

ORIGINAL RESEARCH

On piecewise models and species–area patterns

De Gao¹  | Zhen Cao² | Peng Xu³ | Gad Perry⁴

¹School of Resources and Environmental Science, Hebei Normal University, Shijiazhuang, China

²Department of Chemical and Environmental Engineering, Hebei College of Industry and Technology, Shijiazhuang, China

³Department of Mathematics and Statistics, Eastern Michigan University, Ypsilanti, Michigan, USA

⁴Department of Natural Resources Management, Texas Tech University, Lubbock, Texas, USA

Correspondence

Zhen Cao, Department of Chemical and Environmental Engineering, Hebei College of Industry and Technology, Shijiazhuang, Hebei 050091, China.
Email: caozhen81112@126.com

Funding information

Hebei Normal University, Grant/Award Number: L2018B21; Hebei Education Department, Grant/Award Number: GCC2014037; Hebei Provincial Science and Technology Department, Grant/Award Number: D2019205019

Abstract

Aim: Area thresholds, at which the form of the species–area relationship (SAR) changes abruptly, have played an important role in the theoretical framework of conservation biogeography and biodiversity research. The application of piecewise regressions has been advocated as a rigorous statistical technique to identify such thresholds within SARs, but a large variety of piecewise models remains untested. We explore the prevalence and number of thresholds in SARs and examine whether the currently widely used method for detecting the small island effect (SIE) is robust.

Location: Global.

Taxon: We consider all multicellular taxa based on the criteria of datasets selection.

Methods: We apply 15 regression models, including linear regression and piecewise regressions with two and three segments to 68 global island datasets that are sourced from the literature.

Results: The number of area thresholds in SARs varied among groups and correlated positively with area range of a studied system. Under the AIC or AIC_c criterion, three-segment piecewise models were more prevalent, whereas under the BIC criterion, two-segment piecewise models were more prevalent. From the results of Aegean Sea isopods, West Indies herpetofauna, and Australian Islands mammals, we found evidence that the traditional criteria for detection of SIEs are not robust.

Main conclusions: Our study demonstrates that (a) to detect an SIE, the comparison should use as many models as possible, including not only variants with and without a left-horizontal part, but also those with two and more segments; (b) naive use of the traditional two-segment piecewise regressions may cause poor estimations of both slope and breakpoint values; (c) the number of thresholds increases with the area range of a studied system; (d) conservation biologists and applied ecologists should determine the number of area thresholds when estimating the precise species–area patterns and making management strategies in fragmented landscapes.

KEYWORDS

breakpoint values, conservation biogeography, island biogeography, piecewise regression, small island effect, species–area relationships, thresholds, zero-slope regression

1 | INTRODUCTION

The species–area relationship (SAR), fundamental to the development of MacArthur and Wilson's (1963, 1967) equilibrium theory and island biogeography, has great utility for assessing species diversity and conservation needs (Ladle & Whittaker, 2011; Lomolino, 2000). Insights from SARs have been applied in optimal reserve design (Neigel, 2003), identifying biodiversity hotspots (Fattorini, 2007), and assessing the impact that habitat fragmentation exerts on diversity (Harcourt & Doherty, 2005). However, debate continues as to the applicability of island biogeographic theory in the context of habitat fragmentation (Laurance, 2008), because as species–area studies began to accumulate, scholars have found evidence for non-linearity of log-transformed SARs.

Some studies showing thresholds in SARs depict a steep–shallow relationship, with an initially rapid increase in richness with area until a threshold is reached beyond which a slow rate of increase applies (Drinnan, 2005; Fahrig, 2001; Matthews, Steinbauer, Tzirkalli, Triantis, & Whittaker, 2014). Others depict a shallow–steep relationship, showing an initially slow rate of (or no) increase in richness with increasing area, followed by a steeper phase beyond a threshold, often named the small island effect (SIE; Morrison, 2014; Triantis et al., 2006; Wang, Chen, & Millien, 2018; Wang, Millien, & Ding, 2016). Although around 20 functions have been described for fitting SARs (Triantis, Guilhaumon, & Whittaker, 2012), piecewise regression has been advocated as a rigorous statistical technique suitable for identifying area thresholds (Ficetola & Denoël, 2009).

