PLoS one

Multiscale Analyses of Mammal Species Composition – Environment Relationship in the Contiguous USA

Rafi Kent[1]*, Avi Bar-Massada[2], Yohay Carmel[1]

1 Faculty of Civil and Environmental Engineering, The Technion – Israel Institute of Technology, Haifa, Israel, 2 Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, Wisconsin, United States of America

Abstract

Relationships between species composition and its environmental determinants are a basic objective of ecology. Such relationships are scale dependent, and predictors of species composition typically include variables such as climate, topographic, historical legacies, land uses, human population levels, and random processes. Our objective was to quantify the effect of environmental determinants on U.S. mammal composition at various spatial scales. We found that climate was the predominant factor affecting species composition, and its relative impact increased in correlation with the increase of the spatial scale. Another factor affecting species composition is land-use–land-cover. Our findings showed that its impact decreased as the spatial scale increased. We provide quantitative indication of highly significant effect of climate and land-use–land-cover variables on mammal composition at multiple scales.

Introduction

Understanding the factors that affect the distribution of biodiversity in time and space is a central objective of ecology [1]. Relationships between environmental variables (e.g., climate, topography) and biodiversity patterns are scale-dependent, both spatially and temporally [2]. Species richness, probably the most studied aspect of biodiversity, has often been shown to vary as a function of spatial scale [3,4].

Theories concerning the mechanisms governing distribution patterns of biodiversity measures range from global (latitudinal species richness gradient) to very local scales, and relate to environmental, historical and evolutionary processes [5]. Most studies concentrate on species richness as a measure of biodiversity, resulting in various theories and hypotheses offering mechanistic explanations for patterns of species richness. These explanations are related to interspecific interactions and climate conditions [6], energy levels [7], area effects [8], neutral theory [9] and others. However, the mechanistic explanations for species richness patterns do not necessarily extend to explanations of species composition patterns. Two areas may hold a similar number of species, while the identity of the species might differ considerably, rendering species richness of little value to differentiate between them [10]. Recently, it has been suggested that species richness patterns are largely determined by historical-biogeographical processes [11].

Here we focus on patterns of species composition, rather than species richness. Theories on species composition patterns include a neutral model, which suggests that all variations are caused by random differences in the dispersal of demographically and competitively equal species [9,12]; an environmental model, which relates species distributions to environmental conditions [13]; and a model that claims that species composition is determined by interspecific interactions within and between trophic levels [14].

Although species composition has rarely been studied at multiple spatial scales, there are exceptions such as the studies of Grand and Cushman [15] and Grand and Mello [16], in which scale was defined qualitatively, i.e., plot, patch and landscape scale. However, most studies on species composition were restricted to a single scale [17,18,19,20]. Applied across multiple scales, multivariate analyses may provide a wider picture of the relationships between environmental variables and species composition [21,22]. Understanding species composition – environment relationships, and specifically how they are affected by spatial scale, may improve the ability of conservationists to predict both the spatial distribution of biodiversity, and its reaction to global and regional changes [23].

There are serious conceptual and practical impediments to such analyses. A central conceptual challenge is the nature of scale [24]. Scale is characterized by both grain (grid cell size) and extent [25]. In most studies, a change of only a single element of scale is regarded as a full change of scale [26]. Here we used a "complete" approach, in which both grain and extent are modified together in the process of upscaling (see Appendix S1 for details).

The major practical impediment for such analyses is data availability [27]. Presence-absence data are only available for relatively small extents [27]. Presence-only (occurrence) data in large quantities and for diverse taxonomic groups have become available in the last decade via data portals such as the Global Biodiversity Information Facility (GBIF) and other portals that allow easy access to digitized databases, mostly based on museum

and university collections [28]. However, presence-only data are often considered improper for such analyses, due to a range of inherent biases [29,30]. The validity of using presence-only data in ecological analyses has been studied repeatedly in the context of modeling the distribution of a single species or modeling species richness patterns, but results are inconclusive [31]. In a previous study [32], we evaluated the reliability of using presence-only data for studying multiscale diversity patterns based on taxonomic or functional group composition. The assessment confirmed that presence-only data are sufficient for analyzing the relationships between species composition and environmental determinants. The objective of the current study was to quantify the variation in the relationships between mammal species composition and its environmental determinants, at varying spatial scales. More specifically, we hypothesized that climate is the predominant environmental factor affecting species composition at large scales. An additional hypothesis was that land-use and land-cover (LULC) variables are highly influential at fine spatial scales, however, when grain size is large enough to contain all (or most) of the possible LULC types, the effect of those variables will diminish. Regarding topography and primary productivity, we hypothesized that their effect will be more prominent at small scales.

