Are ecologists conducting research at the optimal scale?

Heather Bird Jackson* and Lenore Fahrig

Geomatics and Landscape Ecology Laboratory, Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6

ABSTRACT

Aim The spatial extent (scale) at which landscape attributes are measured has a strong impact on inferred species–landscape relationships. Consequently, researchers commonly measure landscape variables at multiple scales to select one scale (the 'scale of effect') that yields the strongest species–landscape relationship. Scales of effect observed in multiscale studies may not be true scales of effect if scales are arbitrarily selected and/or are too narrow in range. Miscalculation of the scale of effect may explain why the theoretical relationship between scale of effect and species traits, e.g. dispersal distance, is not empirically well supported.

Location World-wide.

Methods Using data from 583 species in 71 studies we conducted a quantitative review of multiscale studies to evaluate whether research has been conducted at the true scale of effect.

Results Multiple lines of evidence indicated that multiscale studies are often conducted at suboptimal scales. We did not find convincing evidence of a relationship between observed scale of effect and any of 29 species traits. Instead, observed scales of effect were strongly positively predicted by the smallest and largest scales evaluated by researchers. Only 29% of studies reported biological reasons for the scales evaluated. Scales tended to be narrow in range (the mean range is 0.9 orders of magnitude) and few (the mean number of scales evaluated is four). Many species (44%) had observed scales of effect equal to the smallest or largest scale evaluated, suggesting a better scale was outside that range. Increasing the range of scales evaluated decreased the proportion of species with scales of effect equal to the smallest or largest scale evaluated.

Main conclusions To ensure that species–landscape relationships are well estimated, we recommend that the scales at which landscape variables are measured range widely, from the size of a single territory to well above the average dispersal distance.

*Correspondence: Heather Bird Jackson, Geomatics and Landscape Ecology Laboratory, Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6.
E-mail: hjacks15@utk.edu
Present address: Heather Bird Jackson, Department of Ecology and Evolutionary Biology, University of Tennessee, 569 Dabney Hall, Knoxville, TN 37996, USA.*

INTRODUCTION

The scale at which ecological research is conducted can influence study outcomes (Wiens, 1989; Levin, 1992; Horne & Schneider, 1995; Wheatley & Johnson, 2009). At the landscape level, the spatial extent (hereafter scale) at which landscape structure is measured can affect inferred species–landscape relationships (Holland *et al.*, 2004; de Knegt *et al.*, 2010). For example, in a study relating percentage forest cover to the abundance of 12 wood-boring beetle species, Holland *et al.* (2004) found that, depending on the scale at which forest cover was measured (from 20 to 2000 m radius), the correlation between forest cover and beetle abundance ranged from strongly positive to negligible. The implication for researchers is that important species–landscape relationships can be missed if landscape structure is not measured at the scale at which it has its strongest
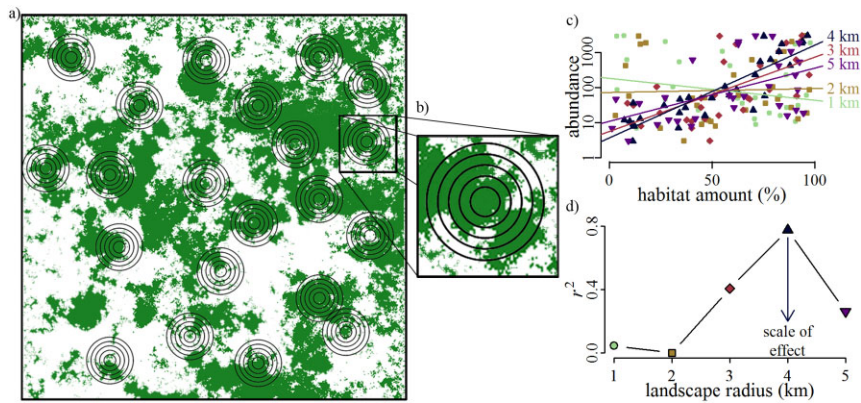
Figure 1 Design of a hypothetical focal site multiscale study like those reviewed in this paper. (a) The *ecological response*, e.g. the abundance of a species, is sampled in *focal sites* (not necessarily distinct patches) across the region of interest. (b) Landscape structure is measured at multiple spatial extents centered on the focal site, each extent including landscape at smaller extents. *Landscape structure* is any measure of landscape composition or configuration (McGarigal *et al.*, 2012). In this example, percentage habitat amount is measured, where habitat is represented in dark green. (c) The relationship between the ecological response and landscape structure is evaluated for each spatial extent. (d) The *scale of effect* is the spatial extent at which landscape structure best predicts the response. The scale of effect is also referred to in the literature as the 'intrinsic scale', the 'characteristic scale' or the 'scale of response'. In this example, the scale of effect is 4 km, where the slope between abundance and landscape structure is steepest (plot c) and the r^2 is highest (plot d). Since spatial scales are nested, landscape metrics are correlated between adjacent scales. Therefore if a landscape variable influences the ecological response, the strength of the effect (r^2 in plot d) increases to the scale of effect (4 km) and then decreases gradually. Note that if habitat amount was only measured at 1 km, one would incorrectly infer a weak negative relationship between abundance and habitat amount.

effect, hereafter referred to as the scale of effect. In this study we determine whether researchers have been measuring landscape structure at the optimal scale (i.e. the scale of effect). From our findings we provide recommendations for improving the success of future species–landscape research.

Usually the scale of effect is unknown to the researcher in advance of the study. To identify the scale of effect, and therefore maximize their ability to detect a species–landscape relationship, researchers can compare the effect of landscape structure on the species response at multiple scales. This is done using the focal site multiscale study design (Brennan *et al.*, 2002) (illustration and definitions in Fig. 1). Briefly, an ecological response (such as population abundance) is measured at multiple focal sites in a region (Fig. 1a), and landscape structure (such as percentage habitat amount) is measured at multiple spatial extents surrounding each focal site (Fig. 1b). The strength of the relationship between landscape structure and an ecological response is evaluated by considering the size of the correlation coefficient, the regression slope or the information value (*sensu* Burnham *et al.*, 2011) of models in which a measure of landscape structure is the predictor and an ecological measure is the response (Fig. 1c). The scale of effect is therefore the spatial extent at which landscape structure has the highest correlation coefficient, the steepest regression slope, or the lowest value of the Akaike information criterion (AIC) or Bayesian information criterion (BIC) when predicting an ecological response (Fig. 1d).

The scale of effect observed in multiscale studies may not equal the true scale of effect if scales evaluated are: (1) too small (Fig. 2a), (2) too large (Fig. 2b), (3) too narrow in range (Fig. 2c), and/or (4) too few (Fig. 2d). It is therefore recommended that researchers base their selection of scales on species traits of the organism under investigation, specifically territory size/home range of a species and/or its dispersal distance (Brennan *et al.*, 2002).