Currently, four piecewise regression functions are found in the literature: (a) continuous left-horizontal function [$Y = c + (\log A > T) z (\log A - T)$] (Lomolino & Weiser, 2001); (b) continuous two-slope function [$Y = c + (\log A \leq T) z_1 \log A + (\log A > T) [z_1 T + z_2 (\log A - T)]$] (Dengler, 2010; Morrison & Spiller, 2008); (c) discontinuous two-slope function [$Y = (\log A \leq T) (c_1 + z_1 \log A) + (\log A > T) (c_2 + z_2 \log A)$] (Gentile & Argano, 2005); and (d) continuous right-horizontal function [$Y = c + (\log A \leq T) z \log A + (\log A > T) z T$] (Dengler, 2010)]. Several shortcomings in the applications of piecewise models persist, however.

First, it is conceivable that discontinuous relationships with disconnected SAR segments may exist in nature (Gentile & Argano, 2005; Maron et al., 2012), and their detection in a few datasets most likely signifies important roles for confounding variables that were not included in the models, such as isolation or matrix effects (Crowe, 1979; Levenson, 1981; Matthews et al., 2014). However, discontinuous function is only allowed in the two-slope method, where two linear segments are fitted onto the data (Gentile & Argano, 2005).

Second, the current continuous models include one segment obtained by regression and the other one obtained by nonregression: slope iteration for the continuous two-slope function (Dengler, 2010; Morrison & Spiller, 2008) or direct inheritance for the continuous left-horizontal (Lomolino & Weiser, 2001) and the continuous

right-horizontal (Dengler, 2010) functions. However, either segment could be obtained by regression, that is, if the upper segment is the regression part, the lower segment will be the nonregression one, and vice versa. So, each method actually has two possible expression and calculation forms rather than one, which is neglected in the literature.

Third, piecewise models with two segments have been widely used in previous studies (e.g., Dengler, 2010; Lomolino & Weiser, 2001; Wang et al., 2016). However, Lomolino and Weiser (2001) and Rosenzweig (2004) proposed three biological scales of species–area curves with three corresponding dominant processes of species addition: (a) stochastic extinction forces structure insular communities on small islands; (b) more deterministic, ecological factors associated with habitat diversity, carrying capacity and extinction/immigration dynamics as envisioned by MacArthur and Wilson (1967) on islands of intermediate size; and (c) in situ speciation on relatively large islands. Therefore, limiting our attention to just one or two species–area patterns is unreasonable.

In recent work on detecting the SIE of herpetofauna of the West Indies, Gao and Perry (2016a) showed that piecewise regressions with three segments performed better than those of two segments, suggesting piecewise models with more than two segments should be considered. Dengler (2010) proposed that model comparisons should include at least one variant without an SIE, one with a left-horizontal function, and one with a two-slope function. And thereafter, datasets fitted best by the left-horizontal function are thought to provide stronger evidence for an SIE than those fitted best by the two-slope function (Morrison, 2014; Wang et al., 2016). However, Dengler's (2010) proposal is based on currently incomprehensive piecewise models, which are restricted to a maximum of two slopes, so we question the reliability of the method for detecting the SIE.

The aim of this paper was thus twofold. First, we provide an updated synthetic analysis of piecewise models to the SARs of 68 island datasets, representing different geographic regions, taxa, and size ranges. We discuss whether the currently widely used method for detecting the SIE is robust when other piecewise models are taken into account. Second, we examine whether larger sample size range and larger species range lead to detection of more area thresholds.

2 | METHODS

2.1 | Data collection

Between May 2015 and July 2017, we searched for true and habitat island studies and relevant datasets within four main abstracting databases (JSTOR, ISI Web of Knowledge, and BIOSIS Biological Abstracts) using the keywords “species richness,” “fragments,” “small island effect,” and “islands” in different combinations. Cross-referenced papers derived from the reference lists of sourced papers were also included. More than a thousand journal papers, books, reports, online databases, and unpublished resources were screened. To increase the statistical power for the

