Article:

Gavish, Y orcid.org/0000-0002-6025-5668 and Ziv, Y (2017) Species-occupancy distribution removes an excessive parameter from species-area relationship. Journal of Biogeography, 44 (4). pp. 937-949. ISSN 0305-0270

https://doi.org/10.1111/jbi.12869

Article type:

Original article

Title:

Species-Occupancy Distribution Removes Excessive Parameter from Species-Area Relationship

Short Title:

SOD removes excessive SAR Parameter

Authors:

Yoni Gavish[1, *, #a], Yaron Ziv[1]

Affiliations:

[1] Spatial Ecology Lab, Department of Life Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva, 84105, Israel.

[#a] Current Address: School of biology, University of Leeds, Leeds, LS2 9JT, United Kingdom

[*] Corresponding author: Yoni Gavish, E-mail: gavishyoni@gmail.com

Word count:

Abstract + main text + references:	7027 words
Tables:	2
Figures:	4
Appendices:	3
Estimated number of pages:	12

ABSTRACT

Aim Although species-occupancy distributions (SODs) and species-area relationships (SARs) arise from the two marginal sums of the same presence/absence matrices, the two biodiversity patterns are usually explored independently. Here, we aim to unify the two patterns for isolate-based data by constraining the SAR to conserve information from the SOD.

Location Widespread

Methods Focusing on the power-model SAR, we first developed a constrained form that conserved the total number of occupancies from the SOD. Next, we developed an additive-constrained SAR that conserve the entire shape of the SOD within the power-model SAR function, using a single parameter (the slope of the endemics-area relationship). We then relate this additive-constrained SAR to multiple-sites similarity measures, based on a probabilistic view of Sørensen similarity. We extend the constrained and additive-constrained SAR framework to 23 published SAR functions. We compare the fit of the original and constrained forms of 12 SAR functions using 154 published datasets, covering various spatial scales, taxa and systems.

Main conclusions In all 23 SAR functions, the constrained form had one parameter less than the original form. In all 154 datasets the model with the highest weight based on the corrected Akaike Information Criteria (wAICc) had a constrained form. The constrained form received higher wAICc than the original form in 98.79% of valid pairwise cases, approaching the wAICc expected under identical log-likelihood. Our work suggests, both theoretically and empirically, that all SAR functions may have one unnecessary parameter, which can be excluded from the function without reduction in goodness-of-fit. The more parsimonious constrained forms are also easier to interpret as they reflect the probability of a randomly chosen occupancy

to be found in an isolate. The additive-constrained SARs accounts for two complimentary turn-over components of occupancies: turnover between species and turnover between sites.

Keywords Biodiversity patterns; islands; Jaccard; landscape; macroecology; multiple-sites similarity; Sørensen; spatial ecology; occupancy-frequency distribution; patches.

INTRODUCTION

Studying biodiversity distribution patterns characterizes a major exploration line in contemporary ecology due to both basic and applied needs. This exploration requires biodiversity data collection of diverse species located at different spatial extents. Consequently, most biodiversity studies end up with a species-by-site table filled with presence/absence data (hereafter we refer to a presence of a species in a site as occupancy). Summing this community matrix for each site over all species yields the total number of species sampled within each site (Fig. 1A). Similarly, when summed for each species over all sites, the marginal sums yield the number of sites in which a species occurred (i.e., the species occupancy level). These two sets of marginal sums give rise to two important biodiversity patterns -- the species occupancy distribution (SOD, the number of species that occurred in each occupancy level, e.g. McGeoch & Gaston, 2002; Jenkins, 2011) and the species-area relationship (SAR, the change in species richness with a change in area).

SARs and SODs can be constructed from data collected in various ways, including nested quadrats, quadrats in a contiguous grid, quadrats in a non-contiguous grid, and non-overlapping areas of various sizes (types I-IV sensu Scheiner, 2003, respectively). Here we focus on type IV SARs, and following Tjørve & Turner (2009), we refer to the sites as isolates (non-overlapping sites with biologically or environmentally defined borders that differ from one another in various attributes such as area, shape, heterogeneity and spatial context). Most SOD studies focused on contiguous or non-contiguous equal-sized quadrat (type II or III), in which SOD shape is highly dependent upon the choice of grain size, while type IV SOD, on which we focus here, has the advantage of working on naturally occurring grains: the isolates. Despite numerous studies of SARs in island-like systems, we are not aware of any manuscript that focused on type IV SODs.

Indeed, among the two biodiversity patterns, SARs have received the most attention, with at least 23 mathematical functions suggested to describe the pattern (Tjørve, 2003, 2009; Williams *et al.*, 2009). In fact, SARs are one of the most fundamental patterns of ecology. Empirically, SARs have been explored in numerous study systems, covering a wide range of scales, focusing on diverse taxa, and using various methods (Rosenzweig, 1995; Scheiner, 2003; Drakare *et al.*, 2006; Triantis *et al.*, 2012). SARs exhibit a consistent pattern: the number of species increases with area, thus considered as a general law of ecology (Rosenzweig, 1995). SARs have also been the subject of extensive theoretical research, either aiming to explain their properties or as a starting rule from which other patterns emerge (e.g., Rosenzweig & Ziv, 1999). The generality and centrality of SARs triggered their usage in applied ecology. Among others, SARs are used to estimate extinction debts (Brooks *et al.*, 2002; Kuussaari *et al.*, 2009), identify biodiversity hotspots (Myers *et al.*, 2000; Gavish, 2011), and optimize reserve design (Bascompte *et al.*, 2007; Tjørve, 2010; Gavish *et al.*, 2012).

In contrast, SODs remained relatively unexplored, perhaps due to the complexity of shapes they can take. Unlike SARs, which are usually described by convex functions with no asymptote (Triantis *et al.*, 2012), SODs may be unimodal, bimodal, random, or uniform, and their modes may occur for satellite (rare), central, or core (common) species (McGeoch & Gaston, 2002;

Jenkins, 2011; Hui, 2012). Furthermore, bimodal SODs may be symmetrical or asymmetrical, and if asymmetric they may have a stronger or weaker mode for rare or common species. Until recently, only one method (Tokeshi, 1992), based on comparison of the size of the satellite and core modes to an expected null model, was used to describe SOD's shape. Recently, Jenkins (2011) introduced the ranked species-occupancy curves (rSOC) as an alternative method and Hui (2012) clarified the direct link between the two patterns. Similar to species-abundance distribution and ranked abundance curves, SOD and rSOC are two alternative ways to present the same information.

Although SODs and SARs arise from the two marginal sums of the same presence/absence table, the two patterns were rarely explored simultaneously (but see: Hui & McGeoch, 2014; Pan, 2015). In fact, in most cases they were explored simultaneously only when both were derived from species abundance data, either through null models (Coleman, 1981), neutral models (Hubbell, 2001), or metapopulation-based models (Ovaskainen & Hanski, 2003). The aim of this paper is to develop the direct link between SODs and SARs for island-like systems within a single framework. Within this framework, the shape of the SOD itself can be explored in relation to species traits, thereby providing a more mechanistic understating of the SAR. Furthermore, mechanistic SAR hypotheses such as the transient hypothesis (MacArthur & Wilson, 1967), rescue effects (Brown & Kodric-Brown, 1977), target area effects (Gilpin & Diamond, 1976) and small island effects (Lomolino, 2000), are mediated through changes in species occupancy levels.

MATERIAL AND METHODS

Mathematical developments

We develop the direct link between SOD and SAR by constraining the SAR to conserve the data encompassed by the SOD. Starting with the power-model (Arrhenius, 1921), we first developed the constrained-SAR model, by forcing the SAR to conserve the observed total number of occupancies (thereby canceling-out one of the power-model parameters). Then, in the additive-constrained SAR model, we fit a separate constrained-SAR model to the species of each occupancy level, and then sum the results over all occupancy levels. By describing the change in SAR parameters with occupancy level we provide a novel one-parameter SAR function that predicts not only the shape of the global SAR, but also the SAR of each occupancy level, while conserving the entire shape of the SOD. The parameter of this additive model is the slope of the endemics area relationship. Subsequently, we relate the SOD to multiple-sites similarity indices and generalize to 23 known SAR functions.

The power-model SAR

We start with a presence/absence matrix of M species in N isolates (Fig. 1A, see notations in Fig. 1F). Each species is notated with m (in the range $\{1,2,\ldots,M\}$), each isolate with i $\{1,2,\ldots N\}$ and each entry as $O_{i,m}$ (that can take the value of 1 or 0). The observed number of species in isolate i (hereafter, S_i) is the sum of $O_{i,m}$ over all M species, and if A_i is the area of isolate i, the global SAR can be constructed (Fig. 1D). The occupancy level of species m (hereafter, j_m) is the sum of $O_{i,m}$ over all N isolate (thus j_m is in the range $\{1,2,\ldots,N\}$). The SOD explores how the number of species in occupancy level j (hereafter R_j) changes with j (Fig. 1C). Thus, summing R_j over all occupancy levels (all j in the range $\{1,2,\ldots,N\}$) yields M. The presence/absence matrix cab be restructured as a square $N\times N$ matrix, with the number of presences from each occupancy-level that were found in each isolate (hereafter $S_{i,j}$, Fig. 1B). The total number of occupancies can be estimated in three ways: by summing $O_{i,m}$ over all M species, by summing S_i over all N isolates, and by summing $j{\cdot}R_j$ over all occupancy levels.

Given the observed S_i and A_i, the original power-model SAR (Arrhenius, 1921) takes the form:

$$E(S_i)_{orig.} = c \cdot A_i^z \qquad \text{(E1)}$$

With $E(S_i)_{orig.}$ as the number of species predicted for isolate i by the power-model AND c and z as scaling parameters. The total number of occupancies predicted by the power-model is the sum of equation 1 over all n isolates. To constrain the power-model SAR such that it will conserve the observed total number of occupancies, we set $\Sigma_j(j{\cdot}R_j)=\Sigma_i c{\cdot}Ai^z$ and multiply equation 1 by $\Sigma_j(j{\cdot}R_j)/\Sigma_i c{\cdot}Ai^z$:

$$E(S_i)_{cons.} = c \cdot A_i^z \times \frac{\sum_{j=1}^{N}(j \cdot R_j)}{\sum_{i=1}^{N}(c \cdot P_i^z)} = \sum_{j=1}^{N}(j \cdot R_j) \cdot \frac{A_i^z}{\sum_{i=1}^{n} A_i^z} \qquad \text{(E2)}$$

with $E(S_i j)_{cons.}$ being the expected number of species in isolate i according to the constrained power-model. Adding the total number of occurrences constraint to the power-model SAR eliminates parameter c, which allows the predicted sum of occupancies to differ from the observed one, leaving only parameter z. Furthermore, $Ai^z/\Sigma_i Ai^z$ is the probability of a single occupancy to be found in isolate i. Although this constrain can be employed with no knowledge of the SOD, we base it on the SOD's arguments to exemplify the effect of focusing only on species from a single occupancy level. In fact, when equation 2 is fitted only to the subset of species from occupancy level j (Fig. 1E), we get:

$$E(S_{i,j})_{cons.} = \left(j \cdot R_j\right) \cdot \frac{A_i^{z_j}}{\sum_{i=1}^{n} A_i^{z_j}} \qquad \text{(E3)}$$

with $E(S_{i,j})_{cons.}$ being the expected number of species from occupancy level j in isolate i, and z_j the slope of the SAR of occupancy level j. If we assume that the SAR of all occupancy levels can be described by a power-model (see below) then equation 3 can be summed to produce a second approximation of the global SAR:

$$E(S_i)_{add.cons.} = \sum_{j=1}^{N}[(j \cdot R_j) \cdot \frac{A_i^{z_j}}{\sum_{i=1}^{n} A_i^{z_j}}] \quad \text{(E4)}$$

with $E(S_i)_{add.cons.}$ being the expected number of species in isolate *i* according to the additive-constrained power-model. We are not aware of any publication that explores the change of z_j with *j*, which we term '*z-occupancy curves*'. However, endemics-area relationships (the SAR when including only species that are endemic to a single isolate, i.e., j=1) usually have relatively high z_j values (Rosenzweig, 1995; Triantis *et al.*, 2008). Eventually, z_j values for $j=N$ are, by definition, zero (the species occur on all isolates, Fig. 1E)). In addition, equation 2 and 3 can estimate the maximal value of *z* that will ensure that none of the isolates contains more species than the actual size of the species pool ($\Sigma_j R_j$), or the number of species in occupancy level *j* (R_j), denoted as z_{max} and $z_{j,max}$, respectively. When setting the monotonically increasing (for z>0) equation 2 and 3 to equal $\Sigma_j R_j$ or R_j (respectively) and solving for the largest isolate (here, isolate $i=N$) we get:

$$\frac{A_N^{z_{max}}}{\sum_{i=1}^{N} A_i^{z_{max}}} = \frac{\sum_{j=1}^{N} R_j}{\sum_{j=1}^{N}(j \cdot R_j)} \quad \text{(E5)}$$

$$\frac{A_N^{z_{j,max}}}{\sum_{i=1}^{N} A_i^{z_{j,max}}} = \frac{R_j}{(j \cdot R_j)} = \frac{1}{j} \quad \text{(E6)}$$

This means that z_{max} is the value of *z* for which the probability of randomly drawn occupancy to be in the largest isolate equals the inverse of the mean occupancy level. Although z_j is unbounded for j=1, z_j of all other occupancy levels have a maximal value ($z_{j,max}$) that is independent of the number of species and depends mainly on the area distribution (A_i values, equation 6). The maximal values result in a decreasing function when plotting $z_{j,max}$ against *j*.

