods can be treated as definitive spatially-explicit predictions. To address the issue of false absences, we present a method to estimate the probability that all species have been listed (and thus that a species' absence from the list represents a true absence), using the species-accumulation curve. This measure of confidence in absence helps both to objectively identify a habitat network for fitting the IFM, and to target areas for further species recording.

1. Introduction

Stochastic patch occupancy models can be useful tools for incorporating biodiversity conservation into city planning because they allow for spatially explicit analysis of species' persistence under habitat fragmentation (Hanski, 1994; Opdam et al., 2002, 2003; Van Teeffelen et al., 2012). Species occurrence data at large spatial and temporal extents are necessary for both biodiversity planning (Williams et al., 2002) and for fitting stochastic patch occupancy models (Hanski, 1999; Opdam et al., 2003; Etienne et al., 2004). The Incidence Function Model (IFM) has been identified as particularly suitable for practical biodiversity planning (Lindenmayer et al., 1999; Graham et al., in press), in part as a result of its low data requirements: widely available species occurrence data can be used (Hanski, 1999; Etienne et al., 2004). Most studies tend to employ the IFM in a single-species approach, where the patch occupancies have been specifically surveyed for the purpose (e.g. Bulman et al., 2007; MacPherson and Bright, 2011; Heard et al., 2013; Dolrenry et al., 2014). For the IFM to be useful for biodiversity assessment within a conservation or planning framework, multiple indicator species need to be studied. However, to collect occupancy data for a suite of species is costly in terms of time and resources and so other strategies are needed. Our contribution is to provide new strategies to address this lack of occupancy data.

Volunteer biological recording, or more broadly citizen science, is a useful source of data for ecological and conservation research over a large spatial extent (Silvertown, 2009; Devictor et al., 2010; Dickinson et al., 2010, 2012; Tulloch et al., 2013; Graham et al., 2014). These kinds of data are also regularly used for biodiversity planning within UK local authorities (Lott et al., 2006). It allows large quantities of occurrence data to be collected at larger spatial and temporal extents than would be feasible through individual field studies. Species-level data are available from local recording schemes, as well as from large repositories, examples of which are Global Biodiversity Information Facility globally (Global Biodiversity Information Facility, 2014) and National Biodiversity Network (NBN) Gateway in the UK (National Biodiversity Network, 2014). There are, however, some problems with volunteer-collected data. There are concerns about the quality of data collected by non-specialists (see Bird et al., 2014; Isaac et al., 2014 for discussions of these issues and some of the potential solutions). Specific to the IFM, there are two prevalent issues in data available from major citizen science schemes. First, the data are typically available at grid-square level (for example the finest resolution of data available on the NBN Gateway

* Corresponding author at: Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA

E-mail address: laurajaneegraham@gmail.com (L.J. Graham).

http://dx.doi.org/10.1016/j.biocon.2015.09.002

is 100 × 100 m, but a greater coverage of data is available at the 2 km resolution), whereas the IFM requires information about patch-level occupancies (Hanski, 1999). Although some patches may cover a 100 × 100 m grid cell, in a highly fragmented landscape such as an urban or heavily managed landscape, the patches are likely to be smaller than this. Additionally, even if the sizes match, the grid cell boundaries are unlikely to be coincident with the patch boundaries. Secondly, the data tend to be presented as species lists, which only give information about species' presences. In a study by Moilanen (2002), it was found that false absences can bias parameter estimates in all components of the model; therefore, the higher the confidence in true absence, the better fitting the model will be (but see Kéry et al., 2010). If volunteer-collected data are to be useful for the IFM, or stochastic patch occupancy models more widely, methods are needed for downscaling these data to patch level, and for determining confidence in species' absences. Here we present methods to address both of these issues.

Current approaches to downscaling atlas data for species tend to fall into three categories: expert opinion, empirical models and spatial processes (Araújo et al., 2005; Keil et al., 2013). The expert opinion approach typically involves matching species to suitable land-cover classes. For a wide range of species, however, the species–habitat relationship is not well known, and so this method can only be applied to well-studied species (Araújo et al., 2005). This approach also operates on the assumption that any suitable habitat is occupied by the species, which is ecologically unrealistic because species range filling is generally discontinuous (Rapoport, 1982). The empirical approach uses environmental variables such as climate, land-cover classes and normalised vegetation difference indices to predict species' occurrences (see Araújo et al., 2005 for an example using general additive modelling and Keil et al., 2013 for one using hierarchical Bayesian modelling). These methods are particularly appropriate for broad-scale species mapping, for example national and continental studies (Stockwell and Peterson, 2002). The spatial-processes approach divides coarse grid cells into finer grid cells and uses statistical point-and-cluster processes to randomly select cells at a fine grain. The environmental attributes from these finer grid cells are used as predictors for species' presences and absences. These methods assume that all fine-grain grid cells within a coarse-grain cell of known occupancy contain suitable habitat. To overcome this problem, Niamir et al. (2011) proposed a method which combines expert knowledge and point sampling.