Theoretical studies support the idea that species traits influence the scale of effect. Using an individual-based multigenerational model of movement, we found strong variation in scale of effect among different simulated species (Jackson & Fahrig, 2012). Average dispersal distance had a strong positive influence on scale of effect, a result that makes intuitive sense and is consistent with the fact that among real species individual use of space, i.e. home ranges in mammals (Bowman *et al.*, 2002) and territory sizes in birds (Bowman, 2003), is also positively correlated with dispersal distance. An expected positive relationship between dispersal distance and scale of effect was corroborated in a simulation study by Ricci *et al.* (2013). Our simulation produced the novel prediction that reproductive rate should have a negative impact on the scale of effect, an idea which has yet to be empirically tested. Although not included in recent simulations, body size and trophic level (herbivores < carnivores) may be positively related to scale of effect due to their correlations with territory size and dispersal distance (Jenkins *et al.*, 2007; Hendriks *et al.*, 2009).

Empirical support for associations between species traits and scale of effect, however, is mixed and largely inconclusive (see Appendix S1 in Supporting Information for a brief review). In a meta-analysis of bird studies, Thornton & Fletcher (2014) found a positive relationship between body size and scale of effect within studies, but across studies the average scale of effect for a
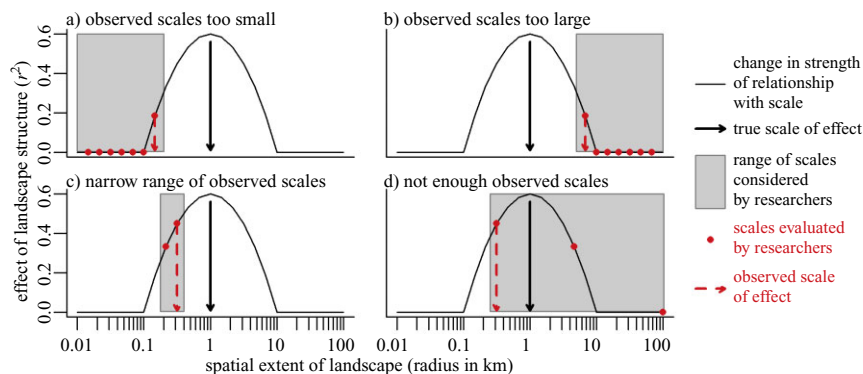
Figure 2 Illustration of ways in which the observed scale of effect can be different from the true scale of effect, even when the influence of landscape structure is evaluated at multiple spatial extents. If landscape scales evaluated are (a) too small, (b) too large, (c) too narrow, or (d) if there are too few scales evaluated, the observed scale of effect (indicated by a dotted arrow) can be different from the true scale of effect (indicated by a continuous arrow).

given body size varied by over three orders of magnitude. Most studies of the relationship between scale of effect and body size have been inconclusive (Appendix S1). In addition, no support for a positive relationship between various correlates of mobility (e.g. home range, dispersal distance) and scale of effect has been found, and there is only modest support for a smattering of other hypotheses (Appendix S1).

Here, we test the hypothesis that the general lack of support for predicted relationships between the scale of effect and species traits is due to generally poor estimation of the scale of effect. We conducted a quantitative, systematic review of focal site multiscale studies to determine whether the ranges of spatial extents evaluated by researchers have included the true scale of effect. Our first step was to compare the relative impact of scales evaluated (largest and/or smallest) versus species traits on the observed scales of effect in multiscale studies. If the true scale of effect is often outside the range of scales evaluated, we expected that the radii of the smallest and largest scales would strongly, positively predict the observed scale of effect, while species traits should have relatively little influence on the observed scale of effect. In other words, as the smallest and/or largest scale evaluated in a study increased, the observed scale of effect would increase, regardless of any species traits. On the other hand, if researchers were evaluating a sufficient number and range of scales to determine the true scale of effect, we expected to find that, consistent with theory, the scale of effect: (1) varied among taxonomic groups, (2) increased with body size, (3) increased with mobility, (4) decreased with reproductive rate, and (5) was greater for carnivores than herbivores.

If the scales selected have included the true scale of effect, we further expected to find that: (1) observed scales of effect were rarely the smallest or largest scale measured in a study; (2) most researchers used species traits to justify the range of scales evaluated in their study; (3) an implicit expectation of a relationship between scale of effect and species traits was present in researchers' selection of scales such that the scales selected were larger for bigger and/or winged species; (4) scales of effect observed for the same species in different studies were more similar than expected by chance; (5) differences in observed scales of effect for the same species in different studies were unrelated to differences in the range of scales evaluated; and (6) selected scales included scales expected to be associated with scale of effect (e.g.

home range or average dispersal distance, or as our previous simulations predict, four to nine times the average dispersal distance; Jackson & Fahrig, 2012).

Although some studies have evaluated the relationship between species traits and scale of effect (Appendix S1), our review is the first to determine how well researchers have been estimating the scale of effect and whether apparent relationships between scale of effect and species traits are compromised by weak estimation of the scale of effect. This question is important because if the scales at which landscape structure is measured are not generally equal to the scale at which landscape structure has its greatest effect (i.e. the true scale of effect), studies will miss or underestimate species–landscape relationships.

METHODS

Study selection

We conducted a thorough search of the literature for focal site multiscale studies that reported the relationship between abundance and landscape structure. We searched ISI Web of Knowledge (accessed 12 August 2011) using the following keywords: ((('spatial scale*' OR 'spatial extent*' OR 'landscape size' OR 'multi-scale' OR 'landscape area' OR 'buffers' OR 'focal patch*' OR 'focal point*') AND ('surrounding landscape*' OR 'landscape context' OR 'habitat loss' OR 'habitat fragmentation' OR 'habitat amount' OR 'landscape structure' OR 'landscape composition' OR 'habitat area') AND (abundance OR occupancy OR incidence)). Data from theses were also used. Criteria for inclusion were as follows.

1. The study measured abundance, density or occurrence of a taxonomic group (pooled taxa or single species) in focal sites. Studies which measured occurrence were included because occurrence is a categorical measure of abundance (zero or greater than zero). Furthermore, density was included because it is equivalent to abundance if focal sites are equally sized (see point 6).

2. Landscape structure was measured within two or more scales surrounding the focal sampling areas so that a 'best' scale (of those evaluated) could be determined.

3. Landscape structure was measured within nested scales of absolute size (e.g. $1 < 2 < 3$ km) as opposed to nested hierarchi-

cal levels (e.g. patch < landscape < region) which can have different absolute sizes.

4. The same landscape attributes were measured within every (or most) scale(s), ensuring that metrics were comparable across scales.

5. Landscape structure was measured within a set spatial extent as opposed to distance measures (e.g. distance to edge).

6. Focal site area was controlled either physically (e.g. using equally sized sample areas) or statistically (e.g. by including focal patch size in the analyses), ensuring that population sampling was not confounded with the amount of habitat in the landscape.