Therefore, we expect z_j to decrease with j in a predictable manner, according to a function $F(z_j|j)$ that intersects the abscissa at $j=n$. Consequently, we get:

$$E(S_i)_{add.cons.} = \sum_{j=1}^{N}[(j \cdot R_j) \cdot \frac{A_i^{F(z_j|j)}}{\sum_{i=1}^{N} A_i^{F(z_j|j)}}] \qquad \text{(E7)}$$

Although various functions may describe the shape of the z-occupancy curve, we focused here on the form given in equation 8, and when plugging it into equation 7 we get:

$$F(z_j|j): \quad z_j = a \cdot (1 - log_N j) \qquad \text{(E8)}$$

$$E(S_i)_{add.cons.} = \sum_{j=1}^{N}\left[(j \cdot R_j) \cdot \frac{A_i^{a\cdot(1-log_N j)}}{\sum_{i=1}^{N} A_i^{a\cdot(1-log_N j)}}\right] \qquad \text{(E9)}$$

We chose equation 8 for three main reasons. First, it is an exponential decay function $z_j=a-b\cdot\ln(j)$ that intersects the point $(N,0)$ (such that $b=a/\ln(N)$), and thus always predict z_j values of 0 when $j=N$. Second, preliminary analysis of several datasets revealed it to be a good candidate model. Third, its only parameter (a) is biologically meaningful- it is the z value of the endemic-area relationship. In fact, equation 9 is a SAR function that incorporates the entire observed shape of the SOD into the SAR, provides predictions for the overall SAR, as well as for the SAR of each occupancy level and has a single, ecologically-meaningful parameter (a). Other $F(z_j|j)$ functions with more complex shapes or with better theoretical grounds can be developed, perhaps after more detailed exploration of the shape of z-occupancy curves is carried.

Finally, if the constrained and additive-constrained model provide comparable predictions, and equation 2 and 9 are divided by the total number of occupancies, we get:

$$\frac{S_i}{\sum_{j=1}^{N}[(j\cdot R_j)]} = \frac{A_i^z}{\sum_{i=1}^{n} A_i^z} \cong \sum_{j=1}^{N}\left[\left(\frac{j\cdot R_j}{\sum_{j=1}^{N}[(j\cdot R_j)]}\right) \cdot \left(\frac{A_i^{a\cdot(1-log_N j)}}{\sum_{i=1}^{N} A_i^{a\cdot(1-log_N j)}}\right)\right] \qquad \text{(E10)}$$

so that the probability of a randomly chosen occupancy to be found in isolate i is similar (up to the error associated with the models) to the sum over all occupancy levels of the multiplication of two probabilities. The first is the probability of a randomly chosen occupancy to be from occupancy level *j*. The second is the conditional probability of this occupancy to be found in isolate *i*, given the SAR of occupancy level *j*. The two probabilities reflect the two marginal sums of the presence/absence data table. In fact, the first probability is the generalization of Sørensen probabilities to multiple-sites, as explained in the next section.

Similarity indices, SODs and weighted SODs

The most commonly used pairwise similarity indices of binary data are *Jaccard* and *Sørensen*. Let S_1 and S_2 be the number of species in isolates 1 and 2, respectively, and let S_{shared} be the number of species shared by the two isolates. *Jaccard* similarity can be expressed as $S_{shared}/(S_1+S_2-S_{shared})$, while *Sørensen* similarity is $2\cdot S_{shared}/(S_1+S_2)$ (Chao *et al.*, 2005). Therefore, *Jaccard* similarity is the ratio of the number of species in occupancy level j=2 and the total number of species. *Sørensen* similarity is the ratio of the number of occupancies in occupancy level j=2 and the total number of occupancies. When viewed as probabilities, *Jaccard* is the probability of randomly selecting a species that is shared by the two isolates. *Sørensen* is the probability of randomly selecting an occupancy from a species shared by two isolates. That is, when n=2, *Jaccard* can be expressed as $R_2/(R_1+R_2)$, while *Sørensen* can be expressed as $2\cdot R_2/(1\cdot R_1+2\cdot R_2)$. *Jaccard* and *Sørensen* dissimilarities are the complimentary of the indices to 1, which can be expressed as $R_1/(R_1+R_2)$ and $1\cdot R_1/(1\cdot R_1+2\cdot R_2)$, respectively. Thus, when there are only two isolates, the additive-constrained SAR (equation 10) explicitly contains *Sørensen* similarity and dissimilarity as weights.

The SOD summarizes the change in R_j with j. If we standardize the SOD by dividing it by $\Sigma_j R_j$, we get for each occupancy level the term $R_j/\Sigma_j R_j$, which is the generalization of *Jaccard* probabilities into multiple isolates. A weighted form of the SOD (wSOD, Fig. 1C) summarizes the change in $j{\cdot}R_j$ with j. When standardizing the wSOD by dividing it with $\Sigma_j(j{\cdot}R_j)$, we get for each occupancy level the term $j{\cdot}R_j/\Sigma_j(j{\cdot}R_j)$, which is the generalization of *Sørensen* probabilities to multiple isolates. Since the basic unit of type IV SARs is occupancy and not species, *Sørensen* probabilities are more relevant to the study of type IV SARs. Therefore, when there are more than two sites, equation 10 incorporates the generalization of *Sørensen* probabilities into the general SAR framework.

We suggest that summary statistics of the standardized SOD and wSOD can be considered as measures of beta diversity, since their constituting values may serve as the building blocks for multiple-sites similarity indices. Such multiple-sites similarity measures may differ from one another in their treatment of the difference between species in occupancy level. For example, the strictest definition of *Jaccard* multiple-sites similarity may be the proportion of species that are found in all isolates from the total number of species, i.e., $R_n/\Sigma_j R_j$, and for *Sørensen,* the equivalent proportion of occupancies from the total number of occupancies, i.e., $n{\cdot}R_n/\Sigma_j(j{\cdot}R_j)$. The least strict may be the proportion of species/occupancies that are found in at least two isolates, i.e., $\Sigma_{j\neq 1}[R_j/\Sigma_j R_j]$ for *Jaccard*, and $\Sigma_{j\neq 1}[j{\cdot}R_j/\Sigma_j\, j{\cdot}R_j]$ for *Sørensen*. In fact, if w_j is the weight given to occupancy level j in the multiple-sites similarity measure (such that $0 \leq w_j \leq 1$), then a general multiple-sites similarity of *Jaccard* and *Sørensen*, which still conserves the 2 isolates interpretation as the proportion of species or occupancies may be:

$$Jac_{mult} = \frac{\sum_{j=1}^{N}[w_j \cdot R_j]}{\sum_{j=1}^{N}[R_j]} \qquad \text{(E11)}$$

$$Sør_{mult} = \frac{\sum_{j=1}^{N}[w_j \cdot j \cdot R_j]}{\sum_{j=1}^{N}[j \cdot R_j]} \qquad \text{(E12)}$$

with $w_1=w_2=\ldots=w_{N-1}=0$ and $w_N=1$ for the most strict example while $w_1=0$ and $w_2=w_3=\ldots=w_N=1$ for the least strict example. A more interesting option for the weights may be the proportion of isolates pairs in which a species co-occur, resulting with

$$Jac_{mult} = \sum_{j=1}^{N}\left[\frac{j(j-1)}{N(N-1)}\cdot\frac{R_j}{\sum_{j=1}^{N}[R_j]}\right] \quad \text{(E13)}$$

$$Sør_{mult} = \sum_{j=1}^{n}\left[\frac{j(j-1)}{N(N-1)}\cdot\frac{j\cdot R_j}{\sum_{j=1}^{N}[j\cdot R_j]}\right] \quad \text{(E14)}$$

which converges to *Jaccard* and *Sørensen* similarities for n=2, while satisfying $w_1=0$ and $w_N=1$ and keeping the original probabilistic interpretation of the indices. We note though, that the multiple-sites similarity indices themselves are not incorporated directly into the SAR, but rather they are built by the same building blocks as the SAR. We further note that published multiple-sites versions of *Jaccard* (Baselga, 2012) and *Sørensen* similarities (Baselga, 2010) can also be restructured using terms from the SOD as:

$$Jac_{mult,Bas} = \frac{\left[\sum_{j=1}^{N}(j\cdot R_j)-\sum_{j=1}^{N}R_j\right]}{\left[\sum_{j=1}^{N}(j\cdot R_j)-\sum_{j=1}^{N}R_j\right]+\sum_{j=1}^{N}\left(R_j\cdot j\cdot(N-j)\right)} = \frac{\sum_{j1}^{N}[(j-1)\cdot R_j]}{\sum_{j1}^{N}[(j-1+j(N-j))\cdot R_j]} \quad \text{(E15)}$$

$$Sør_{mult,Bas} = \frac{2\cdot\left[\sum_{j=1}^{N}(j\cdot R_j)-\sum_{j=1}^{N}R_j\right]}{2\cdot\left[\sum_{j=1}^{N}(j\cdot R_j)-\sum_{j=1}^{N}R_j\right]+\sum_{j=1}^{N}\left(R_j\cdot j\cdot(N-j)\right)} = \frac{\sum_{j1}^{N}[(2j-2)\cdot R_j]}{\sum_{j1}^{N}[(2j-2+j(N-j))\cdot R_j]} \quad \text{(E16)}$$

yet, such extensions to multiple sites do not conserve the total number of species or occupancies in the denominator, and therefore loses the probabilistic interpretation of *Jaccard* and *Sørensen* similarities. Furthermore, although the contribution to the similarity measure increases with occupancy level in the numerators of $Jac_{mult,Bas}$ and $Sør_{mult,Bas}$, the denominator reaches a maximum value for $j=(N+1)/2$ and $j=(N+2)/2$, respectively. If the SOD is indeed the unifying concept between beta-diversity and SARs, we suggest focusing on multiple-sites similarity indices that conserve the probabilistic interpretation of the SOD and the wSOD. In addition to the ecological meaning that of the probabilities, it opens a possible direction to incorporate the effect of unsampled species to multiple-site similarity indices and SARs, as shown for pairwise similarity by Chao *et al.* (2005).

Other SAR functions

Constrained SARs and additively constrained SARs can be based on any SAR function (Tjørve, 2003, 2009; Williams *et al.*, 2009; Triantis *et al.*, 2012). Repeating the steps that led from equation 1 to 2 for other SAR functions has a similar effect – all SAR functions lose one of their parameters (Table 1). Similar additive forms to those shown here for the power-model can be developed for all other SAR functions. Therefore, all SAR functions may have one unnecessary parameter that can be excluded, apparently, without loss of statistical power.

Empirical analysis of 154 datasets

We explored 154 published datasets (see Appendix S2) to examine whether a parameter can be dropped without loss of goodness-of-fit if the SAR is constrained. The datasets cover various spatial scales (from 6 m^2 isolates to inter-provincial SARs), taxa (fungi, plants, invertebrates and vertebrates) and systems (inter-provincials, ecoregions, true islands, fragmented terrestrial landscapes, etc.). Before fitting any model and to ease the search for appropriate starting values for parameters, we first standardized the area units to relative area: $P_i=A_i/\Sigma_i A_i$ (and $\Sigma_i P_i=1$). We fitted each dataset with the original and constrained forms of the twelve functions given in bold face in Table 1, a total of 24 functions. Non-linear least square regressions (using the Levenberg-Marquardt convergence algorithm) were used to fit each dataset with the 24 models, and various parameter-starting values were used to avoid local minima. After convergence, for the original and constrained forms of each SAR function, the estimated parameters of the form that resulted with lower residuals sum-of-squares (RSS) were used as the starting parameters of the second form in an additional non-linear regression and the newly estimated parameters were kept if the fit was improved.

After fitting the 24 models, we calculated for each model the corrected Akaike Information Criteria (AICc; Burnham & Anderson, 2002; see Appendix S1 in Supporting Information). Next, AICc weights (wAICc$_g$, with *g* being the name of the model out of *G* models) were calculated for the entire set of 24 models (i.e., *G*=24). We then focused on each of the 12 SAR functions separately and estimated the wAICc$_g$ of the functions's original and constrained forms of each SAR function (i.e., 12 different sets, each with *G*=2). For the 12 sets, we estimated the expected wAICc for the special case in which the original and constrained form have identical log–likelihood and only differ in the number of parameters (Appendix S1). Finally, we applied a least-square linear regression of the observed wAICc of the constrained form against the expected value under identical log-likelihood and explored whether the confidence intervals of the intercept and slope overlapped with zero and one, respectively.

The 154 datasets were also used to explore the shape of *z*-occupancy curves. Firstly, for each dataset, we fitted equation 3 separately for the species from each occupancy level. This yielded the observed z_j values for every *j* for which some species were observed. Next, we fitted the observed *z*-occupancy curve with equation 8, while recording the explained variance and significance. For the datasets presented in Fig.3, we fitted equation 9 as well, and compared the predicted z occupancy curve to the fitted one. We further compared the AICc values and weights of the original power-model (equation 1), constrained power-model (equation 2) and additive-constrained power-model (equation 9).

Finally, for the 154 datasets we estimated the *Sørensen* multiple-sites similarity index based on equation 14. We used linear regression to explore the relation between the power-model z values and the multiple-sites similarity value. For this analysis we use only dataset where the power-model explained variance was larger than 0.25. All regression analyses were carried out with the minpack.lm package in R (R Development Core Team, 2014).

RESULTS

When comparing the 24 models, in all 154 datasets the model with the highest wAICc had a constrained form. In general, SAR functions with the highest wAICc usually had only two parameters in the original form (80% of datasets), had a convex shape (63%) and had no asymptote (64%). Indeed, in 32% of the datasets, the best SAR function had all three of these characters. The power-model had the highest wAICc for 25.4% of the datasets.

From a total of 1848 (12×154) combinations of SAR models and datasets, the non-linear regression achieved convergence for both the original and constrained forms in 1811 analyses. The constrained form received a higher wAICc than the original form in 1789 out of 1811 pairwise comparisons (98.79%, Table S3 in Appendix S2, Fig. 2). The wAICc of the constrained form approached the expected weight for the special case in which the original and constrained forms had identical log-likelihood (Fig. 2). For ten of the twelve SAR functions the confidence intervals of the intercept and slope of the linear regression between the observed and expected AICc weight of the constrained form overlapped with 0 and 1, respectively (Table S3). The two exceptions were the Monod and Negative Exponential SAR functions. However, in these two SAR functions, large deviation from the expected wAICc occurred in datasets that were not adequately described by the SAR function (Fig. S1 in Appendix S3).