Results

Variable group effects

Climate and Land use – Land cover (LULC) variables explained the largest amount of variance in community composition at all spatial scales of the analyses (Fig. 1). The amount of variance explained by LULC variables decreased gradually until the seventh scale (grain size 1,280 km^2,,extent 1.3*10^6), and then dropped sharply from ~30% of the total explained variance to ~15% between scales 7 and 9. Climate, which explained a slightly smaller proportion of the variance in species composition than LULC variables at the six smaller six scales, also showed a decrease in the amount of explained variance until scale 7, but then exhibited an increase between scales 7 and 10. Topography and primary production explained a relatively small amount of variance in species composition at all the analyzed scales. In general, the proportion of explained variance in species composition decreased with increasing scale, except at the largest scale, where both variable groups exhibited a moderate increase (Fig. 1). The correlation between effective gradient length of the different variable groups and the proportion of variance in species composition explained by each group was intermediate (Pearson's r = 0.44, Fig. 2).

Within - group analyses

Within the group of climatic variables, all four variables had equal contribution to the variance explained by the group at the four smallest scales (Fig. 3a). However, at the larger scales mean annual temperature was the predominant climatic feature. Precipitation seasonality also explained a relatively high proportion of the total explained variance at three of the five largest scales (Fig. 3a).

In the topography group, Altitude generally explained a larger proportion of the variance in species composition compared to altitude range. However differences were relatively small at the small scales and larger at the larger scales (Fig. 3b). Land-cover variables in the LULC group explained most of the variance in that group at all scales (Fig. 3c). However, when we plotted the mean standard deviation in land cover variables, within the LULC groups (i.e., agriculture, forestry etc.), we found that the decrease in the amount of variance explained by LULC variables (Fig. 1)

corresponded to a sharp decrease in the variance in the land cover variables at the eighth scale (Fig. 4).

Discussion

Our analyses revealed that at grain sizes of 10^1 to 10^5 km^2 and extents from 10^5 to 10^8 km^2 respectively, mammal species composition is affected largely by climate and LULC variables. LULC variables had sizeable influence on species composition at the smaller scales, probably via habitat degradation and fragmentation, and ultimately, habitat loss [33]. Topography was not a prominent factor in these analyses, but it is probably more important at finer scales [34]. These results partially corroborate our hypotheses. As we hypothesized, climate is indeed a predominant factor affecting mammal species composition within the contiguous USA. However, at smaller scales, LULC variables are more influential, and explain a larger amount of variance in species composition than climate. This is consistent with theoretical predictions that at fine scales, effects of climatic determinants are obscured by biological interactions and that the effect of climate becomes more evident at larger scales [26]. Also corroborating our hypotheses is our finding that, at the largest scales, as the variance within sampling units increases and the variance between them decreases, LULC become less explanatory. We found intermediate correlation levels between the effective gradient lengths (for an explanation of effective gradient length please see Methods and Materials section) and the amount of variance in species composition explained by the different variable groups. This suggests that there is a change in the reaction of species composition to environmental gradients varies among scales, although some of that change is attributable to differences in gradient lengths. Wiens [26] described a phenomenon called scale-domains, based on a review of studies that used different sized quadrats to study patterns of plant distributions. He suggested that the change in distribution patterns of ecological phenomena observed with scale is monotonous within each scale-domain. In contrast, between domains, pattern variability is chaotic and unpredictable, manifested as high variability between sampling units. Accordingly, our findings indicate that between the seventh and eighth scales there is a possible a shift from one scale domain to another, in both climate and LULC variable groups (Fig. 1). The existence of scale domains is possibly indicated here by the shift in the direction and slope of the line in Figure 2. The high levels of explained variance attributed to land-cover (i.e., forestry, agriculture, urban area, etc.) suggest that at all scales, land cover type is the predominant human related factor affecting mammal species composition.