7. Spatial grain was the same for all scales in a study so that variation in grain was not confounded with variation in extent, because grain can influence the observed scale of effect (Lechner *et al.*, 2012). For example, two large datasets in Hostetler & Holling (2000) were excluded because the grain at which landscape was measured increased with spatial extent.

When a study included species–landscape relationships for multiple species, each species was included as an independent data point. If a statistical model of a species–landscape relationship included landscape predictors from multiple scales, the average of these scales was recorded as the observed scale of effect (we removed averaged data points for a more selective analysis; see below). If species–landscape relationships were reported for the same species for multiple time periods or multiple habitat types within the same study the relationship based on the largest number of landscapes was retained. Major taxonomic groups with few samples (< 10 species, which excluded all fish, viruses, and fungi) were excluded.

Data extraction

For each species, we collected the scale of effect as reported by authors or from tables or graphs in the study. The scale of effect was the scale at which landscape structure predicted abundance with the lowest AIC value, the highest r^2, or the highest correlation coefficient (r). We rated the evidence for scale sensitivity of landscape predictors by comparing the relationships between landscape structure and abundance at the strongest scale (the observed scale of effect) and the weakest scale in the study. We considered the analysis to be scale sensitive if the observed scale of effect improved model fit relative to the weakest scale ($\Delta AIC > 2$, $\Delta r^2 > 0.01$, $\Delta r > 0.1$) or if the observed scale of effect was significantly better than another scale in a significance test performed by authors.

We recorded aspects of study design, including the radius of the smallest scale (if square GIS buffers were used, we recorded half the length of a side), the radius of the largest scale, the authors' justification for the range of scales, independence of replicate landscapes (yes/no; scales were non-overlapping, or if overlapping, non-independence was statistically accounted for in models), the number of replicate landscapes, habitat type in focal sites (forest/open/water), climate (polar/temperate/tropical) and season (breeding/non-breeding).

From published papers and species guides we collected information on species traits that fall into five categories: (1) taxonomic group (major taxonomic groups, order, family, genus); (2) body size (body mass, in g, and body length, in mm); (3) mobility [natal dispersal distance, in km, and other more commonly available data that are associated with dispersal distance, namely breeding home range, in ha, and breeding territory size, in ha (Bowman *et al.*, 2002; Bowman, 2003), and other coarser measures of mobility: migrant/resident; presence of wings]; (4) reproductive rate (mean clutch size for amphibians and birds; mean litter size for bats and ground mammals; or, if available, mean clutch/litter size × the maximum number of broods per season); and (5) trophic level (herbivore/omnivore/carnivore). Values were averaged over the sexes if separate data for sexes were provided. Collected data and sources are available in Appendices S2 & S3.

Statistical analysis

Are observed scales of effect determined more by scale selection than species traits?

We analysed the relative importance of study design (including scale selection) and species traits in two steps (an example is given in Appendix S4): (1) for each of 29 species traits, the best study design model was selected from five models that each included a different study design variable and the same species trait as fixed effects; (2) the evidence supporting the presence of the study design and species trait variables in the best model was calculated. The five study design variables compared were: (1) the radius of the smallest scale evaluated ($\log_{10}$ transformed), (2) the radius of the largest scale evaluated ($\log_{10}$ transformed), (3) the type of habitat in the focal site (forest, open or wetland), (4) climate (temperate, tropical or polar), and (5) response (occurrence is predicted to lead to a larger scale of effect than abundance; Jackson & Fahrig, 2014). For each model the response was scale of effect ($\log_{10}$ transformed for normality), the two fixed effects were a study design variable and a species trait, and study was a random effect (analyses conducted with package 'lme4'; Bates *et al.*, 2013). To ensure variables were meaningful, we removed categorical variables if they did not have at least two levels for a given dataset, where levels with fewer than five samples were eliminated. If the best two study-design variables (i.e. the two variables in models with the lowest AICc scores) were informative when combined (i.e. the AICc score became lower) and were not strongly correlated (Pearson's $r \leq 0.6$), they were both included in the best model.

To summarize the evidence that each fixed effect belonged in the true model, we calculated an evidence ratio (ER; Burnham *et al.*, 2011) comparing the likelihood of the full model and the model in which each predictor was removed. We report model estimates for the full model with confidence intervals estimated using 1000 Markov chain Monte Carlo simulations (package 'languageR'; Baayen, 2011). All analyses were conducted with R 3.0.1 (R Core Team, 2013).

We reran bird and invertebrate analyses using a subset of the data for which we were most confident in the estimated scale of effect (there were insufficient samples for other taxonomic groups to be evaluated). Specifically, we selected species for which: (1) the scale of effect was not averaged across landscape predictors measured at multiple scales (530/583 species), (2) the analysis was clearly scale sensitive (see above; 322/352 species for which this could be ascertained), and (c) the observed scale of effect was neither the smallest (460/583 species) nor the largest (449/583 species) scale in the study. These restrictions left us with 63 birds from 11 studies, 78 invertebrates from 18 studies and species trait data sufficient for a total of 12 analyses.

We addressed constraints in our data in the following ways. First, because species trait data differed among major taxonomic groups we analysed major taxonomic groups (amphibians, bats, birds, invertebrates and ground mammals) in separate analyses, with the exception of one analysis in which major taxonomic group was the predictor. Secondly, we considered the effects of order and family on scale of effect because species traits are probably phylogenetically dependent: major taxonomic groups, orders and families are likely to have similar body sizes, mobility, reproductive rates and, if it truly is a species trait, scales of effect due to common ancestry which could lead to apparent (but not causal) relationships between species traits and scale of effect. Thirdly, we evaluated each species trait separately because species trait data were incomplete. A full model comparing all or even many species traits would have had few samples. Fourthly, many studies had data for only one species, which meant that when mixed models were used with study as random effect, the presence of many so-called singletons could render model estimates less accurate (Bell *et al.*, 2008), particularly estimates for random-effect level variables (i.e. study design variables in our analyses). The accuracy of these estimates must therefore be treated with caution. Fifthly, many species were used in multiple studies, presenting an opportunity for pseudo-replication. On the other hand, these repeated species provided a unique opportunity to consider the extent to which scale of effect is consistent within taxa. We retained these species as separate data points, but evaluated the variation in scales of effect within species directly (see below). Finally, there were numerous variations in study design which were addressed in three ways: (1) we compared the information value of those aspects of study design for which we had enough data; (2) we included study as a random effect in all models to account for unmeasured differences among studies; and (3) we reported the frequency and discussed the effects of aspects of study design which we expected to obscure the scale of effect, namely non-independence of landscapes, small numbers of landscapes, non-breeding season observations, grouped species, averaging scales for different landscape predictors and different average amounts of habitat across regions.

Are observed scales of effect frequently equal to the smallest or largest scale in a study?