The non-linear regression of observed z_j values against j according to the exponential decay function (equation 8) was statistically significant ($p<0.05$) for 138 of the 154 datasets. In some cases, a very clear decay pattern was evident (see a few examples in Fig. 3), while in others the pattern was not that clear. The decay of z_j with j was less well defined when the SAR pattern itself was weak or when the number of species was very low relative to the number of isolates (resulting in poor representativeness in many occupancy levels). The 25th, 50th, and 75th percentiles of parameter a of equation 8 – i.e., the slope of the endemic area relationship – were 0.43, 0.65, and 1.12, respectively. The explained variance of the regressions had a 25th, 50th and 75th percentiles of 0.29, 0.65, and 0.81, respectively. For the datasets presented in Fig. 3, the z values predicted for each occupancy level by fitting equation 9 as the general SAR function (red line) was highly correlated to the z values when fitting each occupancy level separately (black diamonds, equation 3). The predicted z values according to equation 9 was very similar to those achieved by fitting equation 8 to the fitted z values (black line). The additive-constrained SAR received higher AICc weights than the original power-model in 6 out of 9 datasets (table 2), and in two of these cases, it also out-performed the constrained power-model.

We found a statistically significant negative correlation between the power-model z (equation 2) and *Sørensen* multiple-sites similarity index (equation 14, Fig. 4). We observed a strong effect of the number of isolates on the z-similarity trend. Dataset with large number of isolates tended to have lower z values, and lower similarity values, probably since most species remain rare even when a large number of isolates are sampled.

DISCUSSION

We developed the constrained form of 23 known SAR functions, which forces the SAR to conserve the total number of occupancies (Table 1). For all SAR functions, constraining the SAR resulted in a decrease of one parameter in the number of function parameters. The meta-analysis of the 154 datasets revealed that the constrained forms outperformed the original ones. This is evident for the 154 datasets in the wAICc of the constrained form approaching its expected value for the special case in which the original and constrained forms have identical log-likelihood (Fig. 3). In the two SAR functions (Monod and Negative Exponential) for which some deviation from linear correlation were observed, the deviations mainly occurred in datasets for which the SAR function did not describe the pattern well, relative to other SAR function (Fig. S1, Appendix S3). Therefore, the deviations probably resulted from failure to converge to the same global minima, since many local minima have very similar log-likelihoods.

Consequently, for any given SAR function we have two competing models having similar predictions and log-likelihoods, with one of the models having fewer parameters than the other. The basic principle of parsimony requires us to prefer the model with fewer parameters, and therefore for each SAR function to prefer the constrained forms over the original ones. Considering the most common power-model SAR, the parameter which is canceled-out is parameter *c*, the ‘politically ignored’ parameter (sensu, Gould, 1979; Triantis *et al.*, 2012). Our results suggest that it is correctly ignored since it is an unnecessary fitting parameter that comes on the expense of the more informative, process-based component of the SOD. This parameter can be isolated from equation 2, to get: $c=\Sigma_j(j\cdot R_j)/\Sigma_i(A_i^z)$. Therefore, Lomolino (2000) statement that parameter *c* “*varies in a poorly understood manner among taxa and types of systems*” is not surprising, given that even when the area units are standardized, it is a function of the total number of occupancies, the number of isolates, the distribution of area between isolates and the second parameter *z*. Parameter c (and it’s above approximation) is usually interpreted as the number of species in one unit of area. The general SAR is then constructed by multiplying the number of species in one unit area by an area dependent function. This is still true for the constrained SAR (equation 2). However, we show here that it is also true for the constrained SAR of all other SAR models (table 1).

Removing a parameter from a widely used function may seem to present a small technical improvement. However, given its broad use and the importance of SARs for various applications, simplification of SAR models is crucial to understanding patterns and processes, since simpler models are easier to interpret. Although having more parameters allows better fit to data, parameters should be added if the additional goodness-of-fit is needed to better understand the pattern. For SARs, this does not seem to be the case. In fact, the proximity of the wAICc to the expected AICc weight under identical log likelihood (Fig. 2) suggest that the two forms have very similar goodness-of-fit.

By constraining the SAR we have shown that SAR represents the turnover of occupancies between isolates. Although SARs predict the number of species in each isolate, it is more correct to treat occupancy as their basic unit. To claim that the unit of SARs is species is similar to claiming that the unit of the abundance-area relationship (i.e., the total number of individuals per isolate) is species and not individuals. Accepting that SARs represent the turnover of occupancies between isolates suggests that SARs may also be affected by the second occupancies turnover component – i.e., the turnover of occupancies between species. This second component is captured by the SOD and the additive-constrained SAR.

The additive-constrained SAR (equations 7 and 9) sums over all occupancy levels the multiplication of two probabilities. The first is the probability that a random occupancy is from occupancy level j, and the second is the probability of this occupancy being in isolate i, given its occupancy level. The two probabilities represent the two turnover components of occupancies: ***i.*** turnover of occupancies between species, and ***ii.*** turnover of occupancies between isolates. The first turnover component relates to the shape of the wSOD, and as such to the extended Sørensen probabilities. The effect of this turnover component on the SAR's shape is evident in the relation between z and the multiple-sites similarity index (Fig. 4). The second relates to the shape of the *z*-occupancy curves.

Here we explored a very specific additive-constrained SAR that assumes a power-model at all occupancy levels. Of-course, similar to the general SAR, this might not be correct in all datasets. However, even under this strict assumption, the additive-constrained SAR outperformed the original power-model in six out of nine datasets (table 2), while providing excellent prediction to the actual shape of the z-occupancy curves (Fig. 3). Furthermore, even within a given dataset, different models may best describe SARs of different occupancy levels. Unfortunately, occupancy-specific SARs have never been explored before, with the exception of the endemics-area relationship that was mainly explored using a power-model (Triantis *et al.*, 2008). We predict that the best fitting SAR model will change in a consistent manner with occupancy level, (e.g., from a sigmoid curve, to power-law and then to linear models as occupancy increases). Alternatively, additive-constrained SARs can be based on single models with greater flexibility such as the two models suggested by Tjørve (2012).

The constrained form does not suggest any clear ecological interpretation of *z*, yet it is still unclear if any such interpretation will ever arise (Connor & McCoy, 2001; but see: Rosindell & Cornell, 2007; O'Dwyer & Green, 2010; Grilli *et al.*, 2012). Mathematically, *z* is a scaling parameter that changes the proportion $A_i^z/\Sigma_i(A_i^z)$, relative to z=1, for which each isolate receives a proportion from the total occupancies that is identical to its relative area. Therefore, the ecological interpretation added in the constrained form is not in the meaning of *z*, but rather in the meaning of the proportion $A_i^z/\Sigma_i(A_i^z)$. Since equation 2 is structured as the total number of occupancies multiplied by this proportion we can interpret it as the probability of a randomly drawn occupancy to be from isolate *i*. The proportion also explains why *z* has maximal values (equation 5), as no isolate can receive more occupancies than the number of species. This restriction on the values of *z* may be the reason why various theories, in spite of their very different underlying assumptions, also predict it to have a restricted range (e.g., Preston, 1962). Note, that z_{max} cannot be estimated from the original form of SAR, since many different combinations of *c* and *z* may satisfy the maximal proportion criteria. Similar maximal values for parameters can be found for six other SAR functions that have a single parameter in their constrained form (table 1).

In a wider perspective, we used the SOD as a pre-defined pattern that is plugged into the SAR. However, species occupancy levels may be explored in relation to any of the species traits. For example, species dispersal ability is likely to effect the intensity of rescue effects and recolonization rates. Thus, species with higher dispersal abilities are likely to be found in more isolates than species with lower dispersal ability. Alternatively, species with high competitive ability are likely to persist longer in isolates once they are colonized. Thus, species with high competitive abilities are also lankly to occur on more isolates than species with poor competitive abilities. Now, if we can model the probability of a species to have a certain occupancy level (*j*) based on its' dispersal and competitive abilities, we can sum these probabilities over all species for a given *j* to represent R_j. These R_j can then be used in equation 2 , 4 or 9 above. In such analyses, the parameters linking species occupancy levels to species dispersal and competitive abilities (or any other relevant trait) can be estimated simultaneously with the parameters of the SAR, thereby allowing a more mechanistic understanding of SARs. The incorporation of species traits directly into SAR functions may compliment other relations between SARs and traits, such as exploring SAR's slope for various trait values (Franzen *et al.*, 2012), substituting species richness with functional trait diversity as the dependent variable (Whittaker *et al.*, 2014) or building SARs from species-specific incidence functions (Ovaskainen & Hanski, 2003). The advantage here is that one does not need to know in advance the effect of these traits on the probability to occur on j isolates, and can learn on it from the SAR function.

Similarly, SARs are only one of the biodiversity patterns that relate the number of species per isolate with one of the isolate's attributes. Other attributes may include, for example, habitat heterogeneity, degree of isolation or the availability of resources (e.g., species energy relationship). Probably, many of the mathematical functions used to describe SARs (Table 1) may be used to describe other biodiversity patterns, such as species-connectivity relationships and species-heterogeneity relationships. The constrained and additive-constrained forms may be used to explore any of these biodiversity patterns.

Here we show, both theoretically (Table 1) and empirically (Fig. 2), that all known SAR functions have one unnecessary parameter. Simplification of models is crucial to understanding patterns and processes, since simpler models are easier to interpret. By constraining the SAR, we have clarified its basic units, united all functions to a similar general structure (Table 1), introduced Sørensen probabilities into the SAR framework and linked the two sides of presence/absence tables (Fig. 1). SARs are fundamental to the development and testing of many ecological theories (McGill, 2010) and play an important role in conservation and management, including identifying biodiversity hotspots (Guilhaumon *et al.*, 2008) and predicting the effect of habitat loss on species richness (Rosenzweig *et al.*, 2012; Keil *et al.*, 2015). Hopefully, our work will shed new light on this important biodiversity pattern.

ACKNOWLEDGMENTS

We thank Itamar Giladi and Dror Kapota for important remarks and suggestions. Burt Kotler, Zvika Abramsky, Ofer Ovadia, Bob Holt and Bill Kunin provided invaluable comments on earlier versions of this manuscript. Special thanks to Arnost L. Sizling for his extremely constructive review, and to two additional anonymous reviewers and the editor, François Guilhaumon for their comments. YZ was supported by the Israel Science Foundation (grant 751/09).

REFERENCES

Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95-99.

Bascompte, J., Luque, B., Olarrea, J. & Lacasa, L. (2007) A probabilistic model of reserve design. *Journal of Theoretical Biology*, **247**, 205-211.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134-143.

Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223-1232.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. & Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Brown, J.H. & Kodric-Brown, A. (1977) turnover rates in insular biogeography - effect of immigration and extinction. *Ecology*, **58**, 445-449.

Burnham, K.P. & Anderson, R. (2002) *Model selection and multimodel inference - A practical information - theoretic approach*, Second edn. Springer Press.

Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.J. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, **8**, 148-159.

Coleman, B.D. (1981) On random placement and species-area relations *Mathematical Biosciences*, **54**, 191-215.

Connor, E.F. & McCoy, E.D. (2001) Species-Area Relationships. *Encyclopedia of Biodiversity* (ed. by S.A. Levin), pp. 397-411. Academic Press, London, UK.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, **9**, 215-227.

Franzen, M., Schweiger, O. & Betzholtz, P.E. (2012) Species-Area Relationships Are Controlled by Species Traits. *Plos One*, **7**

Gavish, Y. (2011) Questioning israel's great biodiversity - relative to whom? a comment on Roll et al. 2009. *Israel Journal of Ecology & Evolution*, **57**, 183-192.

Gavish, Y., Ziv, Y. & Rosenzweig, M.L. (2012) Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conservation Biology*, **26**, 150-159.

Gilpin, M.E. & Diamond, J.M. (1976) Calcualtion of immigration and extinction curves from species-area-distance relation. *Proceedings of the National Academy of Sciences of the United States of America*, **73**, 4130-4134.

Gould, S.J. (1979) An allometric interpretation of species-area curves: the meaning of the coefficient. *American Naturalist*, **114**, 335-343.

Grilli, J., Azaele, S., Banavar, J.R. & Maritan, A. (2012) Spatial aggregation and the species-area relationship across scales. *Journal of Theoretical Biology*, **313**, 87-97.

Guilhaumon, F., Gimenez, O., Gaston, K.J. & Mouillot, D. (2008) Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15458-15463.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Hui, C. (2012) Scale effect and bimodality in the frequency distribution of species occupancy. *Community Ecology*, **13**, 30-35.

Hui, C. & McGeoch, M.A. (2014) Zeta Diversity as a Concept and Metric That Unifies Incidence-Based Biodiversity Patterns. *American Naturalist*, **184**, 684-694.

Jenkins, D.G. (2011) Ranked species occupancy curves reveal common patterns among diverse metacommunities. *Global Ecology and Biogeography*, **20**, 486-497.

Keil, P., Storch, D. & Jetz, W. (2015) On the decline of biodiversity due to area loss. *Nature Communications*, **6**, 1-11.

Kuussaari, M., Bommarco, R., Heikkinen, R.K., Helm, A., Krauss, J., Lindborg, R., Ockinger, E., Partel, M., Pino, J., Roda, F., Stefanescu, C., Teder, T., Zobel, M. & Steffan-Dewenter, I. (2009) Extinction debt: a challenge for biodiversity conservation. *Trends in Ecology & Evolution*, **24**, 564-571.

Lomolino, M.V. (2000) Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, **27**, 17-26.

MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press.

McGeoch, M.A. & Gaston, K.J. (2002) Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Reviews*, **77**, 311-331.

McGill, B.J. (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters*, **13**, 627-642.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853-858.

O'Dwyer, J.P. & Green, J.L. (2010) Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecology Letters*, **13**, 87-95.

Ovaskainen, O. & Hanski, I. (2003) The species-area relationship derived from species-specific incidence functions. *Ecology Letters*, **6**, 903-909.

Pan, X. (2015) Reconstruct Species-Area Theory Using Set Theory. *National Academy Science Letters*, **38**, 173-177.

Preston, F.W. (1962) The Canonical Distribution of Commonness and Rarity: Part I. . *Ecology*, **13**, 182-215.

R Development Core Team (2014) R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org,*

Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.

Rosenzweig, M.L. & Ziv, Y. (1999) The echo pattern of species diversity: pattern and processes. *Ecography*, **22**, 614-628.

Rosenzweig, M.L., Drumlevitch, F., Borgmann, K.L., Flesch, A.D., Grajeda, S.M., Johnson, G., Mackay, K., Nicholson, K.L., Patterson, V., Pri-Tal, B.M., Ramos-Lara, N. & Serrano, K.P. (2012) An ecological telescope to view future terrestrial vertebrate diversity. *Evolutionary Ecology Research*, **14**, 247-268.