The empirical and spatial-processes approaches to downscaling species atlas data use environmental variables as predictors, drawing from species' distribution modelling. The extent and grain of interest for a city-level biodiversity plan tend to be much smaller than in studies taking a species' distribution modelling approach to downscaling, and the environmental gradients sampled therefore much narrower but with greater habitat heterogeneity. With their very limited variation, environmental factors such as climate are not useful for predicting species' occurrence at smaller extents. Instead, land cover tends to be the most reliable predictor, and thus the empirical and spatial-processes approaches collapse to species–habitat associations at the city level and individual patch characteristics are likely to become important. The method we outline below applies a combination of expert knowledge (through literature review) and spatial factors. The method involves attributing species' presence to a suitable habitat patch based on its spatial characteristics and known ecological patterns (species–area relationships and the distance decay of community similarity).

To return to the second issue with citizen science data – that they tend to report presence only, but the IFM parameters are sensitive to false absences – we show how this can be circumvented. The IFM parameters estimated for a species can be applied to a different patch network (Hanski et al., 1996) or those estimated on a contiguous subset of patches can be applied to the wider landscape (Bulman et al., 2007). If a core area can be identified within the landscape, with a high confidence in the species' absences, parameters can be estimated using the data from this subset. Species-accumulation curves are widely used to estimate species richness in sampled areas (e.g. Soberón and Llorente, 1993; Colwell and Coddington, 1994). This method has also been adapted to give a measure of how well an area has been surveyed (Hortal et al., 2004). Here, we used species accumulation curves to estimate confidence in true absence, and therefore identify subsets of the landscape for use in parameter estimation.

We aim to investigate the extent to which citizen science data are useful as inputs to the IFM. Firstly, we identified well-sampled grid cells within the landscape which can be used to parameterise the IFM. Secondly, we tested several downscaling methods based on spatial characteristics of the landscape and known ecological patterns to convert the species data to an appropriate resolution for the IFM. To achieve our aim, we use the study area of the city of Nottingham, UK and apply the methods to indicator species from the bird, herptile and mammal groups.

2. Methods

2.1. Study area

The Nottingham City unitary authority was used as a case-study area, with a 2 km buffer around its boundary to allow for some effect of dispersal from outside. Nottingham is located in the East Midlands, UK and represents a typical large-to-medium sized urban area in the UK. The unitary authority boundary was chosen as this is the level at which planning decisions are generally made. The location of the study site and a breakdown of the Land Cover Map 2007 classes (Morton et al., 2011) is given in Appendix A (Fig. A1, Table A1) with details for Nottingham, four nearby cities and the aggregate of ten similar-sized UK cities for comparison. This indicates that Nottingham is broadly representative of similarly sized UK cities.

2.2. Citizen science species data

Data for bird species were provided by Nottinghamshire Birdwatchers. These data comprised 12,110 records of 24 species in 44 2 km grid cells for the years 1998–2011. Bat species data were provided by Nottinghamshire Bat Group and further records were downloaded from NBN Gateway. The combined bat datasets, once duplicates had been removed, contained 421 records for 10 species in 109 1 km grid cells from 1983 to 2013. Amphibian and reptile data were downloaded from NBN gateway. There were a total of 1116 records for 11 species in 56 2 km grid cells for the period 1984–2012. All downloads from the NBN Gateway were performed using the R package 'rnbn' (Ball and August, 2013). The full list of data providers is supplied in Supplementary Materials, Appendix A (Table A3).

2.3. Species–habitat associations and dispersal

It is common practise to use indicator species in biodiversity assessments (Caro and O'Doherty, 2013) because constraints on time, funding and taxonomic knowledge make collection of data on all species unfeasible (Blair, 1999; Margules et al., 2002). We selected indicator species for modelling with the IFM where sufficient data and information about habitat requirements and dispersal were available. We ensured that species with a range of habitat specialisms and dispersal abilities were chosen, to maximise the species' validity as indicators.

The bird species chosen for modelling with the IFM included five generalists (*Turdus merula*, *Prunella modularis*, *Carduelis carduelis*, *Carduelis chloris* and *Muscicapa striata*), three farmland specialists (*Emberiza calandra*, *Passer montanus* and *Emberiza citrinella*) and four woodland specialists (*Sylvia atricapilla*, *Dendrocopos major*, *Garrulus glandarius* and *Poecile palustris*). *E. citrinella* also uses heathland. The amphibian species selected were *Rana temporaria* and *Bufo bufo*. Common names for all species are given in Table 1. The species chosen for modelling were those which were from well-sampled groups and