A scale of effect that is equal to the smallest or largest scale is suspect because it may indicate that the true scale of effect is outside the range of those measured in the study (Fig. 2). We expected that a study with a wide range of scales and/or a greater density of scales was likely to have fewer species whose scales of effect were equal to the smallest or largest scale evaluated. We tested these predictions by using the range of scales $[\log_{10}(\text{largest radius}) - \log_{10}(\text{smallest radius})]$ and density of scales $[\log_{10}(\text{number of scales per km radius})]$ to predict the proportion of species with a scale of effect equal to the smallest and largest scales in two separate logistic regressions where study was the unit of replication.

Do authors provide biological justification for their selection of scales?

To evaluate whether authors based their choice of scales on species traits, we grouped the justifications described in methods sections into six categories that loosely belonged to two groups: non-biological and biological. Non-biological justifications included: (1) no justification, (2) physical constraints (i.e. landscape data were not available above or below a scale and/or larger scales would have include too much overlap of landscapes), (3) exploration (i.e. authors explicitly intended to search for the appropriate scale), and (4) precedent (i.e. previous studies of similar taxa used similar scales). Biological justifications included aspects of species biology such as: (1) daily movements (including references to home range, daily commute, nest provisioning, territory size and foraging distance) or (2) dispersal distance (average or maximum). Where two justifications were given, the biological one was recorded.

Supplementary analyses

We further explored the evidence that scale selection included the true scale of effect by evaluating the hypotheses that: (1) scales evaluated were larger for bigger and/or winged species; (2) scales of effect observed for the same species in different studies were more similar than expected by chance; (3) differences in observed scales of effect for the same species in different studies were not associated with differences in scales evaluated for those species; (4) scales evaluated included scales of space use (e.g. home range, territory size, dispersal distance); and (5) observed scales of effect were approximately equal to a scale of space use. Methods and results for these hypotheses are detailed in Appendix S5.

RESULTS

Summary of focal site studies

We found 71 studies that met our criteria and provided sufficient data to determine the scale of effect for 583 species, with observed scales of effect ranging from 10 m to 100 km (Fig. 3). Attributes of study design are summarized in Appendix S5.

Are observed scales of effect determined more by scale selection than species traits?

The largest and smallest scales evaluated in a study were by far the strongest predictors of observed scales of effect. The largest
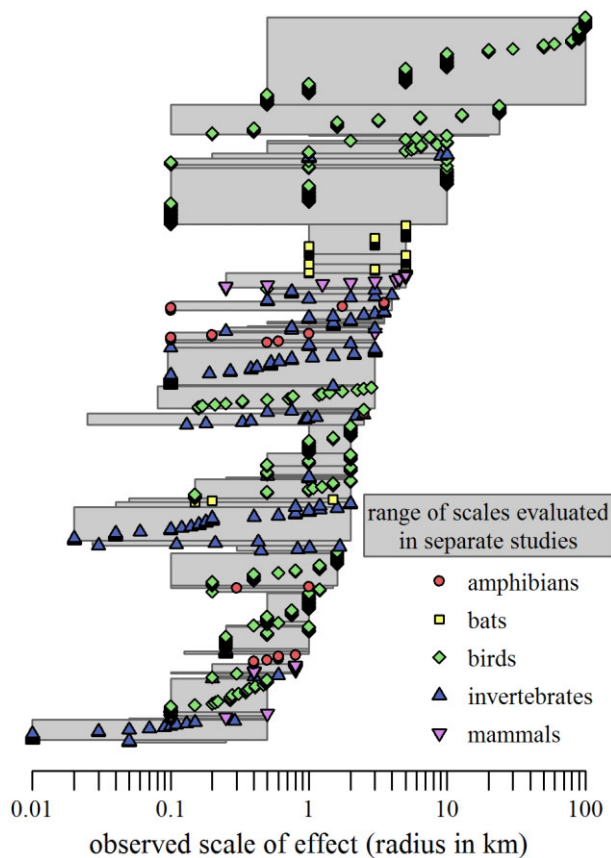
Figure 3 A visual summary of focal site multiscale studies in our review shows that studies tended to measure landscape within a small range of scales (median range 0.9 orders of magnitude, where the range for each study is outlined in grey) and at low densities (median 0.6 per km radius), which can cause inaccurate estimates of scale of effect (see Fig. 2). Scale of effect is on the x-axis (note the $\log_{10}$ scale). From bottom to top, studies are ordered by size of largest scale evaluated. Within each study, species are ordered by scale of effect.

Fig. 4), but when scales were taken into account by analysing major taxonomic group, largest scale and smallest scale together in the full model, differences among taxa were largely removed (i.e. partial regression estimates overlapped), indicating that the scales selected by researchers accounted for any differences among taxa (Fig. 4). The exceptions were amphibians and ground mammals which still had 95% confidence intervals that did not overlap each other's mean estimate. Even so, this difference between amphibians and mammals provided no meaningful information; the ER comparing the full model with a model without major taxonomic groups provided no evidence that major taxonomic group added information to the full model (ER = 0.0004). However, the full model was 187 billion times more likely to be the true model than a model without largest scale evaluated, and was 108 million times more likely to be the true model than one without smallest scale evaluated (Fig. 4, Appendices S4 & S6).

Similarly, the slopes describing the relationship between the other 28 species traits and observed scale of effect were generally weak, with confidence intervals that overlapped zero and with no consistent patterns in terms of their support for our hypotheses (Table 1, Appendix S6). There were some taxonomic groups with 95% confidence intervals that did not overlap with at least one other taxonomic group, but, as with the major taxonomic group analysis, the information provided by these distinctions was not useful (five of the six smallest ERs were for taxonomic group analyses; Table 1).

The smallest and/or largest scale evaluated accounted for almost all of the 'random effect' associated with studies, suggesting that the influence of study attributes other than scales evaluated on the observed scale of effect was minimal. We found large random effect estimates in species-trait-only models (median 0.28, only 13/29 overlapping with zero), but when scale(s) was/were added to models, random effect estimates were much smaller (median 0.08, 28/29 overlapping with zero), indicating that after accounting for scale there was little residual correlation among observed scales of effect within the same study (Appendix S6).

When bird and invertebrate data were restricted to those for which we were most confident in the observed scale of effect, some species traits had a weak relationship with scale of effect but the major outcome was unchanged: scale of effect was largely predicted by the size of the smallest and/or largest scales evaluated in the study (Appendix S6). Smallest and/or largest scale was always informative when added to a species trait in the full model (median ER = 4.1, range 1.0–45.5), but only 3 of the 13 species traits were informative when added to scale(s) in the full model: invertebrate families, bird migratory status and bird territory size. Invertebrate family, however, was so strongly correlated with the largest scale evaluated (r = 0.98), that the effects of scale and species trait were indistinguishable. Mean migrant scale of effect was 2.0 times the mean resident scale of effect (CI 1.1–3.6 times, ER = 1.50; Appendix S6), a pattern that was consistent with our predictions, but opposite to the relationship found when all data were used (Table 1). The negative relationship between observed scale of effect and bird territory size

and/or smallest scale evaluated (as opposed to climate, type of focal habitat, or climate) was/were the most informative study design predictor(s) for all 29 analyses. In 13/29 analyses, models were improved (i.e. AIC scores were lowered) by including both the smallest and largest scale rather than just one; smallest and largest scales were strongly correlated (r > 0.6) in the remaining analyses.