Rosindell, J. & Cornell, S.J. (2007) Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology Letters*, **10**, 586-595.

Scheiner, S.M. (2003) Six types of species-area curves. *Global Ecology and Biogeography*, **12**, 441-447.

Tjørve, E. (2003) Shapes and functions of species-area curves: a review of possible models. *Journal of Biogeography*, **30**, 827-835.

Tjørve, E. (2009) Shapes and functions of species-area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36**, 1435-1445.

Tjørve, E. (2010) How to resolve the SLOSS debate: Lessons from species-diversity models. *Journal of Theoretical Biology*, **264**, 604-612.

Tjørve, E. (2012) Arrhenius and Gleason revisited: new hybrid models resolve an old controversy. *Journal of Biogeography*, **39**, 629-639.

Tjørve, E. & Turner, W.R. (2009) The importance of samples and isolates for species-area relationships. *Ecography*, **32**, 391-400.

Tokeshi, M. (1992) Dynamics of distribution in animal communities - theory and analysis. *Researches on Population Ecology*, **34**, 249-273.

Triantis, K.A., Mylonas, M. & Whittaker, R.J. (2008) Evolutionary species-area curves as revealed by single-island endemics: insights for the inter-provincial species-area relationship. *Ecography*, **31**, 401-407.

Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species-area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215-231.

Whittaker, R.J., Rigal, F., Borges, P.A.V., Cardoso, P., Terzopoulou, S., Casanoves, F., Pla, L., Guilhaumon, F., Ladle, R.J. & Triantis, K.A. (2014) Functional biogeography of oceanic islands and the scaling of functional diversity in the Azores. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 13709-13714.

Williams, M.R., Lamont, B.B. & Henstridge, J.D. (2009) Species-area functions revisited. *Journal of Biogeography*, **36**, 1994-2004.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 – AICc, wAICc and expected AICc

Appendix S2 – references and information on datasets.

Appendix S3 – Linear regressions of *wAICc.*

BIOSKETCH

Yoni Gavish – Is currently a research fellow at the University of Leeds, working on the development of biodiversity analysis tools under the EU-BON project. He is interested in species-distribution models, biodiversity-patterns, community modeling and habitat-classification models.

Yaron Ziv - a spatial and community ecologist, heading the Spatial Ecology Lab at Ben-Gurion University. He is interested in the effect of habitat heterogeneity on processes and patterns of biodiversity at different spatiotemporal scales. In particular, he strives to explore how scale-dependent and multi-scale effects define species occurrence and distribution at various levels of organization. He also involves in conservation activities, through applied research on habitat restoration and management.

EDITOR: François Guilhaumon

Table 1: The original and constrained forms of 23 known species-area relationship (SAR) functions (*c*, *z*, *f*, and *k* are function parameters). The parameters column indicate the change in the number of parameters when moving from the original to the constrained form. Functions given in bold face were used in the empirical analysis of the 154 datasets.

Name	*Original form* (S_i=)	*Constrained form* (S_i=)	*Parameters*
Lin.	$c + z \cdot A_i$	$\sum_j (j \cdot R_j) \cdot \frac{(1+z' \cdot A_i)}{\sum_i [1+z' \cdot A_i]}$	**2 → 1 (z'=z/c)**
Pow.	$c \cdot A_i^z$	$\sum_j (j \cdot R_j) \cdot \frac{A_i^z}{\sum_i [A_i^z]}$	**2 → 1**
Pow.Ros.	$f + c \cdot A_i^z$	$\sum_j (j \cdot R_j) \cdot \frac{(1+c' \cdot A_i^z)}{\sum_i [1+c' \cdot A_i^z]}$	3 → 2 (c'=c/f)
Ext.P1	$c \cdot A_i^{z \cdot A_i^{-f}}$	$\sum_j (j \cdot R_j) \cdot \frac{A_i^{z \cdot A_i^{-f}}}{\sum_i \left[A_i^{z \cdot A_i^{-f}}\right]}$	3 → 2
Ext.P2	$c \cdot A_i^{z-f/A_i}$	$\sum_j (j \cdot R_j) \cdot \frac{A_i^{z-f/A_i}}{\sum_i \left[A_i^{z-f/A_i}\right]}$	3 → 2
P1	$c \cdot A_i^z \cdot exp(-f \cdot A_i)$	$\sum_j (j \cdot R_j) \cdot \frac{A_i^z \cdot exp(-f \cdot A_i)}{\sum_i [A_i^z \cdot exp(-f \cdot A_i)]}$	3 → 2
P2	$c \cdot A_i^z \cdot exp(-f/A_i)$	$\sum_j (j \cdot R_j) \cdot \frac{A_i^z \cdot (1-exp(-f/A_i))}{\sum_i [A_i^z \cdot (1-exp(-f/A_i))]}$	**3 → 2**
Exp.	$c + z \cdot \log(A_i)$	$\sum_j (j \cdot R_j) \cdot \frac{(1+z' \cdot \log(A_i))}{\sum_i [1+z' \cdot \log(A_i)]}$	**2 → 1 (z'=z/c)**
Kob.	$c \cdot log(1 + A_i/z)$	$\sum_j (j \cdot R_j) \cdot \frac{log(1+A_i/z)}{\sum_i [log(1+A_i/z)]}$	**2 → 1**
Mon.	$c/(1 + z/A_i)$	$\sum_j (j \cdot R_j) \cdot \frac{(1/(1+z/A_i))}{\sum_i [1/(1+z/A_i)]}$	**2 → 1**
MMF	$c/(1 + f \cdot A_i^{-z})$	$\sum_j (j \cdot R_j) \cdot \frac{(1/(1+f \cdot A_i^{-z}))}{\sum_i [1/(1+f \cdot A_i^{-z})]}$	3 → 2
Arc.Log.	$c/(f + A_i^{-z})$	$\sum_j (j \cdot R_j) \cdot \frac{(1/(f+A_i^{-z}))}{\sum_i [1/(f+A_i^{-z})]}$	**3 → 2**
Neg.Exp.	$c \cdot (1 - exp(-z \cdot A_i))$	$\sum_j (j \cdot R_j) \cdot \frac{(1-exp(-z \cdot A_i))}{\sum_i [1-exp(-z \cdot A_i)]}$	**2 → 1**
Chp.Ric.	$c \cdot (1 - exp(-z \cdot A_i))^f$	$\sum_j (j \cdot R_j) \cdot \frac{(1-exp(-z \cdot A_i))^f}{\sum_i [1-exp(-z \cdot A_i)]^f}$	3 → 2

Wei.3	$\boldsymbol{c \cdot \left(1 - exp\left(-z \cdot A_i^f\right)\right)}$	$\boldsymbol{\sum_j (j \cdot R_j) \cdot \frac{\left(1-exp\left(-z \cdot A_i^f\right)\right)}{\sum_i \left[1-exp\left(-z \cdot A_i^f\right)\right]}}$	**3 → 2**
Wei.4	$c \cdot \left(1 - exp\left(-z \cdot A_i^f\right)\right)^k$	$\sum_j (j \cdot R_j) \cdot \frac{\left(1-exp\left(-z \cdot A_i^f\right)\right)^k}{\sum_i \left[1-exp\left(-z \cdot A_i^f\right)\right]^k}$	4 → 3
Asy.	$f - c \cdot z^{-A_i}$	$\sum_j (j \cdot R_j) \cdot \frac{\left(1+c' \cdot z^{-A_i}\right)}{\sum_i \left[1+c' \cdot z^{-A_i}\right]}$	3 → 2 (c'=c/f)
Rat.	$\boldsymbol{(c + z \cdot A_i)/(1 + f \cdot A_i)}$	$\boldsymbol{\sum_j (j \cdot R_j) \cdot \frac{\left(1+z' \cdot A_i\right)/(1+f \cdot A_i)}{\sum_i \left[\left(1+z' \cdot A_i\right)/(1+f \cdot A_i)\right]}}$	**3 → 2 (z'=z/c)**
Gom.	$\boldsymbol{c \cdot exp\left(-exp\left(-z \cdot (A_i - f)\right)\right)}$	$\boldsymbol{\sum_j (j \cdot R_j) \cdot \frac{exp \cdot \left(-exp\left(-z \cdot (A_i - f)\right)\right)}{\sum_i \left[exp \cdot \left(-exp\left(-z \cdot (A_i - f)\right)\right)\right]}}$	**3 → 2**
Beta.P	$c \cdot (1 - (1 + (A_i/z)^f)^{-k})$	$\sum_j (j \cdot R_j) \cdot \frac{\left(1-\left(1+(A_i/z)^f\right)^{-k}\right)}{\sum_i \left[e\left(1-\left(1+(A_i/z)^f\right)^{-k}\right)\right]}$	4 → 3
Com.Log.	$\boldsymbol{c/\left(1 + exp(-z \cdot A_i + f)\right)}$	$\boldsymbol{\sum_j (j \cdot R_j) \cdot \frac{\left(1/(1+exp(-z \cdot A_i + f))\right)}{\sum_i \left[1/(1+exp(-z \cdot A_i + f))\right]}}$	**3 → 2**
EVF.	$c \cdot \left(1 - exp\left(-exp(z \cdot A_i + f)\right)\right)$	$\sum_j (j \cdot R_j) \cdot \frac{\left(1-exp\left(-exp(z \cdot A_i + f)\right)\right)}{\sum_i \left[1-exp\left(-exp(z \cdot A_i + f)\right)\right]}$	3 → 2
Lom.	$c/\left(1 + \left(z^{log(f/A_i)}\right)\right)$	$\sum_j (j \cdot R_j) \cdot \frac{\left(1/\left(1+\left(z^{log(f/A_i)}\right)\right)\right)}{\sum_i \left[1/\left(1+\left(z^{log(f/A_i)}\right)\right)\right]}$	3 → 2

Pow. – Power; **Pow.Ros.** – Power Rosenzweig; **Ext.P1**– Extended Power 1; **Ext.P2**– Extended Power 2; **P1** – Persistence Function 1; **P2** – Persistence Function 2; **Exp.** – Exponential; **Kob.** – Kobayashi Logarithmic; **Mon.** – Monod; **MMF.** – Morgan-Mercer-Flodin; **Arc.Log.** – Archibald Logistic; **Neg.Exp.** – Negative Exponential; **Chp.Ric.** – Chapman-Richards; **Wei.3** – Weibull-3; **Wei.4** – Weibull-4; **Asy.** – Asymptotic; **Rat.** – Rational. **Gom.** – Gompertz; **Beta.P.** – Beta-P; **Com.Log. -** Common Logistic; **EVF.** – Extreme-Value Function; **Lom.** – Lomolino function.

Table 2: Corrected Akaike Information Criteria (AICc) values and weights of the original power-model (equation 1), the constrained power-model (equation 2) and the additive-constrained power-model (equation 9) for the nine datasets presented in figure 2. Values in parentheses are the ranking of each model according to the AICc weights.

	AICc			AICc weights		
Data set	Equation 1	Equation 2	Equation 9	Equation 1	Equation 2	Equation 9
DS49	135.89	133.01	135.90	0.161 (2)	0.679 (1)	0.160 (3)
DS104	50.04	45.79	53.27	0.104 (2)	0.875 (1)	0.021 (3)
DS125	107.16	103.49	106.00	0.111 (3)	0.692 (1)	0.197 (2)
DS16	208.25	205.47	207.81	0.160 (3)	0.642 (1)	0.199 (2)
DS38	108.49	101.49	101.26	0.014 (3)	0.465 (2)	0.521 (1)
DS64	114.13	111.39	113.72	0.162 (3)	0.639 (1)	0.199 (2)
DS115	232.96	230.64	236.1	0.227 (2)	0.725 (1)	0.047 (3)
DS136	58.69	53.91	51.19	0.018 (3)	0.201 (2)	0.781 (1)
DS117	1135.18	1133.04	1133.90	0.172 (3)	0.501 (1)	0.326 (2)

Figure 1. General framework for species-occupancy distributions (SODs) and species-area relationships (SARs). The marginal sums of presence/absence tables (A) yields the number of species per isolate which can be used to plot the general SAR (D). The second marginal sums yields the number of isolates per species (i.e., the species occupancy level), which can be used to produce the SOD (C). However, the presence/absence table can be rearranged by grouping species from the same occupancy level (B). From the resulting N×N square matrix the occupancy-specific SARs can be produced (E) as well as a weighted version of the SOD (wSOD, C). The additive-constrained SAR develop here is based on this square matrix. The notations used here and in text are given in (F).

Figure 2. The corrected Akaike Information Criteria weight (wAICc) of the constrained species-area relationships (SAR) form for 154 datasets and 12 SAR functions**.** For 1789 of 1811 valid combinations of 154 datasets and 12 functions, the (wAICc of the constrained form (red, shown here against the number of isolates), was higher than that of the original form. The observed wAICc of the constrained form approaches the expected weight, if the two forms have identical log-likelihood (and as such similar goodness-of-fit) and only differ in the number of parameters (solid black line). As the number of isolates increases, the wAICc approach the expected values under identical log likelihood and infinite number of isolates (horizontal dashed line). See appendix 1 for details.

Figure 3. A few examples of z-occupancy curves. The predicted *z* values in each occupancy level as predicted when fitting equation 9 (red line), compared to the *z* value obtained when fitting each occupancy level separately with a constrained power-model (equation 3, black diamonds). The dashed black line was obtained by fitting equation 8 to the fitted *z* values. Each panel is for one dataset (DS), numbered according to Table S2 (Appendix S2).

Figure 4. The relation between z and *Sørensen* multiple-sites similarity index. The power-model's *z* values decrease with increase in multiple-sites similarity index ($y = 0.53 – 0.61 \times x$, $F_{113, 1} = 14.491$, $p < 0.001$). The size of the points is relative to the square root of the number of isolates in the dataset.