This study, to the best of our knowledge, is among the first to analyze the relationships between species composition and the environmental conditions that affect it at large and multiple spatial scales [but see for example 35]. We found that scale was a prominent factor in these relationships, having a greater impact than that of geographical factors that affect environmental conditions within each scale. This line of research has the potential to contribute much to the understanding of global biodiversity patterns. Studying other taxonomic groups and other regions of the world would be an important step towards establishing a knowledge base of these relationships, which in turn may serve to test general biogeography theories.

Materials and Methods

Our data consisted of all occurrence records found in the GBIF portal [36] of terrestrial mammals (excluding bats) in the contiguous USA. All data were downloaded from GBIF during
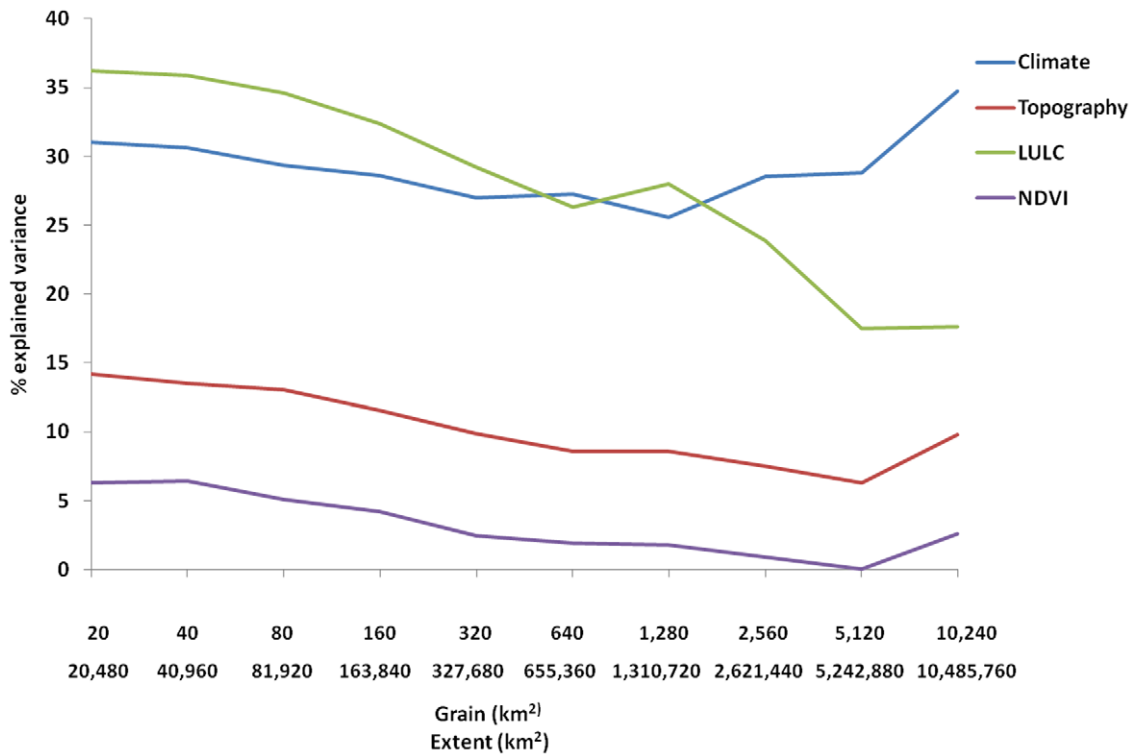
Figure 1. Explained variance rates of environmental variable groups in mammal species composition at varying spatial scales in the contiguous USA (demonstrated using CCA analyses). Scale consists of grain size (upper number on the x-axis) and extent (lower number on the x-axis). Explained variances represent the pure effect of each variable group used in the analyses (see Table 1 for details on the different variable groups).
doi:10.1371/journal.pone.0025440.g001

March–June 2009. Bats were excluded from the analyses under the assumption that the ecological demands and responses of bats to environmental variables may be very different than all other mammals, and thus may decrease the probability of elucidating coherent answers to our questions. Our dataset consisted of ~308,000 records, including 284 species. It originated from ~70 datasets within GBIF. We used all geo-referenced records of specimens and observations in the datasets, with the exception of

records with less than four decimal digits in at least one coordinate (either latitude or longitude).