None of the 29 species traits investigated were informative predictors. The scale-only model was always the best model when compared with the full model (species trait + scale/s) or the species-trait-only model. Evidence ratios comparing full models with models with the species trait removed were always much less than 1 (median 0.13, range 0.00–0.60; Table 1, Appendices S6 & S7).

The analysis of major taxonomic groups had the greatest sample size and provided the most dramatic results. There were apparent differences among taxa when raw data were evaluated (amphibians = invertebrates < ground mammals = birds = bats;

Trait category (hypothesis)	Traits consistent with hypothesis (ER)	Traits inconsistent with hypothesis (ER)
Taxonomic group (any difference among groups)	Mammal orders (0.15) Bat families (0.12) Bird orders (0.01) Invertebrate families (0.01) Major taxonomic groups (0.00) Bird families (0.00) Invertebrate orders (0.00)	
Body size (positive)	Bird length (0.49) Invertebrate length (0.35) Bird mass (0.13) Bat mass (0.12)	Amphibian length (0.60) Mammal mass (0.31) Bat length (0.29)
Mobility (positive)	Bat migration, y/n (0.08)	Bird migration, y/n (0.16) Invertebrate flight, y/n (0.12) Bird home range (0.09) Bird territory size (0.06) Amphibian migration dist. (0.04)
Reproductive rate (negative)	Bird clutch size (0.17)	Bat litter size (0.33) Bird clutch size × no. of broods (0.15) Amphibian clutch size (0.13)
Trophic level (herbivores < carnivores)	Bats (0.48)	Invertebrates (0.46) Birds (0.01)
Geographic range (positive)		Bats (0.00)

y/n, yes/no.

Table 1 Summary of relationships between species traits and observed scales of effect after accounting for scales evaluated (i.e. the smallest and largest GIS buffer radii). In parentheses is the evidence ratio (ER) for each trait calculated by dividing the likelihood of the full model (trait + scale(s) evaluated) by the likelihood of the model without a trait (scales evaluated only). There was little evidence that any of the species traits should be in the best model predicting observed scale of effect (all ER values << 1). The evidence that scales evaluated contributed to the model was always strong (ER values > 1). See Appendix S6 for the complete model output.
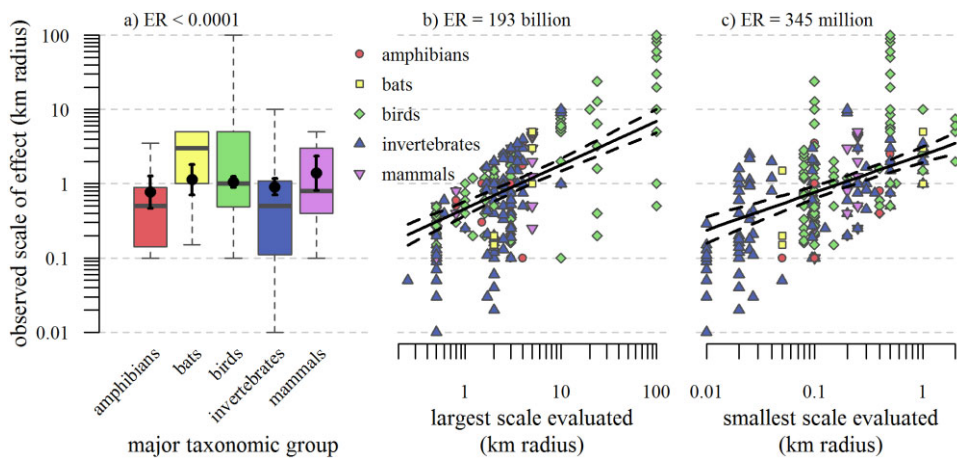

Figure 4 Relationships between observed scale of effect and (a) major taxonomic group, (b) largest scale evaluated, and (c) smallest scale evaluated indicate that there is little difference in predicted scale of effect among major taxonomic groups, but there are large differences in predicted scale of effect as the largest and smallest scales evaluated increase. All three variables are predictors in one linear mixed model with study included as a random effect (71 studies, 583 species). By comparing raw data (grey-outlined dots and box plots) with expected values (thick black dots and lines), it is apparent that any differences among taxonomic groups in the raw data were completely explained by differences in the smallest and largest scales evaluated by researchers for those taxa. The evidence ratios (ER, the likelihood of the full model versus the model without the predictor) indicate no evidence that a model is improved by considering major taxonomic group and overwhelming evidence for the information value of the largest and smallest scale evaluated. An overwhelming effect of the smallest and/or largest scale(s) evaluated on the observed scale of effect is supported by analyses of all other species traits in this review (Appendix S6). Expected values and confidence intervals for each predictor were calculated while other predictors were held at their median values. Box-plot whiskers, box edges and midline bars indicate the most extreme data points that are 1.5 times the length of the box away from the box, the first and third quartiles, and the median, respectively (Tukey, 1977).

(slope −0.33, CI −0.55 to −0.11, ER = 5.08; Appendix S6) was opposite to our predictions, but consistent with our more inclusive analysis (Table 1). Other bird traits (reproductive rate, clutch size, length, mass, family and order) and invertebrate traits (order, length and trophic level) were less informative and model estimates showed little certainty (Appendix S6).

Further evidence that scales evaluated did not include the true scale of effect

A large proportion of species in our review (44%) had an observed scale of effect equal to the smallest (21%) or largest (23%) scale evaluated in a study, making it uncertain whether the true scales of effect for these species were within the observed ranges. Increasing the number and range of scales evaluated removed much of the ambiguity. Studies which evaluated a wide range of scales and had a high density of scales tended to have lower proportions of species with scales of effect equal to the smallest [range, $X_1^2 = 15.5$, $P < 0.001$; $\log_{10}$(GIS buffers per km radius), $X_1^2 = 12.6$, $P < 0.001$] or largest [range, $X_1^2 = 24.1$, $P < 0.001$; $\log_{10}$(GIS buffers per km radius), $X_1^2 = 33.9$, $P < 0.001$] scale evaluated than less optimally designed studies (Fig. 5; $n = 71$ studies including 583 species). From the least optimal to the most optimal study design [least optimal (landscape scale varied by 0.25 orders of magnitude with only 0.03 scales evaluated per km radius) versus most optimal (landscape scales varied by 2.38 orders of magnitude with 10 scales evaluated per km radius)], the expected percentage of species with scales of effect equal to the smallest scale evaluated dropped from 76 to 1%. The comparable drop for species with scales of effect equal to the largest scale evaluated was even more dramatic (94 to 0%).