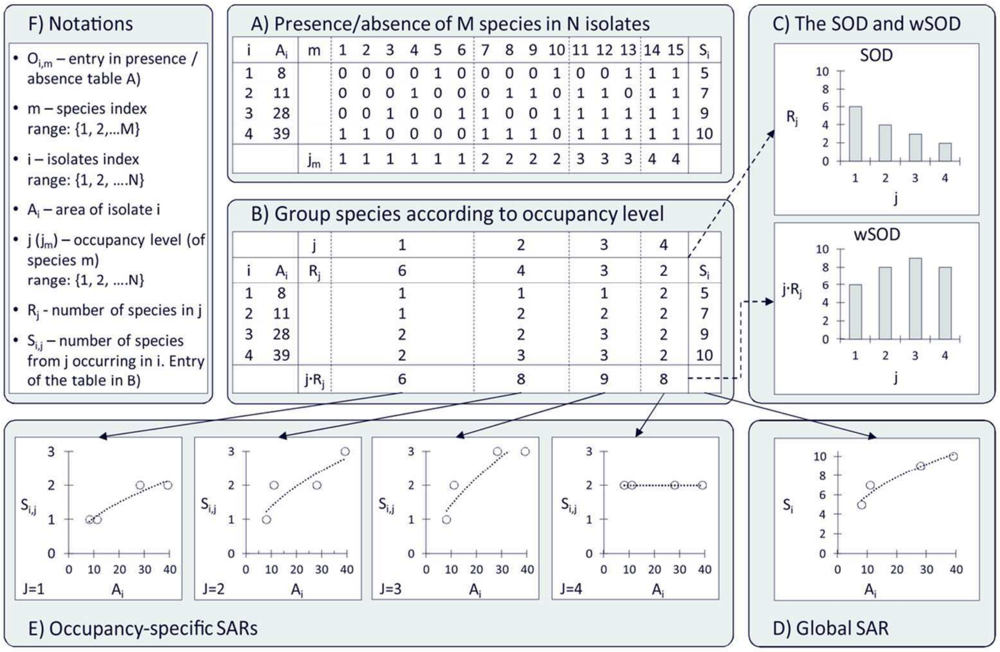
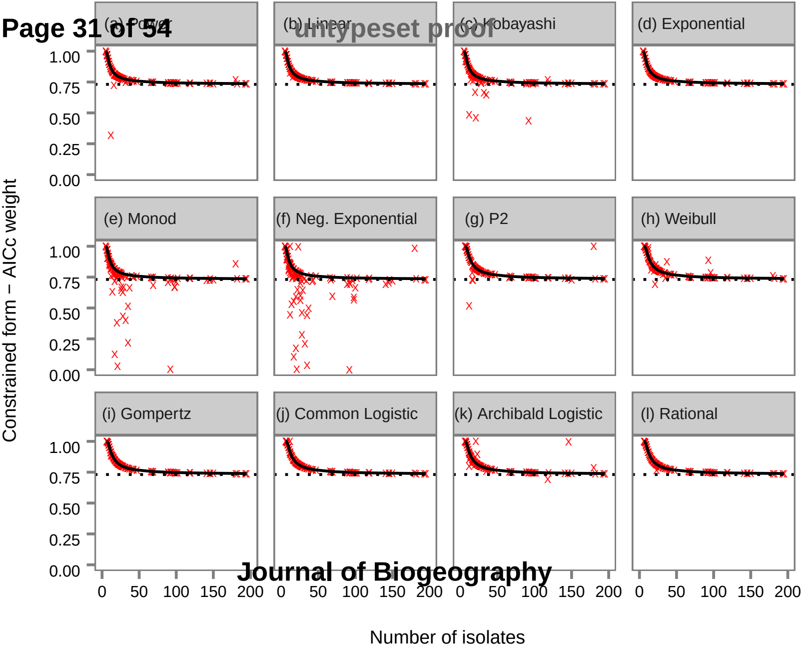
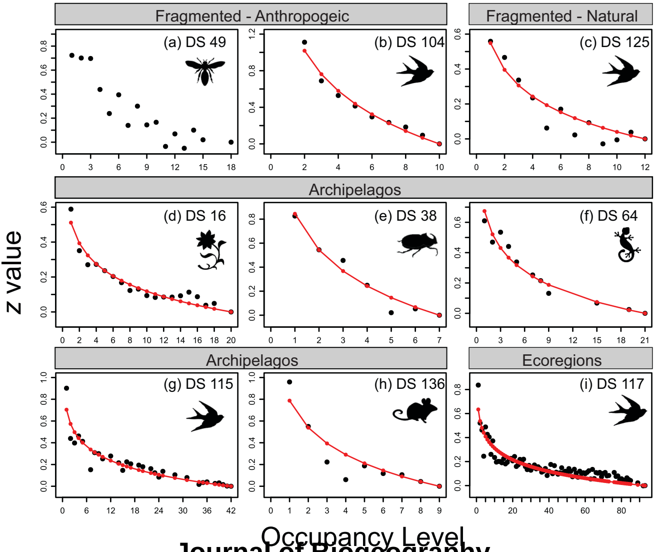


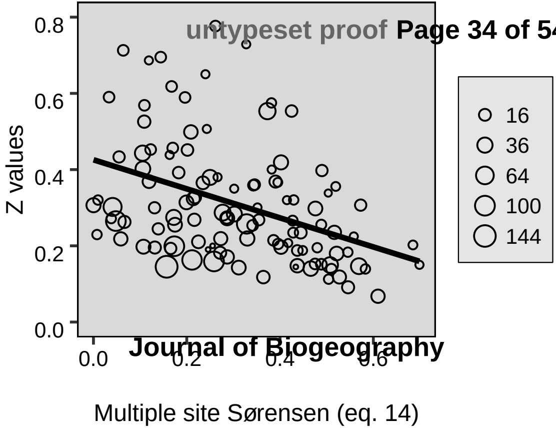
Figure 1. General framework for species-occupancy distributions (SODs) and species-area relationships (SARs). The marginal sums of presence/absence tables (A) yields the number of species per isolate which can be used to plot the general SAR (D). The second marginal sums yields the number of isolates per species (i.e., the species occupancy level), which can be used to produce the SOD (C). However, the presence/absence table can be rearranged by grouping species from the same occupancy level (B). From the resulting N×N square matrix the occupancy-specific SARs can be produced (E) as well as a weighted version of the SOD (wSOD, C). The additive-constrained SAR develop here is based on this square matrix. The notations used here and in text are given in (F).

109x71mm (300 x 300 DPI)


(a) Power
(b) Linear
(c) Kobayashi
(d) Exponential
(e) Monod
(f) Neg. Exponential
(g) P2
(h) Weibull
(i) Gompertz
(j) Common Logistic
(k) Archibald Logistic
(l) Rational
1.00
0.75
0.50
0.25
0.00
0
50
100
150
200
Constrained form − AICc weight
Number of isolates


Fragmented - Anthropogeic
Fragmented - Natural
(a) DS 49
(b) DS 104
(c) DS 125
Archipelagos
(d) DS 16
(e) DS 38
(f) DS 64
Archipelagos
Ecoregions
(g) DS 115
(h) DS 136
(i) DS 117
z value
Occupancy Level

(a) DS 49



0.8
0.6
0.4
0.2
0.0
Z values
0.0
0.2
0.4
0.6
Multiple site Sørensen (eq. 14)
16
36
64
100
144

Journal of Biogeography

SUPPORTING INFORMATION

Species-occupancy distribution removes excessive parameter from species-area relationship

Yoni Gavish and Yaron Ziv

Supplementary Information

Appendix S1 – explaining the AICc, wAICc and expected AICc

Appendix S2 – Reference list of all datasets used in the analysis (Table S1) and basic information on each dataset (table S2).

Appendix S3 – Linear regression between the expected and observed *wAICc* of the constrained form (table S3) with additional focus on the Monod and negative exponential functions (figure S1).

Appendix S1

Explaining the AICc, wAICc and expected AICc

Empirical datasets - data analysis. Non-linear least square regressions (using the Levenberg-Marquardt algorithm) were used to fit each dataset with 24 models (two forms of the 12 functions in Table 1). We used various parameter starting values to avoid local minima. All analyses were carried out with the function nlsLM (Minpack.lm Package) in R (R Development Core Team, 2011). After fitting the 24 models, we calculated Log-likelihood as:

$$LL=-(n/2)\cdot\ln(2\pi)-(n/2)\cdot\ln(RSS/n)-(n/2) \qquad \text{(SI1.1)}$$

where *RSS* is the residuals sum of square and *n* is the number of isolates. Next, AICc values were calculated as:

$$-2\cdot LL+2\cdot F+2\cdot F\cdot(F+1)/(n-F-1) \qquad \text{(SI1.2)}$$

with *F* being the number of parameters of the model plus one for the residuals variance (Burnham & Anderson, 2002). As such, original and constrained forms of SAR function with two function parameters add F=3 and F=2, respectively. Original and constrained forms of SAR function with three function parameters add F=4 and F=3, respectively. We then calculated delta AICc and AICc weights. To avoid mixing the frequentist approach with the model-selection approach we employed in this study, we have not checked for normality with commonly used methods (e.g., Kolmogorov-Smirnov). Instead, we repeated the entire analysis using a Poisson error distribution, with: $LL=\sum_{i=1}^{n} ln\left[exp(-\mu_i\cdot\mu_i^{y_i})/y_i\,!\right]$

(where y_i and μ_i are the observed and expected number of species for isolate *i*). Using the Poisson error had no qualitative effect on the results shown in the paper.

Empirical datasets - expected AICc weights. If constraining the SAR has no effect on the model's goodness-of-fit, the log-likelihood of the two forms should be identical. Under identical log-likelihoods, the constrained form (with one parameter less) will have a lower AICc value than the original form. Therefore, the delta AICc of the constrained form will be 0, and that of the original form will be:

$$\Delta AICc_{reg}=AICc_{reg}-AICc_{con}=\left[2\cdot F+\frac{2\cdot F\cdot(F+1)}{(n-F-1)}\right]-\left[2\cdot(F-1)+\frac{2\cdot(F-1)\cdot F}{(n-F)}\right] \qquad \text{(SI1.3)}$$

The expected AICc weight of the constrained form (solid black line in Fig. 2) can then be calculated as:

$$wAICc_{con}=\frac{\exp(-0.5\cdot 0)}{(\exp(-0.5\cdot 0)+\exp(-0.5*\Delta AICc_{reg}))} \qquad \text{(SI1.4)}$$

depending only on the number of isolates (n) and the number of parameters of the original form (F). When the number of isolates approaches infinity, the second term within the brackets of equation SI1.3 can be omitted,

$\Delta AICc_{reg}$ = 2, and $wAICc_{con}$ = 0.731 (dashed horizontal line in Fig. 1 of the main text). Finally, for each of the 12 SAR functions we explored the relation between the observed and expected $wAICc_{con}$ using linear regressions. If no information is lost, we expected the 154 datasets to fall on the unity line (Supplementary Table S3).

Burnham K.P. & Anderson R. (2002). Model selection and multimodel inference - A practical information - theoretic approach. Second edn. Springer Press.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Appendix S1

Reference list of all datasets used in the analysis (Table S1) and basic information on each dataset (table S2).

The 154 dataset explored in this manuscript were collected using several methodologies. Our criteria for inclusion in the meta-analysis was that the reference:

- Reported the entire presence/absence data and not only the total number of species per site
- The area of the isolates was reported or can be extracted from online sources such as the island directory (http://islands.unep.ch/isldir.htm).
- Sampling effort increased with area.
- The pdf was available online or was received from the authors upon request.

We looked for dataset using several sources:

1. Manuscript known by the authors from there general reading in the fields.
2. Data collected by the authors
3. Dataset lists of other meta-analyses, mainly:
    a. Triantis et al., 2012, The island species-area relationship: biology and statistics, Journal of Biogeography, 39 (2): 215-231
    b. Drakare et al., 2006, The imprint of the geographical, evolutionary and ecological context on species–area relationships, Ecology Letters 9 (2): 215-227
    c. Boecklen, W. J., 1997, Nestedness, biogeographic theory, and the design of nature reserves, Oecologia, 112 (1): 123-142
4. General google scholar/web of knowledge search using different crossing of the terms (and closely related terms):

System term	Crossed with	Data type
fragmentation	×	species list
fragmented landscapes		presence/absence
patchy landscape		occurrence
islands		abundance
archipelagos		Species atlas

Table S1 contains the list of references for all the datasets (with some references providing more than 1 dataset), while table S2 provides additional information (as well as some analytical results for each dataset)

Table S1:
Reference details for the 154 datasets used in the analysis. Reference number in supplementary table 2 refers to the numbers here.

1 Aldasoro, J. J., Cabezas, F. & Aedo, C. Diversity and distribution of ferns in sub-Saharan Africa, Madagascar and some islands of the South Atlantic. *Journal of Biogeography* **31**, 1579-1604, doi:10.1111/j.1365-2699.2004.01106.x (2004).
2 Andrade, R. D. & Marini, M. A. Bird species richness in natural forest patches in southeast Brazil. *Lundiana* **3**, 141-149 (2002).
3 Anjos, L. D. & Bocon, R. Bird communities in natural forest patches in southern Brazil. *Wilson Bull.* **111**, 397-414 (1999).
4 Arechavaleta, M., Zurita, N., Marrero, M. C. & Martin, J. L. Lista preliminary de especies silvestres de Cabo Verde (hongos plantas y animals terrestres) - Consejeria de Medio Ambiente y Ordenacion Territorial, Gobierno de Canarias. . (2005).