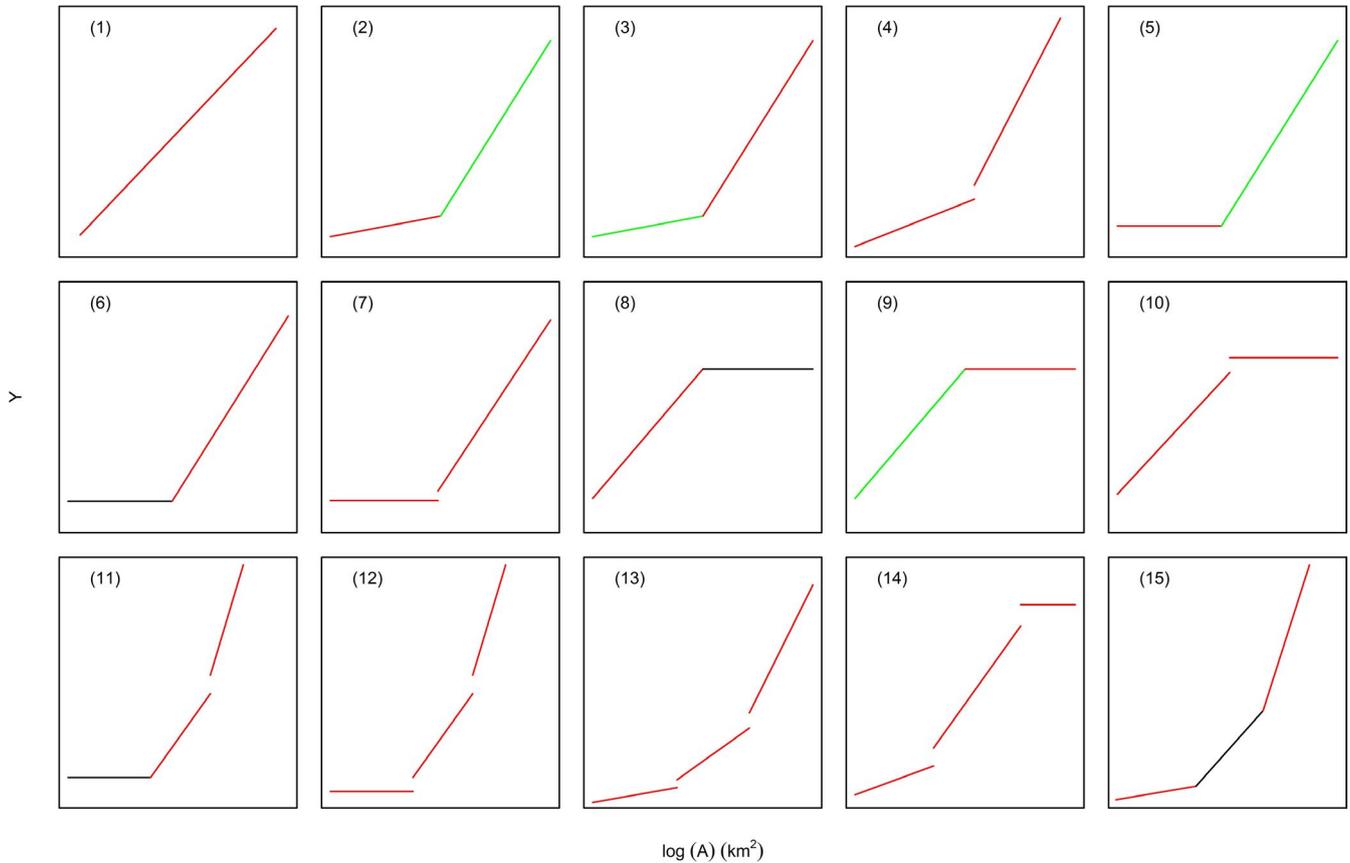


FIGURE 1 Schematic illustration of the function and derivation of each segment in each model. Red line signifies the segment is obtained by regression; green line signifies segment is obtained by slope iteration; and black line signifies the segment is obtained by direct inheritance. The illustrated slope of each segment (excluding the horizontal segment) in each model does not imply the actual relative shallowness or steepness in data analysis. We denote the x-axis as $\log(A)$ (km^2), and the y-axis as Y, which stands for species richness S or log S for semi-log and log-log versions of the model, respectively

robustness of piecewise models, however, we only included datasets that met the following criteria: (a) islands constituted discrete patches of habitat surrounded by contrasting habitat; (b) there were at least 40 islands within each dataset; (c) the area and species richness of each island were accessible either in the source publication or from the authors of the source papers; and (d) the dataset derived from a source did not overlap with those from any other sources (data for different taxa within the same study system were accepted).

For Aegean Sea isopods, we removed the largest island from the dataset and formed a new dataset to facilitate further comparison. Finally, 68 datasets from 38 sources (see Appendix S1 in Supporting Information) were retained for further analyses.