Researchers did not usually explicitly select scales based on species traits. Most (51/71; 71%) justifications for landscape scales were non-biological (Fig. 6). The remaining 29% of studies selected scales that overlapped with some aspect of space use.

Similarly, there was little evidence that researchers implicitly considered biology when selecting landscape scales: the sizes of the smallest and largest scales were unrelated to the average body size of species within studies. Studies of winged species, however, tended to use larger scales than studies of flightless species of the same size (Appendix S5).

The scales of effect measured in multiple studies for the same bird species were no more similar to each other than to the scales of effect of randomly selected birds ($P = 0.85$, 67 species). Observed scales of effect varied widely among different studies of the same species (median difference 9.6 km, range 0.0–99.8 km). The difference in scales of effect for the same species in different studies was best predicted by the differences in radii of the largest scale evaluated in those studies (ΔAICc > 10 compared with models considering the effect of difference in the smallest scale evaluated, body length, clutch size or territory size; $r^2 = 33$%; Appendix S5).

Scales selected by researchers often included average dispersal distance (65% of 60 species with known dispersal distances) but
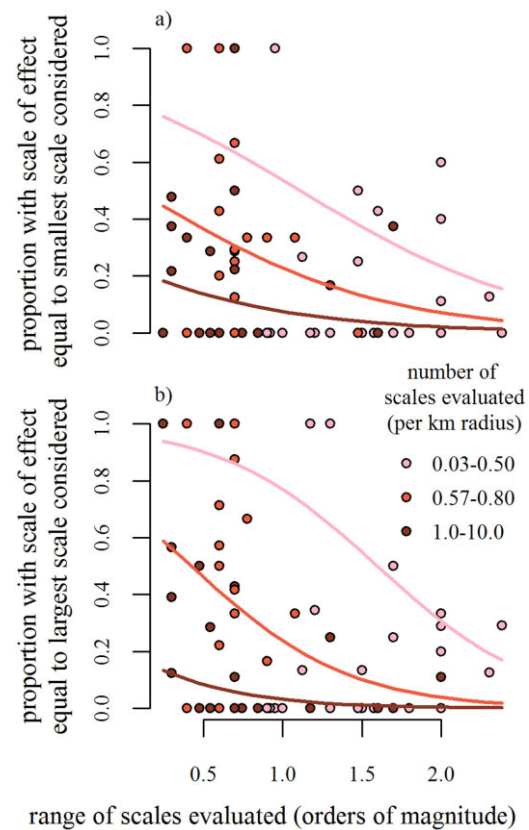

Figure 5 Associations between the range of scales evaluated and (a) the proportion of species with observed scale of effect equal to the smallest scale evaluated or (b) the proportion of species with scale of effect equal to the largest scale evaluated show that, as predicted, studies which measured landscape structure at a wider range of scales and/or measured a greater density of scales tended to have fewer species whose scale of effect was equal to the smallest or largest scale evaluated. Almost half the species in our review had scales of effect equal to the smallest (21%) or largest (23%) scales, suggesting the possibility that the true scales of effect were commonly outside the range of scales evaluated (see Fig. 2a–d). Lines indicate least squares means predictions from logistic regression models in which range in scales evaluated ($\log_{10}$) and GIS buffer density (number per km radius, $\log_{10}$) are predictors, where GIS buffer densities at the minimum (0.03), median (0.6) and maximum (10) are shown. Each dot represents a single study.

did not usually include other scales of space use. The effect of landscape structure was evaluated at home range (18/50), territory size (25/246), four times the average dispersal distance (28/60) or nine times the average dispersal distance (22/60) for fewer than 50% of species (Appendix S5). Of those species for which scales of space use were included in the range of scales evaluated, the average observed scale of effect was significantly larger than home range size or territory size, but was not significantly different from average natal dispersal distance or from the expectations produced by our simulation models (four or nine times the average dispersal distance; Jackson & Fahrig, 2012).
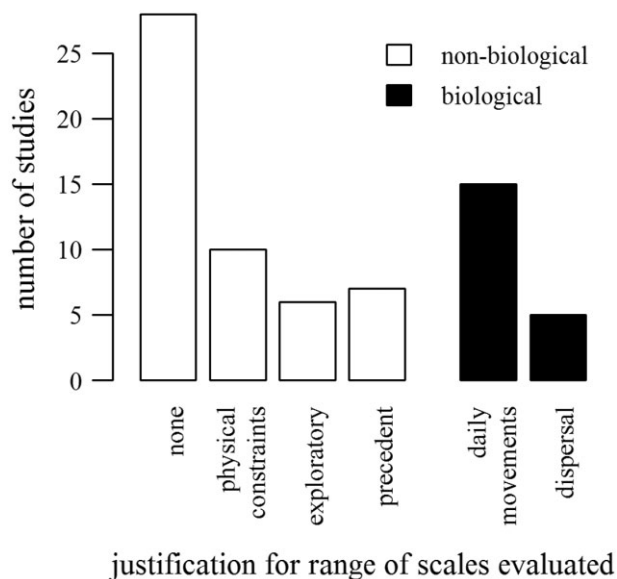
Figure 6 A summary of researchers' verbal justifications for the range of scales evaluated shows that most researchers did not mention biological reasons for their choice of scales. Physical constraints included data limitations and a desire to avoid overlapping and potentially non-independent GIS buffers.

DISCUSSION

Using multiple lines of evidence, we found strong support for the hypothesis that multiscale studies do not usually measure landscape structure at the true scale of effect. This implies that these studies do not accurately estimate species–landscape relationships because they do not measure landscape structure at the scales where these relationships are strongest. Our first line of evidence is that the largest or smallest scales at which landscape structure was measured were consistent, strong, positive predictors of the observed scale of effect, whereas none of the 29 species traits investigated were good predictors of observed scale of effect (Fig. 4, Table 1, Appendix S6). This suggests that limited ranges of scales evaluated in studies were driving observed scales of effect, rather than innate attributes of species. One might argue that a strong relationship between the observed scale of effect and landscape scales indicates accurate a priori selection of scales (i.e. landscape scales predict observed scales of effect because landscape scales were close to the true scale of effect), but this was not supported. Few researchers (29%; Fig. 6) provided biological explanations for their selection of scales and there was no relationship between the most easily measured species trait (median body size) and the scales selected by researchers (Appendix S5).