In addition to mammal occurrence data, we compiled environmental data related to 15 variables, which we categorized according to 4 groups: climate; topography; land-use/land-cover (LULC); and primary productivity (Table 1). The spatial resolution of all environmental layers was (or was reduced to) $0.0833°$ (~10 km). As a measure of anthropogenic disturbance we
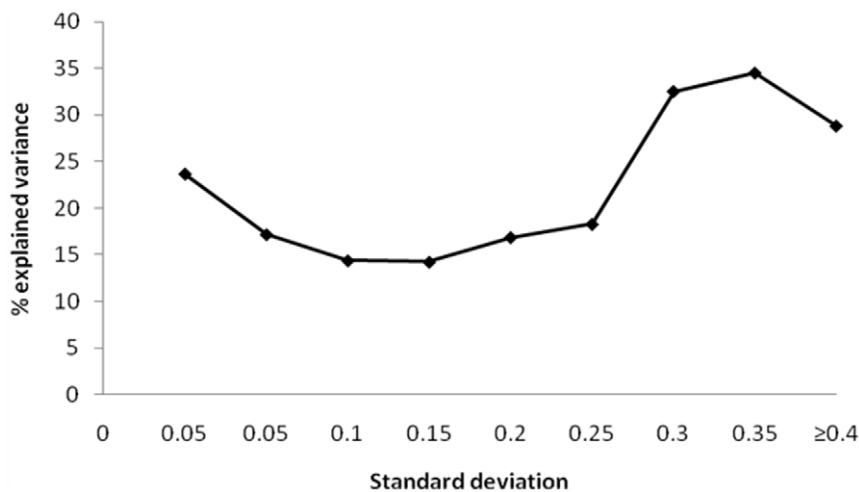

Figure 2. Mean values of standard deviation in environmental variables (for detailed description of variables see Table 1). For convenience, values are presented as averages in intervals of 0.05. Pearson's correlation between SD and % explained variance is 0.44.
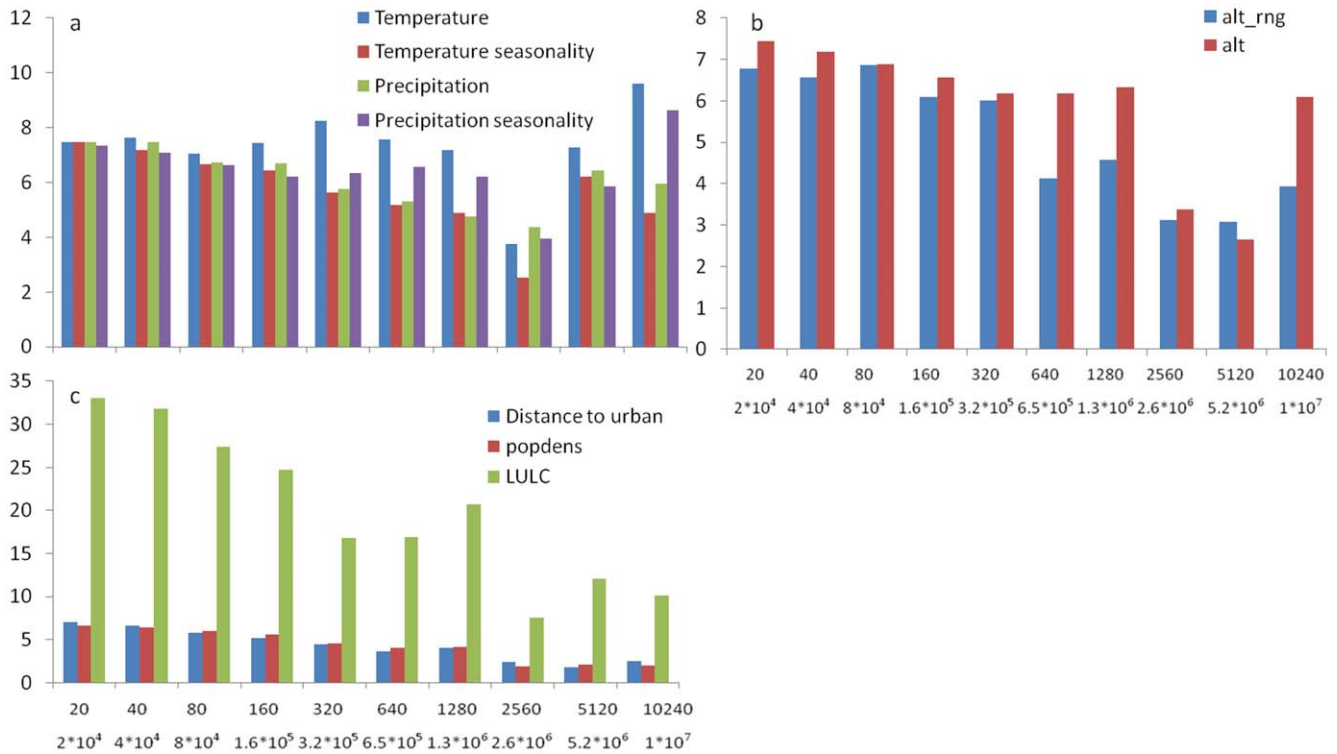doi:10.1371/journal.pone.0025440.g002