2.2 | Statistical analyses

We fitted a linear model (Equation 1) alongside 14 piecewise models (Equations 2–15; Figure 1) to each dataset.

$$Y = c_1 + z_1 \log A \quad (1)$$

$$Y = c_1 + (\log A \leq T_1)z_1 \log A + (\log A > T_1)[z_1 T_1 + z_2(\log A - T_1)] \quad (2)$$

$$Y = c_1 + (\log A \leq T_1)[z_1 \log A + (z_2 - z_1)T_1] + (\log A > T_1)z_2 \log A \quad (3)$$

$$Y = (\log A \leq T_1)(c_1 + z_1 \log A) + (\log A > T_1)(c_2 + z_2 \log A) \quad (4)$$

$$Y = c_1 + (\log A > T_1)z_1(\log A - T_1) \quad (5)$$

$$Y = c_1 + (\log A \leq T_1)z_1 T_1 + (\log A > T_1)z_1 \log A \quad (6)$$

$$Y = (\log A \leq T_1)c_1 + (\log A > T_1)(c_2 + z_1 \log A) \quad (7)$$

$$Y = c_1 + (\log A \leq T_1)z_1 \log A + (\log A > T_1)z_1 T_1 \quad (8)$$

$$Y = c_1 + (\log A \leq T_1)z_1(\log A - T_1) \quad (9)$$

$$Y = (\log A \leq T_1)(c_1 + z_1 \log A) + (\log A > T_1)c_2 \quad (10)$$

$$Y = (\log A \leq T_2)[c_1 + (\log A \leq T_1)z_1 T_1 + (\log A > T_1)z_1 \log A] + (\log A > T_2)(c_2 + z_2 \log A) \quad (11)$$

$$Y = (\log A \leq T_1)c_1 + (\log A > T_1 \text{ AND } \log A \leq T_2)(c_2 + z_1 \log A) + (\log A > T_2)(c_3 + z_2 \log A) \quad (12)$$

$$Y = (\log A \leq T_1)(c_1 + z_1 \log A) + (\log A > T_1 \text{ AND } \log A \leq T_2)(c_2 + z_2 \log A) + (\log A > T_2)(c_3 + z_3 \log A) \quad (13)$$

$$Y = (\log A \leq T_1)(c_1 + z_1 \log A) + (\log A > T_1 \text{ AND } \log A \leq T_2)(c_2 + z_2 \log A) + (\log A > T_2)c_3 \quad (14)$$

$$Y = (\log A \leq T_1)(c_1 + z_1 \log A) + (\log A > T_1 \text{ AND } \log A \leq T_2) [(c_1 - c_2 + z_1 T_1 - z_2 T_2)(\log A - T_1)/(T_1 - T_2) + c_1 + z_1 T_1] + (\log A > T_2)(c_2 + z_2 \log A) \quad (15)$$

In these equations, Y stands for species richness S or $\log S$ for semi-log and log-log versions of the model, respectively, A for area (in the unit of km^2), while c_i (intercept), z_i (slope), and T_i (breakpoint) are fitted parameters. The logical AND operator combines two logical operands that have a value true or false. The expression combined by logical AND evaluates to true if both operands $\log A > T_1$ and $\log A \leq T_2$ evaluate to true; if either or both of the operands for the logical AND operator are false, the result of the expression is false. The logical expressions in brackets return value 1 if they are true and 0 if they are false. Islands that have no species recorded were involved in multiple datasets (see Table S1 in Appendix S1). In some datasets, there were lots of small islands, very few large ones, and a large gap in between. To improve statistical rigor and avoid outlier effects, all datasets, with the exception of the three large datasets mentioned below, were analyzed with the log-log versions of the model. Since removing observations is not favored (Williamson, 1981), we used the traditional $\log(S + 1)$ for the datasets that includes empty islands, which is consistent with the treatment applied by Suarez, Bolger, and Case (1998), Lavin et al. (2001), and Gao and Perry (2016a). For Aegean Sea isopods, Aegean Sea centipedes, and Italian islands centipedes (see Table S1 in Appendix S1), we used S rather than $\log S$ based on two considerations: (a) Species richness values distribute pretty continuously, and log transformation is unnecessary and (b) both Sfenthourakis and Triantis (2009) and Dengler (2010) have used the untransformed version in their analyses, so we follow this to facilitate the comparison among our results as well as to those of Dengler (2010).