5 Arechavaleta, M., Zurita, N. & Garcia, A. Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. *Gobierno de Canarias*, 1-577 (2009).
6 Azeria, E. T. Terrestrial bird community patterns on the coralline islands of the Dahlak Archipelago, Red Sea, Eritrea. *Glob. Ecol. Biogeogr.* **13**, 177-187, doi:10.1111/j.1466-882X.2004.00079.x (2004).
7 Baker, R. J. & Genoways, H. H. Zoogeography of Antillean bats. *Zoogeography in the Caribbean* **Special publication 13**, 53-97 (1978).
8 Baldi, A. & Kisbenedek, T. Orthopterans in small steppe patches: an investigation for the best-fit model of the species-area curve and evidences for their non-random distribution in the patches. *Acta Oecol.-Int. J. Ecol.* **20**, 125-132, doi:10.1016/s1146-609x(99)80025-3 (1999)
9 Baz, A. & GarciaBoyero, A. The SLOSS dilemma: A butterfly case study. *Biodivers. Conserv.* **5**, 493-502 (1996).
10 Bolger, D. T. *et al.* Response of rodents to habitat fragmentation in coastal southern California. *Ecol. Appl.* **7**, 552-563, doi:10.2307/2269520 (1997).
11 Brown, J. H. Mammals on mountaintops - nonequilibrium insular biogeography. *Am. Nat.* **105**, 467-&, doi:10.1086/282738 (1971).
12 Case, T. J., Cody, M. L. & Ezcurra, E. (editors) *A new island biogeography of the sea of Cortes*. (Oxford University Press, 2002).
13 Castelletta, M., Thiollay, J. M. & Sodhi, N. S. The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biol. Conserv.* **121**, 135-155, doi:10.1016/j.biocon.2004.03.033 (2005).
14 Chinnaraj, S. Higher marine fungi from mangroves of Andaman and Nicobar Islands. *Sydowia* **45**, 109-115 (1993).
15 Conroy, C. J., Demboski, J. R. & Cook, J. A. Mammalian biogeography of the Alexander Archipelago of Alaska: a north temperate nested fauna. *Journal of Biogeography* **26**, 343-352, doi:10.1046/j.1365-2699.1999.00266.x (1999).
16 Corti, C., Masseti, M., Delfino, M. & Perez-Mellado, V. Man and herpatofauna of the mediterranean islands. *Revista Española de Herpetología* **13**, 83-100 (1999).
17 Cramer, K. L. New mammal record for Fremont Island with an updated checklist of mammals on islands in the Great-Salt-Lake. *Gt. Basin Nat.* **54**, 287-289 (1994).
18 Crowell, K. L. A comparison of relict versus equilibrium-models for insular mammals of the gulf of Maine. *Biol. J. Linnean Soc.* **28**, 37-64, doi:10.1111/j.1095-8312.1986.tb01748.x (1986).
19 Dapporto, L. & Cini, A. Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *Eur. J. Entomol.* **104**, 497-503 (2007).
20 Dapporto, L. & Dennis, R. L. H. Species richness, rarity and endemicity on Italian offshore islands: complementary signals from island-focused and species-focused analyses. *Journal of Biogeography* **35**, 664-674, doi:10.1111/j.1365-2699.2007.01812.x (2008).
21 de Castro, E. B. V. & Fernandez, F. A. S. Determinants of differential extinction vulnerabilities of small mammals in Atlantic forest fragments in Brazil. *Biol. Conserv.* **119**, 73-80, doi:10.1016/jbiocon.2003.10.023. (2004).
22 Dennis, R. L. H., Shreeve, T. G., Olivier, A. & Coutsis, J. G. Contemporary geography dominates butterfly diversity gradients within the Aegean archipelago (Lepidoptera : Papilionoidea, Hesperioidea). *Journal of Biogeography* **27**, 1365-1383, doi:10.1046/j.1365-2699.2000.00514.x (2000).
23 Donazar, J. A., Gangoso, L., Forero, M. G. & Juste, J. Presence, richness and extinction of birds of prey in the Mediterranean and Macaronesian islands. *Journal of Biogeography* **32**, 1701-1713, doi:10.1111/j.1365-2699.2005.01294.x (2005).
24 Dueser, R. D. & Brown, W. C. Ecological correlates of insular rodent diversity. *Ecology* **61**, 50-56, doi:10.2307/1937154 (1980).
25 Enderson, E. F., Quijada-Mascareñas, A., Turner, D. S., Rosen, P. C. & Bezy, R. L. The herpetofauna of Sonora, Mexico, with comparisons to adjoining states. *Check List* **5**, 632-672 (2009).

26 Fattorini, S. Biogeography of the tenebrionid beetles (Coleoptera, Tenebrionidae) on the Aegean Islands (Greece). *Journal of Biogeography* **29**, 49-67 (2002).
27 Fernandez-Juricic, E. Bird community composition patterns in urban parks of Madrid: The role of age, size and isolation. *Ecol. Res.* **15**, 373-383, doi:10.1046/j.1440-1703.2000.00358.x (2000).
28 Fuller, R. J. & Glue, D. E. Sewage works as bird habitats in Britain. *Biol. Conserv.* **17**, 165-181, doi:10.1016/0006-3207(80)90054-3 (1980).
29 Galli, A. E., Leck, C. F. & Forman, R. T. T. Avian distribution patterns in forest islands of different sizes in central New-Jersey. *Auk* **93**, 356-364 (1976).
30 Gavish, Y., Ziv, Y. & Rosenzweig, M. L. Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conserv. Biol.* **26**, 150-159 (2012).
31 Gavish, Y. & Ziv, Y. Unpublished data. Avaialble from the authors upon request (gavishyoni@gmail.com)
32 Giladi, I. & Ziv, Y. Unpublished data. Availble upon request from: itushgi@bgu.ac.il
33 Gotmark, F., Ahlund, M. & Eriksson, M. O. G. Are indexes reliable for assessing conservation value of natural areas - an avian case-study. *Biol. Conserv.* **38**, 55-73, doi:10.1016/0006-3207(86)90019-4 (1986).
34 Hausdorf, B. & Hennig, C. The influence of recent geography, palaeogeography and climate on the composition of the fauna of the central Aegean Islands. *Biol. J. Linnean Soc.* **84**, 785-795, doi:10.1111/j.1095-8312.2005.00467.x (2005).
35 Heaney, L. R. Biogeography of mammals in SE Asia - estimates of rates of colonization, extinction and speciation. *Biol. J. Linnean Soc.* **28**, 127-165, doi:10.1111/j.1095-8312.1986.tb01752.x (1986).
36 Howe, R. W. & Jones, G. Avian utilization of small woodlots in Dane county, Wisconsin. *Passenger Pigeon*, 313-319 (1977).
37 Kitchener, D. J., Chapman, A., Dell, J. & Muir, B. G. Lizard assemblage and reserve size and structure in the western Australian wheatbelt - some implications for conservation. *Biol. Conserv.* **17**, 25-62, doi:10.1016/0006-3207(80)90024-5 (1980).
38 Kitchener, D. J., Chapman, A., Muir, B. G. & Palmer, M. The conservation value for mammals of reserves in the western Australian wheatbelt. *Biol. Conserv.* **18**, 179-207, doi:10.1016/0006-3207(80)90033-6 (1980).
39 Krystufek, B. & Kletecki, E. Biogeography of small terrestrial vertebrates on the Adriatic landbridge islands. *Folia Zool.* **56**, 225-234 (2007).
40 Lomolino, M. V. Mammalian community structure on islands - the importance of immigration, extinction and interactive effects. *Biol. J. Linnean Soc.* **28**, 1-21, doi:10.1111/j.1095-8312.1986.tb01746.x (1986).
41 Lomolino, M. V., Brown, J. H. & Davis, R. Island biogeography of montane forest mammals in the American southwest. *Ecology* **70**, 180-194, doi:10.2307/1938425 (1989).
42 Lomolino, M. V. & Perault, D. R. Island biogeography and landscape ecology of mammals inhabiting fragmented, temperate rain forests. *Glob. Ecol. Biogeogr.* **10**, 113-132, doi:10.1046/j.1466-822x.2001.00221.x (2001).
43 Marshall, A. R. *et al.* The Species Area Relationship and Confounding Variables in a Threatened Monkey Community. *Am. J. Primatol.* **72**, 325-336, doi:10.1002/ajp.20787 (2010).
44 Miyashita, T., Shinkai, A. & Chida, T. The effects of forest fragmentation on web spider communities in urban areas. *Biol. Conserv.* **86**, 357-364 (1998).
45 Oliveros, C. *et al. An avifaunal survey of the Babuyan islands, northern Philippines, with notes on mammals, reptiles and amphibians. 29 March – 6 June 2004. Final Report*. (Unpublished report. Manila, Philippines., 2004).
46 Panitsa, M. & Tzanoudakis, D. Floristic diversity on small islands and islets: Leros islets' group (East Aegean area, Greece). *Phytologia Balcanica* **16**, 271-284 (2010).
47 Powell, R. & Henderson, R. W. Island lists of west-Indian amphibians and reptiles. *Bulletin of the Florida Museum of Natural History* **51**, 85–166 (2012).

48 Presley, S. J. & Willig, M. R. Bat metacommunity structure on Caribbean islands and the role of endemics. *Glob. Ecol. Biogeogr.* **19**, 185-199, doi:10.1111/j.1466-8238.2009.00505.x (2010).
49 Ribas, C. R. *et al.* How large is large enough for insects? Forest fragmentation effects at three spatial scales. *Acta Oecol.-Int. J. Ecol.* **27**, 31-41, doi:10.1016/j.actao.2004.08.008 (2005).
50 mberloff, D. Experimental zoogeography of islands - effects of island size. *Ecology* **57**, 629-648, doi:10.2307/1936179 (1976).
51 Smith, G. T., Arnold, G. W., Sarre, S., AbenspergTraun, M. & Steven, D. E. The effects of habitat fragmentation and livestock-grazing on animal communities in remnants of gimlet Eucalyptus salubris woodland in the Western Australian wheatbelt .2. Lizards. *J. Appl. Ecol.* **33**, 1302-1310, doi:10.2307/2404771 (1996).
52 Soule, M. E. *et al.* Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat Islands. *Conserv. Biol.* **2**, 75-92 (1988).
53 Urban, M. C. Disturbance heterogeneity determines freshwater metacommunity structure. *Ecology* **85**, 2971-2978 (2004).
54 Wang, Y. P., Bao, Y. X., Yu, M. J., Xu, G. F. & Ding, P. Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions* **16**, 862-873, doi:10.1111/j.1472-4642.2010.00682.x (2010).
55 Watson, D. M. Long-term consequences of habitat fragmentation - highland birds in Oaxaca, Mexico. *Biol. Conserv.* **111**, 283-303, doi:10.1016/s0006-3207(02)00271-9 (2003).
56 World Wildlife Fund. *WildFinder: Online database of species distributions, ver. Jan-06*, <www.worldwildlife.org/WildFinder> (2006). Access date: October 2012.

Table S2: Basic information on each dataset used in the analysis. The best model is the model that received the highest AICc weight from the 24 models (the constrained and regular forms of 12 functions). Abbreviations: *n* – number of isolates; Models abbreviations follow table 1 and are given below; Ori – original form; Con – constrained form; Ref' – the reference number according to supplementary table S1. A version of the table in .xls format is available from the authors upon request (YG at gavishyoni@gmail.com).

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
1	Fungi	Archipelago	Andaman and Nicobar Islands	6	127-1536 (km^2)	63 (24 , 38.1%)	180	2.86	Lin. (Con)	14
2	Fungi	Archipelago	Canary islands	7	290.5-2007 (km^2)	1825 (963 , 52.8%)	3599	1.97	Lin. (Con)	5
3	Fungi	Archipelago	Cape-Verde	12	1.4-991 (km^2)	58 (36 , 62.1%)	93	1.60	Pow. (Con)	4
4	Fungi (Lichens)	Archipelago	Canary islands	7	290.5-2007 (km^2)	1438 (618 , 43%)	3364	2.34	Lin. (Con)	5
5	Fungi (Lichens)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	244 (77 , 31.6%)	773	3.17	Pow. (Con)	4
6	Plants (Bryophyta)	Archipelago	Canary islands	7	290.5-2007 (km^2)	351 (88 , 25.1%)	1263	3.60	Lin. (Con)	5
7	Plants (Bryophyta)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	139 (45 , 32.4%)	323	2.32	Neg.Exp. (Con)	4
8	Plants (Ferns)	Other	Sub-Saharan Africa	27	17-311040 (km^2)	687 (180 , 26.2%)	3115	4.53	Pow. (Con)	1
9	Plants (Marchantio-phyta)	Archipelago	Canary islands	7	290.5-2007 (km^2)	138 (27 , 19.6%)	502	3.64	Lin. (Con)	5
10	Plants	Anthropogenic fragmented landscape	Lachish, Israel	40	0.06-7.93 (ha)	408 (77 , 18.9%)	6012	14.74	Exp. (Con)	32
11	Plants	Archipelago	Leros islets' group, east Aegean, Greece	17	0.6-124 (ha)	290 (99 , 34.1%)	958	3.30	Neg.Exp. (Con)	46
12	Plants (Pteridophyta)	Archipelago	Canary islands	7	290.5-2007 (km^2)	49 (5 , 10.2%)	218	4.45	Mon. (Con)	5
13	Plants (Pteridophyta)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	35 (11 , 31.4%)	118	3.37	P2 (Con)	4

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
14	Plants (Spermatophyta)	Archipelago	Canary islands	7	290.5-2007 (km^2)	1314 (416 , 31.7%)	4461	3.39	Lin. (Con)	5
15	Plants (Spermatophyta)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	757 (195 , 25.8%)	3196	4.22	Pow. (Con)	4
16	Plants (Vascular)	Archipelago	Sea of Cortes	20	0.6-1223 (km^2)	707 (281 , 39.8%)	2734	3.87	Kob. (Con)	12
17	Plants (Vascular)	Archipelago	Sea of Cortes, Bahia de Los Angeles	14	0.02-9.13 (km^2)	99 (23 , 23.2%)	396	4.00	Kob. (Con)	12
18	Plants (Vascular)	Archipelago	Sea of Cortes, small gulf islands	10	0.02-2.26 (km^2)	153 (96 , 62.8%)	271	1.77	Lin. (Con)	12
19	Invertebrates	Archipelago	mangrove islands, Florida Bay, USA (1969)	9	264-1263 (m^2)	205 (75 , 36.6%)	732	3.57	P2 (Con)	50
20	Invertebrates	Archipelago	mangrove islands, Florida Bay, USA (1970)	9	104-779 (m^2)	179 (67 , 37.4%)	652	3.64	Kob. (Con)	50
21	Invertebrates (Freshwater invertebrates)	Inland water bodies	Connecticut, USA	14	6.03-8318 (m^2)	89 (43 , 48.3%)	232	2.61	Pow. (Con)	53
22	Invertebrates (Mollusca)	Archipelago	Canary islands	7	290.5-2007 (km^2)	259 (219 , 84.6%)	354	1.37	Lin. (Con)	5
23	Invertebrates (Mollusca)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	50 (13 , 26%)	192	3.84	Pow. (Con)	4
24	Invertebrates (Mollusca-Land snails)	Archipelago	Aegean Islands	34	3.4-842 (km^2)	152 (47 , 30.9%)	996	6.55	Pow. (Con)	34
25	Invertebrates (Acarina)	Archipelago	Canary islands	7	290.5-2007 (km^2)	391 (236 , 60.4%)	620	1.59	Neg.Exp. (Con)	5
26	Invertebrates (Araneae)	Anthropogenic fragmented landscape	Dvir, Israel	12	0.11-3.90 (ha)	114 (43 , 37.7%)	389	3.41	Exp. (Con)	30
27	Invertebrates (Araneae)	Anthropogenic fragmented landscape	Galon, Israel	8	0.16-4.24 (ha)	99 (39 , 39.4%)	308	3.11	Mon. (Con)	31
28	Invertebrates (Araneae)	Anthropogenic fragmented landscape	Lachish, Israel	12	0.06-2.81 (ha)	115 (35 , 30.4%)	447	3.89	Pow. (Con)	30
29	Invertebrates (Araneae)	Anthropogenic fragmented	Tokyo, Japan	7	0.2-27	34 (12 ,	103	3.03	Mon. (Con)	44