Other lines of evidence support the idea that the range and number of scales selected by researchers were too limited and/or arbitrary to allow for accurate estimates of the true scale of effect. Besides being selected for non-biological reasons, the ranges of scales evaluated were narrow (median of 1.96 km or a 0.90 order of magnitude difference between the smallest and largest scale evaluated) and/or imprecise (a median of four scales were evaluated per study or one scale per 1.67 km radius; Fig. 3). One reason for the small numbers of scales evaluated in studies may be that researchers incorrectly assume that between-scale correlations in landscape variables need to be low. In the context of finding the scale of effect (Fig. 1), this criterion is unnecessary and can lead to an imprecise estimate of scale of effect (Fig. 2d). Almost half of species had scales of effect equal to the smallest or largest scale evaluated in a study, suggesting that for many species the true scale of effect was outside the range of observed scales (Fig. 2). Indeed, as the range and number of scales in a study increased, the proportion of species with scales of effect equal to the smallest or largest scale decreased (Fig. 5). Note that even when the observed scale of effect is not equal to the smallest or largest scale evaluated, the true scale of effect may be outside the range evaluated if random variation in species–landscape relationships causes an apparent peak in explanatory power at an intermediate scale. This may explain why in analyses from which studies with observed scales of effect equal to the smallest or largest scale evaluated were removed, the radii of the smallest and largest scales evaluated continued to be the best predictors of observed scale of effect.

Finally, a lack of a consistent observed scale of effect for a species could indicate that the scale of effect is not an innate attribute of a species, but our data suggest that a more likely explanation is the lack of consistency in the scales evaluated by different researchers. When scales of effect for the same species were compared across studies, they were no more similar to each other than to those of randomly selected species (Appendix S5). The difference in observed scales of effect for the same species in different studies was strongly, positively predicted by the difference between the largest scales evaluated in those studies.

Inconsistencies in observed scales of effect across studies could also indicate that scale of effect depends on environmental context. Home range size, for example, can vary with environmental context (Gompper & Gittleman, 1991; Kie *et al.*, 2002; Börger *et al.*, 2006; Morellet *et al.*, 2013), and scale of effect might be likewise expected to do so. However, neither latitude nor focal habitat type were informative predictors of observed scale of effect when compared with scales evaluated. After the smallest and largest scales evaluated were included in mixed models, random effects were almost zero, indicating that the primary factors causing scales of effect within the same study to be similar to each other were the scales evaluated in that study.

That the scales evaluated in a study will largely determine the scales that are considered important may seem obvious, but if the intention behind evaluating the effect of landscape at multiple scales is to find a biologically meaningful scale for a species, our results provide the surprising conclusion that we have not been successful. The true scale of effect is often either larger than the largest or smaller than smallest scale evaluated by researchers. By measuring landscape structure at a suboptimal scale researchers are underestimating the effects of landscape structure on abundance.

Effects of species traits on scale of effect, if any, are obscured

Without accurate estimates of scale of effect, adequate empirical assessment of a relationship between scale of effect and species traits is impaired. Of all the species traits, body size was the one for which we had most expected a relationship with observed scale of effect, because body size data are the most accurate and widely available, because some empirical studies have indicated a positive relationship (Hostetler & Holling, 2000; Holland *et al.*, 2005; Thornton & Fletcher, 2014) and because body size has so often been shown to be related to important aspects of a species' biology (Peters, 1986). Among our data, however, the evidence supporting a relationship between body size and scale of effect was weak, the effect sizes were small and the direction of the effect (positive or negative) varied among major taxonomic groups and even within the same taxonomic group (birds), depending on whether all data were used or just a selective subset (Table 1, Appendix S6). The evidence suggests that any relationship between body size and scale of effect in our data, if there is one, was obscured by variation in the scales evaluated across studies. In Appendix S8 we discuss the species traits for which we found some weak evidence for a relationship between the scale of effect and a trait. These traits include bird territory size (negative relationship with observed scale of effect) and bird migratory status (scale of effect larger for migrants than for residents).

Other factors influencing scale of effect

Despite our null results regarding the influence of species traits on scale of effect, we nevertheless suspect that the true scale of effect is primarily determined by species traits (especially mobility), based on simulations (Jackson & Fahrig, 2012; Ricci *et al.*, 2013), the strong scaling relationships among other aspects of species biology (e.g. body size, home range size, dispersal distance, geographic range and reproductive rate; Brown *et al.*, 1993; Hendriks *et al.*, 2009) and the intuition we share with other researchers (Fig. 6).

Before any hypothesis concerning the mechanisms underlying scale of effect can be evaluated, the scale of effect itself must be accurately estimated. We leave a discussion of other potential complications in the relationship between species traits and scale of effect to the Supporting Information (Appendix S8), where we discuss the following possibilities: (1) environmental context has a strong influence on scale of effect, thereby masking the effects of species traits; (2) species have multiple scales of effect; (3) variation in study design is obscuring the relationship between scale of effect and species traits; and (4) species trait data are flawed.

How can we ensure landscape research is conducted at the best spatial scale?

The obvious practical recommendations from our results are that researchers should select scales based on expected scales of space use, and they should increase the range and number of scales at which landscape is measured. Ideally, a study would evaluate landscape at scales that range from less than the home range of a species to greater than nine times the average dispersal distance (because scale of effect is predicted by simulations to be equal to four to nine times the average dispersal distance; Jackson & Fahrig, 2012). Most studies thus far have not included the relevant scales of space use (i.e. home range, territory size, average dispersal distance, or four to nine times the average dispersal distance; Appendix S5). The inclusion of a larger range and/or density of scales, however, improved study outcomes (Fig. 4).

If measuring landscape structure at large scales is not practical, we suggest that researchers should be clear about the limitations of their work. If the goal of research is to provide practical information for landscape managers, for example, it may be appropriate to limit measurements of landscape structure to within the range of scales for which management can be accomplished. However, if the goal of research is to detect true species–landscape relationships, authors should admit the limits of their data and the resulting inferences, namely that a different species–landscape relationship may occur at scales not evaluated in their study.

An excellent reason for researchers to continue to consider the relationship between species abundance and landscape structure at multiple scales, even if they cannot consider the ideal range of scales, is simply to increase the probability of finding a relationship if there is one. As Holland *et al.* (2004) found when evaluating beetle–landscape relationships across scales that ranged from 20 to 2000 m, a strong relationship between landscape structure and abundance can be found at one scale even if there is no relationship at other scales. The greatest danger if research is conducted at the wrong scale, therefore, is to fail to recognize an important species–landscape relationship.

CONCLUSION

We provide quantitative evidence that landscape ecologists are generally conducting research at scales that are too few, too narrow in range and at landscape sizes that are not biologically justified. This has impaired the detection of species–landscape relationships. To mischaracterize important species–landscape relationships simply because the landscape is measured at the wrong spatial extent is at best unproductive and at worst can lead to mismanagement of diversity. In addition, although theory and intuition support a relationship between the scale of effect and species traits such as dispersal, poor estimation of the scale of effect has prevented us from identifying those relationships, assuming that they exist.

ACKNOWLEDGEMENTS

REFERENCES

Baayen, R.H. (2011) languageR: data sets and functions with 'Analyzing Linguistic Data: a practical introduction to statistics'. R package version 1.4. Available at: http://cran.r-project.org/package=languageR (accessed 1 June 2012).