Equations 2–10 are the comprehensive possible function forms for the two-segment piecewise method, in which Equations 2 and 3 comprise a regression segment and a slope-iteration segment; Equation 4 comprises two regression segments; Equations 5 and 9 comprise a zero-slope regression segment and a slope-iteration segment; Equations 7 and 10 comprise a regression segment and a zero-slope regression segment; and Equations 6 and 8 comprise a regression segment and a direct-inheritance segment (Figure 1). Equations 2–4, 5–7, and 8–10 are the two-slope, the left-horizontal, and the right-horizontal variants, respectively, under the two-segment condition. As for the three-segment condition, due to the option of continuity or discontinuity at each breakpoint and performance (regression/slope iteration/direct inheritance) of each segment, there will be many more possible forms than the two-segment condition. So, here, we just randomly choose five forms (Equations 11–15) based on the consideration of computational load. Equations

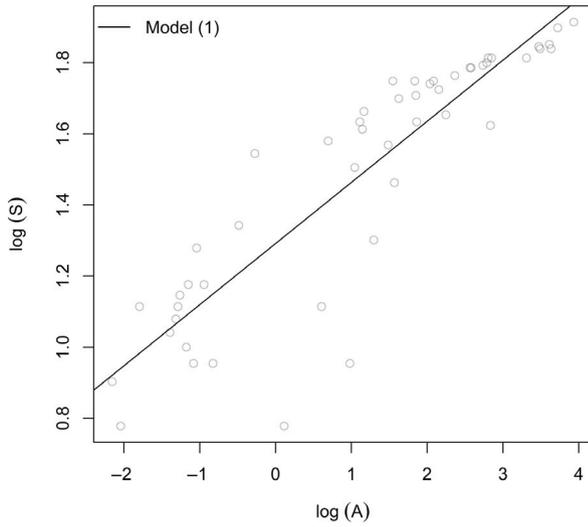
12–14 comprise three regression segments, in which, a zero-slope regression segment constitutes Equations 12 and 14; Equations 11 and 15 comprise two regression segments and a direct-inheritance segment (Figure 1).

If a piecewise model is discontinuous, the model expression will be relatively easier, and each segment can be depicted as ($Y = c$) for a zero-slope regression or ($Y = c + z \log A$) for a normal regression. However, if a piecewise model is continuous, the model expression will be more complicated, as the right end of a lower segment has to be connected to the left end of an upper segment. And thus, to unify the expressions, we give priority to the expression of regression segments and denote ($Y = c$) for zero-slope regression segments and ($Y = c + z \log A$) for linear regression segments. Therefore, Model 6 (Figure 2; expressed as [$Y = c + (\log A > T) z (\log A - T)$ (Lomolino & Weiser, 2001)]) is denoted here as [$Y = c_1 + (\log A \leq T_1) z_1 T_1 + (\log A > T_1) z_1 \log A$], because, in this way, $Y = c_1 + z_1 T_1$ when $\log A \leq T_1$ and $Y = c_1 + z_1 \log A$ when $\log A > T_1$, and it can be clearly seen that the upper, rather than the lower, part is a regression segment. Instead, the function provided by Lomolino and Weiser (2001) is consistent with Equation 5 here, denoting a lower zero-slope regression segment and an upper slope-iteration segment (Figure 1).

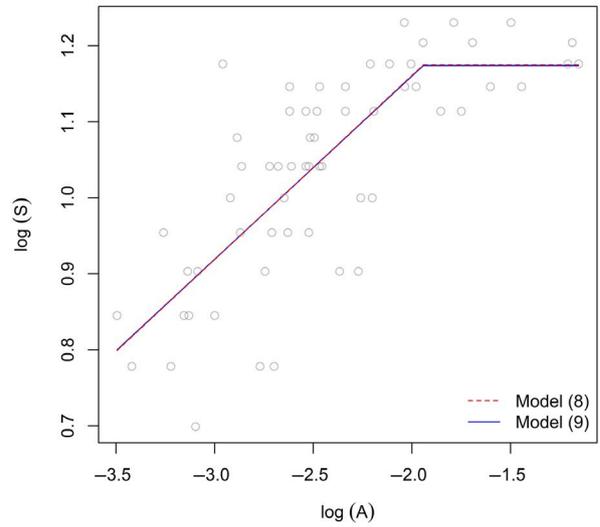
We used a minimum residual sum of squares (RSS) method (Gao & Perry, 2016a) to estimate the threshold values. For Equations 2–15, the parameters were estimated by using nonlinear estimation procedures based on iteration. If a piecewise model is continuous at a breakpoint, then the breakpoint lying between two adjacent data points will influence the RSS value of the model, we incremented such breakpoint values (T) by 0.001; however, if a piecewise model is discontinuous at a breakpoint, then the breakpoint lying between two adjacent data points will not influence the RSS value of the model, so we assigned the breakpoint values (T) to the log-transformed area values of each island (see Fig. S2–S10 in Appendix S2). Equations 12–14 are discontinuous, so we assigned the first breakpoint values (T_1) to the log-transformed area values of each island, and at any particular value of T_1 , T_2 was assigned to the log-transformed area values between T_1 and the maximum log-transformed area value. We recorded the minimum RSS value produced by the iteration of T_2 for each particular value of T_1 , so that the first breakpoint (T_1) was determined prior to the second one (T_2). After T_1 was determined, we run iteration of T_2 again to look for the T_2 that produced the minimum RSS value (see Fig. S12–S14 in Appendix S2). For Equations 11 and 15, we assigned the second breakpoint values (T_2) to the log-transformed area values of each island or incremented T_2 by 0.001 for Equations 11 and 15, respectively. And T_1 was incremented from the minimum log-transformed area value to T_2 by 0.001. We recorded the minimum RSS value produced by the iteration of T_1 for each particular value of T_2 , so that the second breakpoint (T_2) was determined prior to the first one (T_1). After T_2 was determined, we run iteration