Figure 3. Explained variance rates of individual environmental variables in mammal species composition: a) climatic variables; b) topographic variables and c) LULC variables. Land cover is the combined effect of six land cover categories (agriculture, forestry, open herbaceous, urban, water and wetland). Prec_sea and Temp_sea stand for *precipitation seasonality* and *temperature seasonality* respectively; Altitude – rng stands for altitude range; DTU stands for *distance to nearest urban area*; and Pop density stands for *population density*.
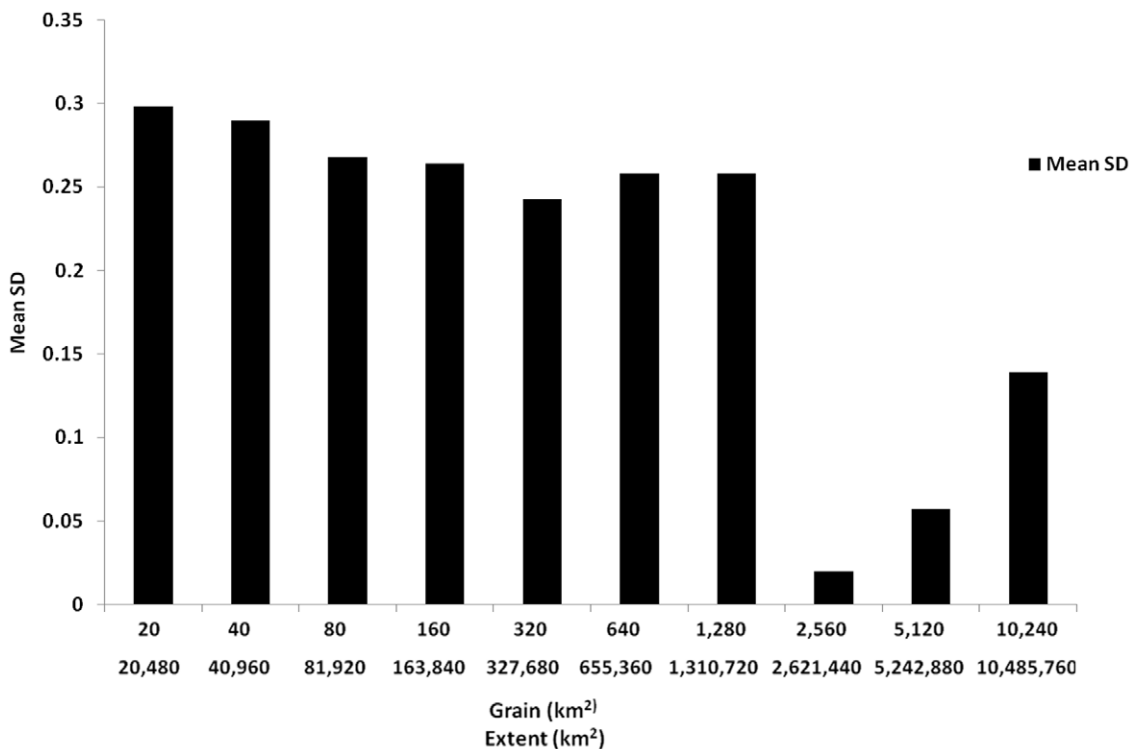doi:10.1371/journal.pone.0025440.g003