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
		landscape			(ha)	35.3%)				
30	Invertebrates (Araneae)	Anthropogenic fragmented landscape	Yokohama, Japan	9	0.4-15.1 (ha)	52 (12 , 23.1%)	223	4.29	Pow. (Con)	44
31	Invertebrates (Araneae)	Archipelago	Canary islands	7	290.5-2007 (km^2)	455 (258 , 56.7%)	962	2.11	Lin. (Con)	5
32	Invertebrates (Araneae)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	114 (44 , 38.6%)	233	2.04	Kob. (Con)	4
33	Invertebrates (Crustacea)	Archipelago	Canary islands	7	290.5-2007 (km^2)	99 (63 , 63.6%)	170	1.72	Arc.Log. (Con)	5
34	Invertebrates (Crustacea-Land isopods)	Archipelago	Aegean Islands	23	3.4-476 (km^2)	67 (14 , 20.9%)	569	8.49	Pow. (Con)	34
35	Invertebrates (Chilopoda)	Archipelago	Canary islands	7	290.5-2007 (km^2)	30 (11 , 36.7%)	90	3.00	Lin. (Con)	5
36	Invertebrates (Diplopoda)	Archipelago	Canary islands	7	290.5-2007 (km^2)	67 (56 , 83.6%)	91	1.36	Rat. (Con)	5
37	Invertebrates (Nematoda)	Archipelago	Canary islands	7	290.5-2007 (km^2)	98 (61 , 62.2%)	195	1.99	Rat. (Con)	5
38	Invertebrates (Coleoptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	1926 (950 , 49.3%)	4638	2.41	Lin. (Con)	5
39	Invertebrates (Coleoptera)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	474 (171 , 36.1%)	1408	2.97	Lin. (Con)	4
40	Invertebrates (Coleoptera, Tenebrionidae)	Archipelago	Aegean Islands	32	3.8-8260 (km^2)	165 (92 , 55.8%)	514	3.12	Lin. (Con)	26
41	Invertebrates (Coleoptera, Tenebrionidae)	Archipelago	Sea of Cortes	18	0.004-9.13 (km^2)	31 (12 , 38.7%)	120	3.87	Pow. (Con)	12
42	Invertebrates (Collembola)	Archipelago	Canary islands	7	290.5-2007 (km^2)	113 (53 , 46.9%)	249	2.20	Rat. (Con)	5
43	Invertebrates (Diptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	989 (350 , 35.4%)	2439	2.47	Lin. (Con)	5
44	Invertebrates (Diptera)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	220 (97 , 44.1%)	546	2.48	Pow. (Con)	4

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
45	Invertebrates (Hemiptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	725 (254 , 35%)	2052	2.83	Lin. (Con)	5
46	Invertebrates (Hemiptera)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	308 (105 , 34.1%)	937	3.04	Pow. (Con)	4
47	Invertebrates (Hymenoptera)	Archipelago	Canary islands	7	290.5-2007 (km^2))	960 (434 , 45.2%)	2086	2.17	Rat. (Con)	5
48	Invertebrates (Hymenoptera)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	229 (142 , 62%)	370	1.62	Rat. (Con)	4
49	Invertebrates (Hymenoptera-Ants)	Anthropogenic fragmented landscape	south-eastern Brazil	18	3-299 (ha)	120 (26 , 21.7%)	535	4.46	Gom. (Con)	49
50	Invertebrates (Hymenoptera-Ants)	Archipelago	Sea of Cortes	13	0.02-8.68 (km^2)	24 (9 , 37.5%)	84	3.50	Kob. (Con)	12
51	Invertebrates (Lepidoptera)	Anthropogenic fragmented landscape	Southern Spain	13	3.6-2115 (ha)	81 (21 , 25.9%)	481	5.94	Mon. (Con)	9
52	Invertebrates (Lepidoptera)	Archipelago	Aegean Islands	31	9-9254 (km^2)	127 (41 , 32.3%)	1052	8.28	Com.Log. (Con)	22
53	Invertebrates (Lepidoptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	606 (222 , 36.6%)	1576	2.60	Lin. (Con)	5
54	Invertebrates (Lepidoptera)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	163 (68 , 41.7%)	450	2.76	Lin. (Con)	4
55	Invertebrates (Lepidoptera)	Archipelago	Italian islands	10	40-22352 (ha)	76 (18 , 23.7%)	307	4.04	Pow. (Con)	20
56	Invertebrates (Lepidoptera)	Archipelago	Sardinian–Corsican islands	11	40-11559 (ha)	32 (5 , 15.6%)	175	5.47	Kob. (Con)	20
57	Invertebrates (Lepidoptera)	Archipelago	Sicilian islands	10	250-24600 (ha)	30 (5 , 16.7%)	160	5.33	Pow. (Con)	20
58	Invertebrates (Lepidoptera)	Archipelago	Tuscan islands	8	220-6030 (ha)	67 (23 , 34.3%)	198	2.96	Mon. (Con)	19
59	Invertebrates (Orthoptera)	Anthropogenic fragmented landscape	Small steppe patches, Buda Hills, Hungary	26	0.018-10.117 (ha)	32 (7 , 21.9%)	224	7.00	Lin. (Con)	8
60	Invertebrates (Orthoptera)	Anthropogenic fragmented	south-eastern Brazil	18	3-299	16 (9 ,	43	2.69	Kob. (Con)	49

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
		landscape			(ha)	56.3%)				
61	Invertebrates (Orthoptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	82 (26 , 31.7%)	236	2.88	Lin. (Con)	5
62	Invertebrates (Thysanoptera)	Archipelago	Canary islands	7	290.5-2007 (km^2)	68 (28 , 41.2%)	189	2.78	Pow. (Con)	5
63	Vertebrates (Herpatofauna)	Archipelago	Abaco Cays, West-Indian islands	21	0.05-16.4 (km^2)	8 (3 , 37.5%)	56	7.00	Pow. (Con)	47
64	Vertebrates (Herpatofauna)	Archipelago	Anguilla bank, West-Indian islands	21	0.007-90.7 (km^2)	40 (19 , 47.5%)	129	3.23	Com.Log. (Con)	47
65	Vertebrates (Herpatofauna)	Archipelago	Baubyan islands, Northern Philippines	5	0.7-196 (km^2)	44 (20 , 45.5%)	83	1.89	Exp. (Con)	45
66	Vertebrates (Herpatofauna)	Archipelago	Berry Islands, West-Indian islands	15	0.09-25.9 (km^2)	18 (4 , 22.2%)	69	3.83	Pow. (Con)	47
67	Vertebrates (Herpatofauna)	Archipelago	Bimini Islands, West-Indian islands	15	0.003-8.8 (km^2)	22 (9 , 40.9%)	71	3.23	Lin. (Con)	47
68	Vertebrates (Herpatofauna)	Archipelago	Caicos Cays, West-Indian islands	29	0.01-144 (km^2)	14 (2 , 14.3%)	130	9.29	Pow. (Con)	47
69	Vertebrates (Herpatofauna)	Archipelago	Central Exuma Cays, West-Indian islands	26	0.018-12.3 (km^2)	16 (2 , 12.5%)	108	6.75	Com.Log. (Con)	47
70	Vertebrates (Herpatofauna)	Archipelago	Crooked-Acklins bank , West-Indian islands	11	0.03-497 (km^2)	10 (2 , 20%)	35	3.50	Pow. (Con)	47
71	Vertebrates (Herpatofauna)	Archipelago	Grenada bank, West-Indian islands	35	0.01-32 (km^2)	28 (10 , 35.7%)	160	5.71	Com.Log. (Con)	47
72	Vertebrates (Herpatofauna)	Archipelago	Guadeloupe bank, West-Indian islands	14	0.005-22 (km^2)	17 (8 , 47.1%)	45	2.65	Pow. (Con)	47
73	Vertebrates (Herpatofauna)	Archipelago	Hispaniola bank, West-Indian islands	17	0.09-692 (km^2)	65 (29 , 44.6%)	154	2.37	Pow. (Con)	47
74	Vertebrates (Herpatofauna)	Archipelago	Jamaica bank, West-Indian islands	14	0.02-2.2 (km^2)	10 (3 , 30%)	27	2.70	Com.Log. (Con)	47
75	Vertebrates (Herpatofauna)	Archipelago	Keys of the northern coast of Cuba, West-	35	0.02-680	59 (15 ,	346	5.86	Pow. (Con)	47

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
			Indian islands		(km^2)	25.4%)				
76	Vertebrates (Herpatofauna)	Archipelago	Martinique bank, West-Indian islands	28	0.002-0.8 (km^2)	41 (4 , 9.8%)	77	1.88	Pow. (Con)	47
77	Vertebrates (Herpatofauna)	Archipelago	Mediterranean sea	14	84.9-25662 (km^2)	89 (40 , 44.9%)	234	2.63	Exp. (Con)	16
78	Vertebrates (Herpatofauna)	Archipelago	Puerto-Rico bank, West-Indian islands	92	0.001-137 (km^2)	57 (20 , 35.1%)	581	10.19	Pow. (Con)	47
79	Vertebrates (Herpatofauna)	Archipelago	Sea of Cortes, major islands	23	0.6-1173 (km^2)	85 (49 , 57.7%)	212	2.49	Pow. (Con)	12
80	Vertebrates (Herpatofauna)	Archipelago	Southern Exuma Cays, West-Indian islands	20	0.019-9.26 (km^2)	18 (8 , 44.4%)	68	3.78	Lin. (Con)	47
81	Vertebrates (Herpatofauna)	Archipelago	Turks bank, West-Indian islands	10	0.01-17.39 (km^2)	13 (6 , 46.2%)	42	3.23	Lin. (Con)	47
82	Vertebrates (Herpatofauna)	Political	Sonora, Mexico and adjoining states	7	58238-315194 (km^2)	416 (169 , 40.6%)	994	2.39	Neg.Exp. (Con)	25
83	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Australasia	66	1600-823000 (km^2)	528 (211 , 40%)	1712	3.24	Mon. (Con)	56
84	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Indo-Malaysia	89	2600-663600 (km^2)	711 (382 , 53.7%)	2220	3.12	Lin. (Con)	56
85	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Neoarctic	102	3900-753800 (km^2)	267 (87 , 32.6%)	1640	6.14	Neg.Exp. (Con)	56
86	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Neotropics	141	100-1916900 (km^2)	2167 (919 , 42.4%)	8190	3.78	Pow. (Con)	56
87	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Palearctic	182	2900-4639900 (km^2)	377 (117 , 31%)	1920	5.09	Neg.Exp. (Con)	56
88	Vertebrates (Herpatofauna-Amphibia)	Ecoregions	Sub-Saharan Africa	96	1000-3053200 (km^2)	629 (237 , 37.7%)	2669	4.24	Pow. (Con)	56
89	Vertebrates (Herpatofauna-Amphibia)	Inter-provincial	Global	9	16800-52731900 (km^2)	4587 (4355 , 94.9%)	4826	1.05	P2 (Con)	56
90	Vertebrates (Herpatofauna-Reptiles)	Archipelago	Canary islands	7	290.5-2007 (km^2)	15 (5 , 33.3%)	27	1.80	P2 (Con)	5

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
91	Vertebrates (Herpatofauna-Reptiles)	Archipelago	Islands, north-eastern Adriatic coast	14	15-410 (km^2)	28 (6 , 21.4%)	170	6.07	Lin. (Con)	39
92	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Australasia	69	100-823000 (km^2)	1216 (412 , 33.9%)	6042	4.97	Gom. (Con)	56
93	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Indo-Malaysia	92	100-663600 (km^2)	1252 (537 , 42.9%)	5978	4.77	Pow. (Con)	56
94	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Neoarctic	89	100-753800 (km^2)	474 (135 , 28.5%)	3226	6.81	Mon. (Con)	56
95	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Neotropics	145	100-1916900 (km^2)	2164 (771 , 35.6%)	11768	5.44	Pow. (Con)	56
96	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Palearctic	180	1400-4639900 (km^2)	789 (242 , 30.7%)	4931	6.25	Mon. (Con)	56
97	Vertebrates (Herpatofauna-Reptiles)	Ecoregions	Sub-Saharan Africa	98	200-3053200 (km^2)	1330 (478 , 35.9%)	6913	5.20	Wei.3 (Con)	56
98	Vertebrates (Herpatofauna-Reptiles)	Inter-provincial	Global	9	16800-52731900 (km^2)	6856 (6168 , 90%)	7607	1.11	Neg.Exp. (Con)	56
99	Vertebrates (Herpatofauna-Lizards)	Anthropogenic fragmented landscape	Reserves, Western Australia	23	34-5119 (ha)	69 (22 , 31.9%)	384	5.57	Mon. (Con)	37
100	Vertebrates (Herpatofauna-Lizards)	Anthropogenic fragmented landscape	Western Australia	26	0.5-174 (ha)	15 (2 , 13.3%)	106	7.07	Pow. (Con)	51
101	Vertebrates (Herpatofauna-Lizards)	Archipelago	Sea of Cortes	9	0.6-187 (km^2)	13 (1 , 7.7%)	77	5.92	Neg.Exp. (Con)	12
102	Vertebrates (Aves)	Anthropogenic fragmented landscape	Brazil	12	0.09-1.02 (ha)	19 (4 , 21.1%)	73	3.84	Lin. (Con)	2
103	Vertebrates (Aves)	Anthropogenic fragmented landscape	Canyon habitats, San Diego County, California, USA	37	0.4-102 (ha)	9 (1 , 11.1%)	73	8.11	Lin. (Con)	52
104	Vertebrates (Aves)	Anthropogenic fragmented landscape	Forest islands, central New-Jersey, USA	10	0.01-24 (ha)	35 (0 , 0%)	205	5.86	Pow. (Con)	29
105	Vertebrates (Aves)	Anthropogenic fragmented landscape	Sewage works, Britain	12	3-400 (ha)	24 (5 , 20.8%)	109	4.54	Mon. (Con)	28
106	Vertebrates (Aves)	Anthropogenic fragmented	Singapore	17	7-935	166 (26 ,	1234	7.43	Mon. (Con)	13