FIGURE 2 Results of model selection for six sample datasets: (a) Solomon Islands birds (sea-level species) by BIC, (b) Austria semi-natural grassland Orthoptera species by BIC, (c) Australian Islands mammals by AIC_c, (d) and (e) Aegean Sea isopods by AIC_c, and (f) West Indies herpetofauna by AIC. We analyzed Aegean Sea isopods for both with (d) and without (e) the largest island

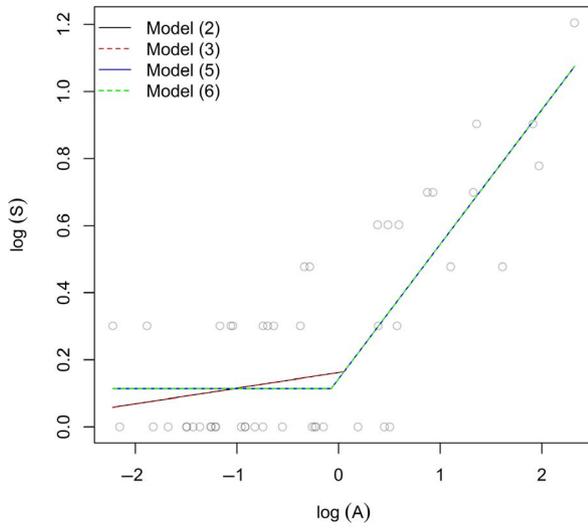
(a) Solomon Islands birds (sea-level species)



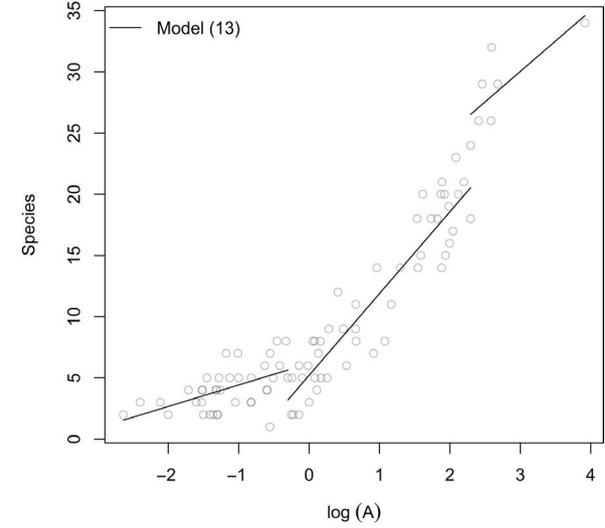
(b) Austria semi-natural grassland Orthoptera