Figure 4. Mean value of standard deviation in land-cover variables (for detailed description of variables see Table 1) per spatial scale analyzed.
doi:10.1371/journal.pone.0025440.g004

measured the average distance to nearest urban area in 0.00833°
(~1 km) grid in the entire study area, using the Euclidean distance
function in ArcMap [37] with a polygonal urban area layer (see
Table 1). We then calculated the mean value of the distance to
nearest urban area in each cell in the grid, at each spatial scale.
Seasonality in climatic variables, i.e. temperature and precipitation
(Table 1), was represented by inter-month variance. The data were
downloaded as GIS layers from Worldclim [38]. The coefficient of
variation (CV) was the measure of variance used to represent
precipitation seasonality. Seasonality in temperature was repre-
sented as standard deviation, as CV makes no sense when values
are between −1 and 1. For more details see the Worldclim website
http://www.worldclim.org/bioclim.

To analyze the effects of scale on community composition, we
wrote an ArcGIS python script that generated sets of rectangular
sampling units of extent E and grain g at each scale (Table 2). At
each scale, the area of a grid cell is twice the area of a cell in the
previous scale. The value of each environmental variable in a cell
was calculated as the average of the respective values in all pixels
contained within that cell. In each sampling unit, the script
counted the number of pixels that had species observations in
them and the total number of species in those observations. We
then identified sampling units that had sufficient information for a
Canonical Correspondence Analysis (the CCA) analysis by setting
thresholds for numbers of species and pixels with observations. For
the subsequent statistical analysis, we only used sampling units that
had more than five species and at least 30 pixels with non-
singleton observations. For each sampling unit that complied with
the thresholds, and at each scale, we ran a partial Canonical
Correspondence Analysis (pCCA) using the vegan package [39] in
the R statistical software package, version 2.12 [40]. The
difference between CCA and pCCA is that pCCA decomposes
the explained variance to its components, i.e. it allows determining
how much of the variance is explained by individual variables or
variable groups. This is accomplished by using the variable(s) of
interest as constraints (i.e. explanatory variables) and the rest of the
environmental variables as conditioning variables (also termed co-

variables). Thus, the proportions of the variance explained by the
conditional variables alone and by the interactions between the
variable(s) of interest and the conditioning variables, are accounted
for [for a detailed description of pCCA see 21]. We ran pCCA for
each variable using the vegan package (Oksanen et al. 2011) in the
R statistical software package, version 2.12 (R Development Core
Team 2010). For pCCA, we split the environmental variables into
four groups: climate (mean annual temperature, temperature
seasonality, mean annual precipitation and precipitation season-
ality), topography (elevation and elevation range); land-use land-
cover (distance to urban areas, population density, and percent-
ages of agriculture, forest, grasslands, urban, surface waters, and
wetland areas); and NDVI. We then ran pCCA for each group
separately, using its variables as the constraints, while using all
other variables as conditioning variables [13,21,22,41]. In addition
we analyzed each individual variable as the constraint, using all
other variables as conditioning variables, in order to differentiate
the various variables within each group. To calculate the amount
of variance in species composition explained by each variable and
each group, we divided the inertia of each group in each sampling
unit by the overall inertia in the respective sampling unit, and
multiplied it by 100. Total inertia is an expression of the amount of
variance in the species data within the sampling units [22], and
individual inertia is equivalent to the amount of variance that is
related solely to the specific variable or group of variables, after
accounting for the variance explained by other variables and the
interaction between the different variables [21].