Bates, D., Maechler, M. & Dai, B. (2013) lme4: linear mixed-effects models using S4 classes. R package version 0.999999-2. Available at: http://cran.r-project.org/package=lme4 (accessed 1 May 2013).

Bell, B.A., Ferron, J.M. & Kromrey, J.D. (2008) Cluster size in multilevel models: the impacts of sparse data structures on point and interval estimates in two-level models. *JSM 2008: Proceedings of the section on survey research methods. Denver, CO*, pp. 1122–1129. American Statistical Association, Alexandria, VA.

Börger, L., Franconi, N., Ferretti, F., Meschi, F., De Michele, G., Gantz, A. & Coulson, T. (2006) An integrated approach to identify spatiotemporal and individual-level determinants of animal home range size. *The American Naturalist*, **168**, 471–485.

Bowman, J. (2003) Is dispersal distance of birds proportional to territory size? *Canadian Journal of Zoology*, **81**, 195–202.

Bowman, J., Jaeger, J.A.G. & Fahrig, L. (2002) Dispersal distance of mammals proportional to home range size. *Ecology*, **83**, 2049–2055.

Brennan, J.M., Bender, D.J., Contreras, T.A. & Fahrig, L. (2002) Focal patch landscape studies for wildlife management: optimizing sampling effort across scales. *Integrating landscape ecology into natural resource management* (ed. by J. Liu and W.W. Taylor), pp. 68–91. Cambridge University Press, Cambridge.

Brown, J.H., Marquet, P.A. & Taper, M.L. (1993) Evolution of body size – consequences of an energetic definition of fitness. *The American Naturalist*, **142**, 573–584.

Burnham, K.P., Anderson, D.R. & Huyvaert, K. (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, **65**, 23–35.

Gompper, M.E. & Gittleman, J.L. (1991) Home range scaling – intraspecific and comparative trends. *Oecologia*, **87**, 343–348.

Hendriks, A.J., Willers, B.J.C., Lenders, H.J.R. & Leuven, R.S.E.W. (2009) Towards a coherent allometric framework for individual home ranges, key population patches and geographic ranges. *Ecography*, **32**, 929–942.

Holland, J.D., Bert, D.G. & Fahrig, L. (2004) Determining the spatial scale of species' response to habitat. *Bioscience*, **54**, 227–233.

Holland, J.D., Fahrig, L. & Cappuccino, N. (2005) Fecundity determines the extinction threshold in a Canadian assemblage of longhorned beetles (Coleoptera: Cerambycidae). *Journal of Insect Conservation*, **9**, 109–119.

Horne, J.K. & Schneider, D.C. (1995) Spatial variance in ecology. *Oikos*, **74**, 18–26.

Hostetler, M. & Holling, C.S. (2000) Detecting the scales at which birds respond to structure in urban landscapes. *Urban Ecosystems*, **4**, 25–54.

Jackson, H.B. & Fahrig, L. (2012) What size is a biologically relevant landscape? *Landscape Ecology*, **27**, 929–941.

Jackson, N.D. & Fahrig, L. (2014) Landscape context affects genetic diversity at a much larger spatial extent than population abundance. *Ecology*, **95**, 871–881.

Jenkins, D.G., Brescacin, C.R., Duxbury, C.V., Elliott, J.A., Evans, J.A., Grablow, K.R., Hillegass, M., Lyon, B.N., Metzger, G.A., Olandese, M.L., Pepe, D., Silvers, G.A., Suresch, H.N., Thompson, T.N., Trexler, C.M., Williams, G.E., Williams, N.C. & Williams, S.E. (2007) Does size matter for dispersal distance? *Global Ecology and Biogeography*, **16**, 415–425.

Kie, J.G., Bowyer, R.T., Nicholson, M.C., Boroski, B.B. & Loft, E.R. (2002) Landscape heterogeneity at differing scales: effects on spatial distribution of mule deer. *Ecology*, **83**, 530–544.

de Knegt, H.J., van Langevelde, F., Coughenour, M.B., Skidmore, A.K., de Boer, W.F., Heitkonig, I.M.A., Knox, N.M., Slotow, R., van der Waal, C. & Prins, H.H.T. (2010) Spatial autocorrelation and the scaling of species–environment relationships. *Ecology*, **91**, 2455–2465.

Lechner, A.M., Langford, W.T., Jones, S.D., Bekessy, S.A. & Gordon, A. (2012) Investigating species–environment relationships at multiple scales: differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, **11**, 91–102.

Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.

McGarigal, K., Cushman, S.A. & Ene, E. (2012) FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at: http://www.umass.edu/landeco/research/fragstats/fragstats.html

Morellet, N., Bonenfant, C., Borger, L., Ossi, F., Cagnacci, F., Heurich, M., Kjellander, P., Linnell, J.D.C., Nicoloso, S., Sustr, P., Urbano, F. & Mysterud, A. (2013) Seasonality, weather and climate affect home range size in roe deer across a wide latitudinal gradient within Europe. *Journal of Animal Ecology*, **82**, 1326–1339.

Peters, R.H. (1986) *The ecological implications of body size*. Cambridge University Press, Cambridge.

R Core Team (2013) R: a language and environment for statistical computing. R version 3.0.1. Available at: http://www.R-project.org/ (accessed 16 May 2013).

Ricci, B., Franck, P., Valantin-Morison, M., Bohan, D.A. & Lavigne, C. (2013) Do species population parameters and landscape characteristics affect the relationship between local population abundance and surrounding habitat amount? *Ecological Complexity*, **15**, 62–70.

Thornton, D.H. & Fletcher, R.J. (2014) Body size and spatial scales in avian response to landscapes: a meta-analysis. *Ecography*, **37**, 454–463.

Tukey, J.W. (1977) *Exploratory data analysis*. Addison Wesley, Reading, MA.

Wheatley, M. & Johnson, C. (2009) Factors limiting our understanding of ecological scale. *Ecological Complexity*, **6**, 150–159.

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 List of previously derived relationships between species traits and observed scales of effect.
Appendix S2 Review data.
Appendix S3 Metadata for Appendix S2 and list of sources for review data.
Appendix S4 Analysis of the effect of major taxonomic groups on scale of effect.
Appendix S5 Supplemental analyses investigating whether studies were conducted at the best scale.

Appendix S6 Full models predicting scale of effect.
Appendix S7 Metadata and supplementary information for Appendix S6.
Appendix S8 Supplementary discussion.

Additional references may be found at the end of Appendices S1, S3, S4, S5 & S7.

BIOSKETCHES

Heather Bird Jackson is an ecologist who investigates the most appropriate spatial scales of observation for species as well as how these ideal scales can be reconciled with scales that are practical for researchers and managers.

Lenore Fahrig is an ecologist who studies the effects of landscape structure on abundance, distribution and persistence of organisms using a combination of spatial simulation modelling and field studies on a wide range of organisms.

Editor: Joshua Lawler