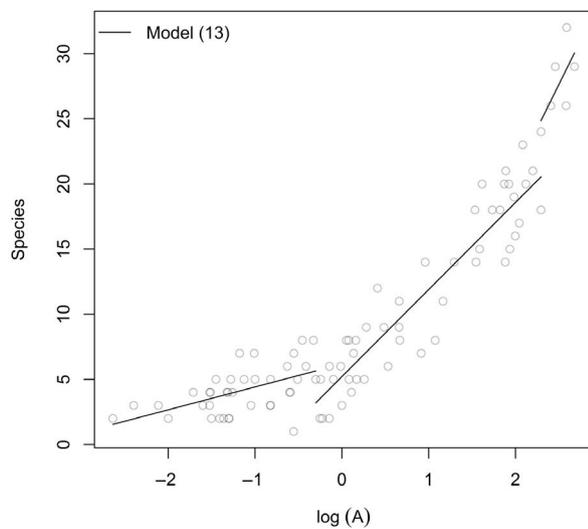
(c) Australian Islands mammal



(d) Aegean Sea isopod



(e) Aegean Sea isopod (the largest island removed)



(f) West Indies herpetofauna

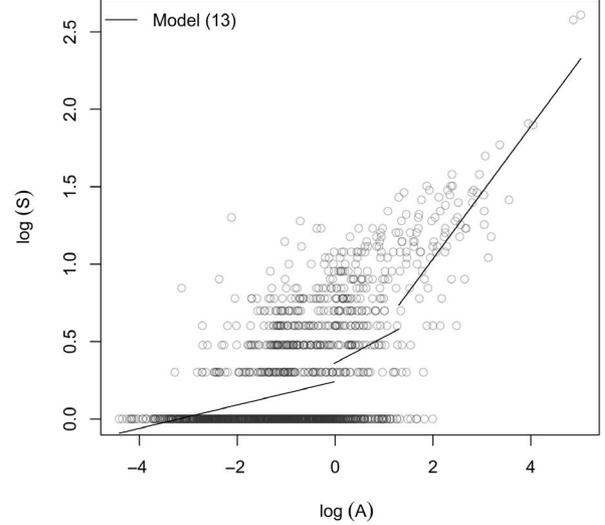


TABLE 1 The distribution of occurrence frequency across model segments and threshold forms determined under $AIC_{(c)}$ and BIC model selection criteria, respectively, for the 68 global datasets. In our study, models with $\Delta IC \leq 2$ were equally supported, and therefore, the sum of the numbers did not equal 68 (total sample size)

Model segments	$AIC_{(c)}$			BIC		
	SIE threshold	Steep-shallow threshold	None threshold	SIE threshold	Steep-shallow threshold	None threshold
Linear model	—	—	12	—	—	31
Two-segment models	35	11	—	36	13	—
Three-segment models	42	26	—	16	6	—
Total cases	58	31	12	46	19	31

of T_1 again to look for the T_1 that produced the minimum RSS value (see Fig. S11; Fig. S15 in Appendix S2).

Both the Akaike's information criterion (AIC; AIC was applied for Worldwide terrestrial mammals and West Indies herpetofauna and Akaike's information criterion corrected (AIC_c) for the other taxonomic groups; Burnham & Anderson, 2002) and the Bayesian information criterion (BIC) were applied, respectively, for each taxonomic group as criteria for model selection (Burnham & Anderson, 2002). For each model in each taxonomic group, we calculated the log-likelihood ($\log L$), which was used to determine AIC or AIC_c (hereafter referred to as $AIC_{(c)}$ for brevity), and BIC, respectively. We also calculated the difference in $AIC_{(c)}$ ($\Delta AIC_{(c)}$) and BIC (ΔBIC) for the model selection. The model with lower $AIC_{(c)}$ or BIC (hereafter referred to as IC for brevity) value indicates stronger evidence for being better over the others. Models with ICs less than two apart ($\Delta IC \leq 2$) are more or less equivalent (equally supported); those with ICs four to seven apart are clearly distinguishable; and models with ICs more than 10 apart are definitely different (Burnham & Anderson, 2002).

We evaluated the existence of an SIE within an island system by comparing the fit of a linear model (Equation 1) with the other Equations (2–15). When either of Equations 2–15 were supported as the best model ($\Delta IC \leq 2$), an SIE was considered to be present (Lomolino & Weiser, 2001). Equations 5, 6, 7, 11, and 12 being the best model is considered to provide stronger support for the SIE than Equations 2, 3, 4, 13, 14, and 15 (Morrison, 2014; Wang et al., 2016). We evaluated the existence of a steep-shallow area threshold by comparing the fit of a linear model (Equation 1) with Equations 8, 9, 10, and 14. If Equations 8, 9, 10, or 14 are supported ($\Delta IC \leq 2$), the evidence of a steep-shallow area threshold is found (Matthews et al., 2014). We evaluated the threshold number of a dataset by comparing all the equations. Equation 1 has zero threshold, Equations 2–10 one threshold, and Equations 11–15 two thresholds. We depicted the threshold number of a dataset as the average number of thresholds represented by all corresponding models that had equal support ($\Delta IC \leq 2$). For example, if Equation 1 is supported, the threshold number is calculated as $0/1 = 0$; if Equations 1, 2, and 11 are supported, the threshold number is calculated as $(0 + 1 + 2)/3 = 1$. We calculated area range as $\log A_{(max)} - \log A_{(min)}$, which reflected how many orders of magnitude that island size varies across a studied system. We applied the log-transformed version of species range, calculated as $\log (S_{(max)} - S_{(min)})$, to avoid outlier