In order to discriminate between the effect of effective
environmental gradient length and the amount of variance in
species composition that is explained by that gradient, we
calculated effective gradient lengths of the different variable
groups within the sampling units. Effective gradient length is
related to the amount of variance of a variable within the entire
dataset. We calculated the effective gradient length by standard-
izing all variables so that they ranged between 0 and 1, and
calculating the cumulative standard deviation within each variable
group, which is the standard deviation in each variable group

Table 1. Environmental variables used in the analyses, and their source.

Variable name	Description	Source
Temperature		Worldclim [38]
Temperature seasonality	Standard deviation of monthly temperature values	
Precipitation		
Precipitation seasonality	Coefficient of variation of monthly precipitation values	
Altitude		Worldclim [38]
Alt_Range		
NDVI		MODIS – http://glcf.uniacs.umd.edu/data/ndvi
Pop-Density	Population density	FAOGeoNetwork http://www.fao.org/geonetwork/em/mainhome
Urban*	Urban area	
Forestry*	Forest	
Open-Herbaceous*	Herbaceous vegetation	
Agriculture*	Agricultural area	
Water*	Large water body	
Wettland*	Wetland area	
Distance to Urban	Distance to nearest urban area calculated at a fine resolution (0.0083°) and averaged for each grid-cell	Data were extracted from ESRI data files

*denotes values based on % coverage of categorical variables.
doi:10.1371/journal.pone.0025440.t001

Table 2. A list of sampling units used in the analyses, including grain size (side length in km) and extent (area in km²).

Scale	grain (km)	extent (km²)	valid samples (#)
1	4.47	20,480	9
2	6.32	40,960	12
3	8.94	81,920	12
4	12.65	163,840	14
5	17.89	327,680	13
6	25.30	655,360	9
7	35.78	1,310,720	7
8	50.60	2,621,440	7
9	71.55	5,242,880	3
10	101.19	10,485,760	2

doi:10.1371/journal.pone.0025440.t002

across all scales. Next we correlated that group's variance with the amount of variance that was explained by that group. High correlation coefficient values indicate strong effect of effective gradient length while low correlation values suggest effect of scale independent of differences in gradient length. In addition, in order to understand the sharp decrease in the amount of species composition variance explained by land-cover variables within the LULC variable group, we calculated the mean value, over the five different categories of land-cover (agriculture, forestry, open-herbaceous, urban and water), and then over the different units, of the standard deviation in those variables. A decrease in that value would indicate that the level of variance within the units is decreasing, i.e. each unit is composed of more components. A sharp decrease in intra-unit variance of land-cover variables should manifest as a decrease in the explanatory power of that group, in explaining the variance in species composition.

Supporting Information

Appendix S1 **An explanation of the concept of simultaneous alteration of grain and extent when using multiple spatial scales in ecological studies.**
(DOCX)

Author Contributions

Conceived and designed the experiments: RK YC. Performed the experiments: RK. Analyzed the data: RK AB-M. Contributed reagents/materials/analysis tools: RK YC. Wrote the paper: RK YC.