	Major Taxon (Taxon)	*system*	*Location*	*n*	*Area range (units)*	*Species (endemics, % endemics)*	*Number of occupancies*	*mean occupancy level*	*Best Model (Ori/Con)*	*Ref'*
		landscape			(ha)	15.7%)				
107	Vertebrates (Aves)	Anthropogenic fragmented landscape	Small woodlots, Wisconsin, USA	9	0.2-4.41 (ha)	26 (8 , 30.8%)	108	4.15	Neg.Exp. (Con)	36
108	Vertebrates (Aves)	Anthropogenic fragmented landscape	Urban parks, Madrid, Spain	25	1-118.2 (ha)	32 (6 , 18.8%)	293	9.16	Exp. (Con)	27
109	Vertebrates (Aves)	Archipelago	Baubyan islands, Northern Philippines	5	0.7-196 (km^2)	137 (46 , 33.6%)	332	2.42	Kob. (Con)	45
110	Vertebrates (Aves)	Archipelago	Canary islands	7	290.5-2007 (km^2)	78 (9 , 11.5%)	364	4.67	Lin. (Con)	5
111	Vertebrates (Aves)	Archipelago	Cape-Verde	12	1.4-991 (km^2)	55 (4 , 7.3%)	344	6.25	Pow. (Con)	4
112	Vertebrates (Aves)	Archipelago	Dahlak Archipelago	26	2-2143 (ha)	38 (16 , 42.1%)	162	4.26	Com.Log. (Con)	6
113	Vertebrates (Aves)	Archipelago	Northern islands, Sea of Cortes	16	0.03-15.03 (km^2)	32 (10 , 31.3%)	147	4.59	Rat. (Con)	12
114	Vertebrates (Aves)	Archipelago	Southern islands, Sea of Cortes	16	0.05-187 (km^2)	28 (0 , 0%)	218	7.79	Kob. (Con)	12
115	Vertebrates (Aves)	Archipelago	Thousand Island lake, china	42	0.3-1289 (ha)	93 (25 , 26.9%)	1193	12.83	Arc.Log. (Con)	54
116	Vertebrates (Aves)	Ecoregions	Australasia	69	100-823000 (km^2)	1605 (409 , 25.5%)	16038	9.99	Mon. (Con)	56
117	Vertebrates (Aves)	Ecoregions	Indo-Malaysia	93	300-663600 (km^2)	1781 (161 , 9%)	30768	17.28	Exp. (Con)	56
118	Vertebrates (Aves)	Ecoregions	Neoarctic	118	100-1032800 (km^2)	728 (74 , 10.2%)	21702	29.81	Com.Log. (Con)	56
119	Vertebrates (Aves)	Ecoregions	Neotropics	150	100-1916900 (km^2)	3687 (569 , 15.4%)	54165	14.69	Exp. (Con)	56
120	Vertebrates (Aves)	Ecoregions	Palearctic	195	1400-4639900 (km^2)	1570 (159 , 10.1%)	46043	29.33	Mon. (Con)	56
121	Vertebrates (Aves)	Ecoregions	Sub-Saharan Africa	100	100-3053200 (km^2)	2046 (239 , 11.7%)	36905	18.04	Exp. (Con)	56

	Major Taxon (Taxon)	*system*	*Location*	*n*	*Area range (units)*	*Species (endemics, % endemics)*	*Number of occupancies*	*mean occupancy level*	*Best Model (Ori/Con)*	*Ref'*
122	Vertebrates (Aves)	Inland water bodies	Bogs, Wetland habitats, SW Sweden	47	8-66 (ha)	13 (0 , 0%)	161	12.38	Neg.Exp. (Con)	33
123	Vertebrates (Aves)	Inland water bodies	Wet meadows, Wetland habitats, SW Sweden	15	2-22 (ha)	11 (2 , 18.2%)	60	5.45	Neg.Exp. (Con)	33
124	Vertebrates (Aves)	Inter-provincial	Global	9	16800-52731900 (km^2)	9008 (6965 , 77.3%)	11861	1.32	Neg.Exp. (Con)	56
125	Vertebrates (Aves)	Naturally fragmented	Forest patches, Southern Brazil	12	0.5-840 (ha)	189 (36 , 19.1%)	938	4.96	Com.Log. (Con)	3
126	Vertebrates (Aves)	Naturally fragmented	Oaxaca, Mexico	17	2-159246 (ha)	60 (0 , 0%)	478	7.97	Gom. (Con)	55
127	Vertebrates (Aves-Birds of prey)	Archipelago	Mediterranean sea	43	143-25662 (km^2)	25 (0 , 0%)	307	12.28	Pow. (Con)	23
128	Vertebrates (Mammals)	Anthropogenic fragmented landscape	Atlantic forest fragments, Brazil	8	1.2-13.3 (ha)	12 (1 , 8.3%)	62	5.17	Neg.Exp. (Con)	21
129	Vertebrates (Mammals)	Anthropogenic fragmented landscape	Reserves, Western Australia	23	34-5119 (ha)	24 (5 , 20.8%)	171	7.13	Neg.Exp. (Con)	38
130	Vertebrates (Mammals)	Anthropogenic fragmented landscape	Temperate rain forest, Olympic Peninsula, Washington, USA	20	0.93-58.91 (ha)	18 (2 , 11.1%)	142	7.89	Lin. (Con)	42
131	Vertebrates (Mammals)	Archipelago	Alexander archipelago, Alaska	24	10.1-5777 (km^2)	23 (4 , 17.4%)	199	8.65	Com.Log. (Con)	15
132	Vertebrates (Mammals)	Archipelago	Baubyan islands, Northern Philippines	5	0.7-196 (km^2)	20 (13 , 65%)	35	1.75	Pow. (Con)	45
133	Vertebrates (Mammals)	Archipelago	Great Salt Lake, Utah, USA	7	9-10767 (ha)	27 (11 , 40.7%)	63	2.33	Lin. (Con)	17
134	Vertebrates (Mammals)	Archipelago	Islands, Gulf of Maine, USA	8	1.243-279 (km^2)	35 (11 , 31.4%)	122	3.49	Lin. (Con)	18
135	Vertebrates (Mammals)	Archipelago	Islands, north-eastern Adriatic coast	14	15-410 (km^2)	13 (2 , 15.4%)	86	6.62	Lin. (Con)	39
136	Vertebrates (Mammals)	Archipelago	Philippine Trench	9	22-99078 (km^2)	35 (15 , 42.9%)	123	3.51	Lin. (Con)	35
137	Vertebrates (Mammals)	Archipelago	Sea of Cortes	28	0.32-1173	77 (75 ,	79	1.03	Arc.Log. (Con)	12

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
					(km^2)	97.4%)				
138	Vertebrates (Mammals)	Archipelago	Thousand Islands Region, New-York, USA	20	0.04-591 (ha)	10 (0 , 0%)	50	5.00	Gom. (Con)	40
139	Vertebrates (Mammals)	Ecoregions	Australasia	67	100-823000 (km^2)	674 (166 , 24.6%)	3684	5.47	Com.Log. (Con)	56
140	Vertebrates (Mammals)	Ecoregions	Indo-Malaysia	93	300-663600 (km^2)	830 (126 , 15.2%)	11697	14.09	Exp. (Con)	56
141	Vertebrates (Mammals)	Ecoregions	Neoarctic	119	100-1032800 (km^2)	481 (65 , 13.5%)	7658	15.92	Neg.Exp. (Con)	56
142	Vertebrates (Mammals)	Ecoregions	Neotropics	146	100-1916900 (km^2)	1227 (127 , 10.4%)	23383	19.06	P2 (Con)	56
143	Vertebrates (Mammals)	Ecoregions	Palearctic	193	2900-4639900 (km^2)	905 (120 , 13.3%)	14046	15.52	Mon. (Con)	56
144	Vertebrates (Mammals)	Ecoregions	Sub-Saharan Africa	98	200-3053200 (km^2)	1039 (195 , 18.8%)	10964	10.55	Pow. (Con)	56
145	Vertebrates (Mammals)	Inter-provincial	Global	9	16800-52731900 (km^2)	4541 (3846 , 84.7%)	5295	1.17	Neg.Exp. (Con)	56
146	Vertebrates (Mammals)	Naturally fragmented	Mountain-tops, great basin of north America, USA	17	31.1-3051 (km^2)	13 (1 , 7.7%)	97	7.46	Com.Log. (Con)	11
147	Vertebrates (Mammals)	Naturally fragmented	Montane islands, American Southwest	27	6.89-11134 (km^2)	23 (5 , 21.7%)	154	6.70	Kob. (Con)	41
148	Vertebrates (Mammals-Bats)	Archipelago	Antillean islands	22	13-105805 (km^2)	57 (31 , 54.4%)	189	3.32	Kob. (Con)	7
149	Vertebrates (Mammals-Bats)	Archipelago	Islands, Bahamas	23	2.18-5959 (km^2)	13 (4 , 30.8%)	117	9.00	Pow. (Con)	48
150	Vertebrates (Mammals-Bats)	Archipelago	Islands, Greater Antilles	19	5.2-105805 (km^2)	37 (13 , 35.1%)	209	5.65	P2 (Con)	48
151	Vertebrates (Mammals-Bats)	Archipelago	Islands, Lesser Antilles	23	5.49-1628 (km^2)	24 (7 , 29.2%)	225	9.38	Pow. (Con)	48
152	Vertebrates (Mammals-Rodents)	Anthropogenic fragmented landscape	Coastal Southern California	24	0.41-84 (ha)	9 (0 , 0%)	67	7.44	Neg.Exp. (Con)	10

	Major Taxon (Taxon)	***system***	***Location***	***n***	***Area range (units)***	***Species (endemics, % endemics)***	***Number of occupancies***	***mean occupancy level***	***Best Model (Ori/Con)***	***Ref'***
153	Vertebrates (Mammals-Rodents)	Archipelago	Virginia barrier islands, USA	9	29-2197 (ha)	5 (0 , 0%)	19	3.80	Pow. (Con)	24
154	Vertebrates (Mammals-Monkeys)	Anthropogenic fragmented landscape	Udzungwa Mountains of Tanzania	22	0.06-526.32 (km^2)	7 (1 , 14.3%)	69	9.86	Kob. (Con)	43

Pow. – Power; **Pow.Ros.** – Power Rosenzweig; **Ext.P1**– Extended Power 1; **Ext.P2**– Extended Power 2; **P1** – Persistence Function 1; **P2** – Persistence Function 2; **Exp.** – Exponential; **Kob.** – Kobayashi Logarithmic; **Mon.** – Monod; **MMF.** – Morgan-Mercer-Flodin; **Arc.Log.** – Archibald Logistic; **Neg.Exp.** – Negative Exponential; **Chp.Ric.** – Chapman-Richards; **Wei.3** – Weibull-3; **Wei.4** – Weibull-4; **Asy.** – Asymptotic; **Rat.** – Rational. **Gom.** – Gompertz; **Beta.P.** – Beta-P; **Com.Log.** - Common Logistic; **EVF.** – Extreme-Value Function; **Lom.** – Lomolino function.

Appendix S3

Linear regression between the expected and observed *wAICc* of the constrained form (table S3) with additional focus on the Monod and negative exponential functions (figure S1).

Table S3:

For each of the twelve functions, the result of linear regression of the observed wAICc of the constrained form against its expected wAICc, if the original and constrained forms have identical log likelihoods. *N* is the number of datasets used for the regression, while 'Low' and 'High' stand for the lower and higher values of the 95% confidence intervals around the intercept and slope. Cases in which the confidence interval of the intercept or slope did not overlap with 0 and 1 (respectively) are given in bold face.

				Constant			*Slope*		
Model	***N***	***R^2***	***Sig.***	***Estimate***	***Low***	***High***	***Estimate***	***Low***	***High***
Power	154	0.772	<0.001	-0.002	-0.076	0.071	0.997	0.910	1.084
Linear	154	0.998	<0.001	-0.001	-0.007	0.005	1.000	0.993	1.008
Kobayashi	154	0.755	<0.001	-0.069	-0.152	0.014	1.068	0.970	1.165
Exponential	154	1.000	<0.001	0.000	-0.001	0.000	1.000	1.000	1.001
Monod	149	0.441	<0.001	**-0.329**	**-0.538**	**-0.120**	**1.338**	**1.092**	**1.584**
Negative Exponential	151	0.389	<0.001	**-0.470**	**-0.726**	**-0.214**	**1.481**	**1.180**	**1.782**
P2	151	0.820	<0.001	0.021	-0.044	0.086	0.972	0.898	1.045
Weibull	151	0.951	<0.001	0.011	-0.021	0.043	0.988	0.952	1.025
Gompertz	145	0.999	<0.001	0.000	-0.004	0.005	0.999	0.994	1.004
Common Logistic	149	0.994	<0.001	-0.003	-0.014	0.008	1.004	0.991	1.017
Archibald Logistic	151	0.903	<0.001	0.035	-0.009	0.080	0.961	0.910	1.012
Rational	148	0.998	<0.001	-0.002	-0.008	0.004	1.002	0.994	1.009

Figure S1: The difference between the observed and expected (under identical log-likelihood) AICc weight of the constrained form, plotted against the cumulative AICc weight of the SAR function in the 24 SAR models analysis, for the (a) Monod and (b) Negative Exponential SAR functions. Note that deviation from zero difference occurs when the model poorly describes the empirical data (relative to other SAR functions).