References

1. Shmida A, Wilson M (1985) Biological determinants of species diversity. Journal of Biogeography. pp 1–20.
2. Levin SA (2000) Multiple scales and the maintenance of biodiversity. Ecosystems 3: 498–506.
3. Rahbek C, Graves G (2001) Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences 98: 4534.
4. Whittaker RJ, Willis KJ, Field R (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. Journal of Biogeography 28: 453–470.
5. Rosenzweig ML (1995) Species diversity in space and time. Cambridge: Cambridge University Press. 436 p.
6. MacArthur RH (1972) Geographical ecology: patterns in the distribution of species. New York: Harper and Row.
7. Allen AP, Brown JH, Gillooly JF (2002) Global biodiversity, biochemical kinetics, and the Energetics-Equivalence Rule. Science 297: 1545–1548.
8. Colwell RK, Lees DC (2000) The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology & Evolution 15: 70–76.
9. Hubbell S (1997) A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs 16: 9–21.
10. Patrick R (1963) The structure of Diatom communities under varying ecological conditions. Annals of the New York Academy of Sciences 108: 359–365.
11. Pyron R, Burbrink F (2009) Can the Tropical Conservatism Hypothesis explain temperate species richness patterns? An inverse latitudinal biodiversity gradient in the New World snake tribe Lampropeltini. Global Ecology and Biogeography 18: 406–415.
12. Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73: 1045–1055.
13. Legendre P, Borcard D, Peres-Neto P (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs 75: 435–450.
14. May R (1984) An overview: real and apparent patterns in community structure. Ecological communities: conceptual issues and the evidence. PrincetonNew Jersey: Princeton University Press. pp 3–16.
15. Grand J, Cushman SA (2003) A multi-scale analysis of species-environment relationships: breeding birds in pitch pine-scrub oak (Pinus rigida-Quercus ilicifolia) community. Biological Conservation 112: 307–317.
16. Grand J, Mello MJ (2004) A multi-scale analysis of species-environment relationships: rare moths in a pitch pine-scrub oak (Pinus rigida-Quercus ilicifolia) community. Biological Conservation 119: 495–506.
17. Jones M, Tuomisto H, Borcard D, Legendre P, Clark D, et al. (2008) Explaining variation in tropical plant community composition: influence of environmental and spatial data quality. Oecologia 155: 593–604.
18. Rodriguez J, Hortal J, Nieto M (2006) An evaluation of the influence of environment and biogeography on community structure: the case of Holarctic mammals. Journal of Biogeography 33: 291–303.
19. Svenning J, Skov F (2005) The relative roles of environment and history as controls of tree species composition and richness in Europe. Journal of Biogeography 32: 1019–1033.
20. Vieira MV, Olifiers N, Delciellos AC, Antunes VZ, Bernardo LR, et al. (2009) Land use vs. fragment size and isolation as determinants of small mammal composition and richness in Atlantic Forest remnants. Biological Conservation 142: 1191–1200.
21. Cushman SA, McGarigal K (2002) Hierarchical, multi-scale decomposition of species-environment relationships. Landscape Ecology 17: 637–646.
22. Ter Braak CJF (1986) Canonical corespondence analysis: A new eigenvector technique for multivariate direct gradient analysis. Ecology 67: 1167–1179.
23. Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405: 243–253.
24. Allen T, Hoekstra T (1992) Toward a Unified Ecology. New York: Columbia University Press.
25. Willig M, Kaufman D, Stevens R (2003) Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution, and Systematics 34: 273–309.
26. Wiens J (1989) Spatial scaling in ecology. Functional Ecology 3: 385–397.
27. Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43: 393–404.
28. Graham C, Ferrier S, Huettman F, Moritz C, Peterson A (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19: 497–503.
29. Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14: 401–413.
30. Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, et al. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography 35: 105–116.
31. Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. (2006) Novel methods improve prediction of species' distributions from occurence data. Ecography 29: 129–151.
32. Kent R, Carmel Y (2011) Presence-only vs. Presence-absence data in species composition determinant analyses. Diversity and Distributions 17: 474–479.
33. Wilson RJ, Thomas CD, Fox R, Roy DB, Kunin WE (2004) Spatial patterns in species distributions reveal biodiversity change. Nature 432: 393–396.
34. Pianka ER (1966) Latitudinal gradients in species diveristy: a review of concepts. The American Naturalist 100: 33–46.
35. Kadmon R, Danin A (1999) Distribution of plant species in Israel in relation to spatial variation in rainfall. Journal of Vegetation Science 10: 421–432.
36. GBIF (2008) GBIF Training Manual 1: Digitisation of History Collections Data, version 1.0. Copenhagen: Global biodiversity information facility. 518 p.
37. ESRI (1999) ArcView GIS. 8.3 ed ESRI.

38. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolates climate surfaces for global land areas. International Journal of Climatology 25: 1965–1978.

39. Oksanen J, Blanchet GF, Kindt R, Legendre p, O'Hara RB, et al. (2011) vegan: Community Ecology Package. R package version 117-8. http://CRANR-projectorg/package = vegan.

40. Team RDC (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing http://wwwR-projectorg/. Vienna, Austria.

41. Ter Braak CJF, Verdonschot PFM (1995) Cannonical correspondence analysis and related multivariate methods in aquatic ecology. Aquatic Sciences 57: 255–289.