effects and improve linearity. Last, we ran linear regression to show the relationship between threshold number and area range as well as the relationship between threshold number and species range, using $AIC_{(c)}$ and BIC for model selection, respectively. We performed all analyses using R 3.1.1 (R Development Core Team, 2014).

3 | RESULTS

Among the 68 global island datasets, SIE thresholds were detected in 58 cases (85%) and 46 cases (68%) using $AIC_{(c)}$ and BIC, respectively, whereas steep-shallow thresholds were detected in 31 cases (46%) and 19 cases (28%) using $AIC_{(c)}$ and BIC, respectively (Table 1). Among the 58 cases in which SIE thresholds were detected, two-segment piecewise models were supported in 35 cases (60%) and three-segment piecewise models were supported in 42 cases (72%) under $AIC_{(c)}$ criterion, whereas under BIC criterion, two-segment piecewise models were supported in 36 cases (78%) and three-segment piecewise models were supported in 16 cases (35%) among the 46 cases in which SIE thresholds were detected (Table 1). Among the 31 cases in which steep-shallow thresholds were detected, two-segment piecewise models were supported in 11 cases (35%) and three-segment piecewise models were supported in 26 cases (84%) under $AIC_{(c)}$ criterion, whereas under BIC criterion, two-segment piecewise models were supported in 13 cases (68%) and three-segment piecewise models were supported in six cases (32%) among the 19 cases in which steep-shallow thresholds were detected (Table 1).

The number of area thresholds in SARs varied among groups (Figure 2; Table S2 in Appendix S1). Model selection based on BIC identified Equation 1 as the most parsimonious model for Solomon Islands birds (sea-level species; Figure 2a); Equations 8 and 9 as equally the most parsimonious models for Austria semi-natural grassland Orthoptera species (Figure 2b). Model selection based on $AIC_{(c)}$ identified Equations 2, 3, 5, and 6 as the most parsimonious model for Australian Islands mammals (Figure 2c); and Equation 13 as the most parsimonious model for Aegean Sea isopods (Figure 2d) and West Indies herpetofauna (Figure 2f). Moreover, different groups showed different forms of the SAR with increasing area: steep-shallow, shallow-steep, shallow-steep-shallow, and

shallow–steep–steeper relationships for Austria semi-natural grassland Orthoptera species (Figure 2b), Australian Islands mammals (Figure 2c), Aegean Sea isopods (Figure 2d), and West Indies herpetofauna (Figure 2f), respectively.

We compared the first breakpoint value (T_1) between the best performing three-segment method and the best performing two-segment method for Aegean Sea isopods and West Indies herpetofauna, and found Equations 2, 3, and 4 were equally the most parsimonious models (mutual $\Delta AIC_{(c)} \leq 2$) for the two-segment method; moreover, T_1 estimated by Equations 2, 3, and 4 were larger than by Equation 13 for Aegean Sea isopods (0.227–0.914 vs. –0.301; Figure 2d; Table S2 in Appendix S1) but smaller for West

Indies herpetofauna (–0.603–0.111 vs. –0.022; Figure 2f; Table S2 in Appendix S1).

We examined whether the traditional method of detecting the SIE is still robust here for Australian Islands mammals, Aegean Sea isopods, and West Indies herpetofauna using $AIC_{(c)}$ as the criterion for model selection, which is widely used in other SIE studies (e.g., Morrison, 2014; Wang et al., 2016). For Aegean Sea isopods and West Indies herpetofauna, any choice of Equations 2, 3, 4 (one threshold without a left-horizontal segment) were better than Equations 5, 6, 7 (one threshold with a left-horizontal segment) but worse than Equation 11 (two thresholds with a left-horizontal segment); however, Equation 11 was worse than Equation 13 (two

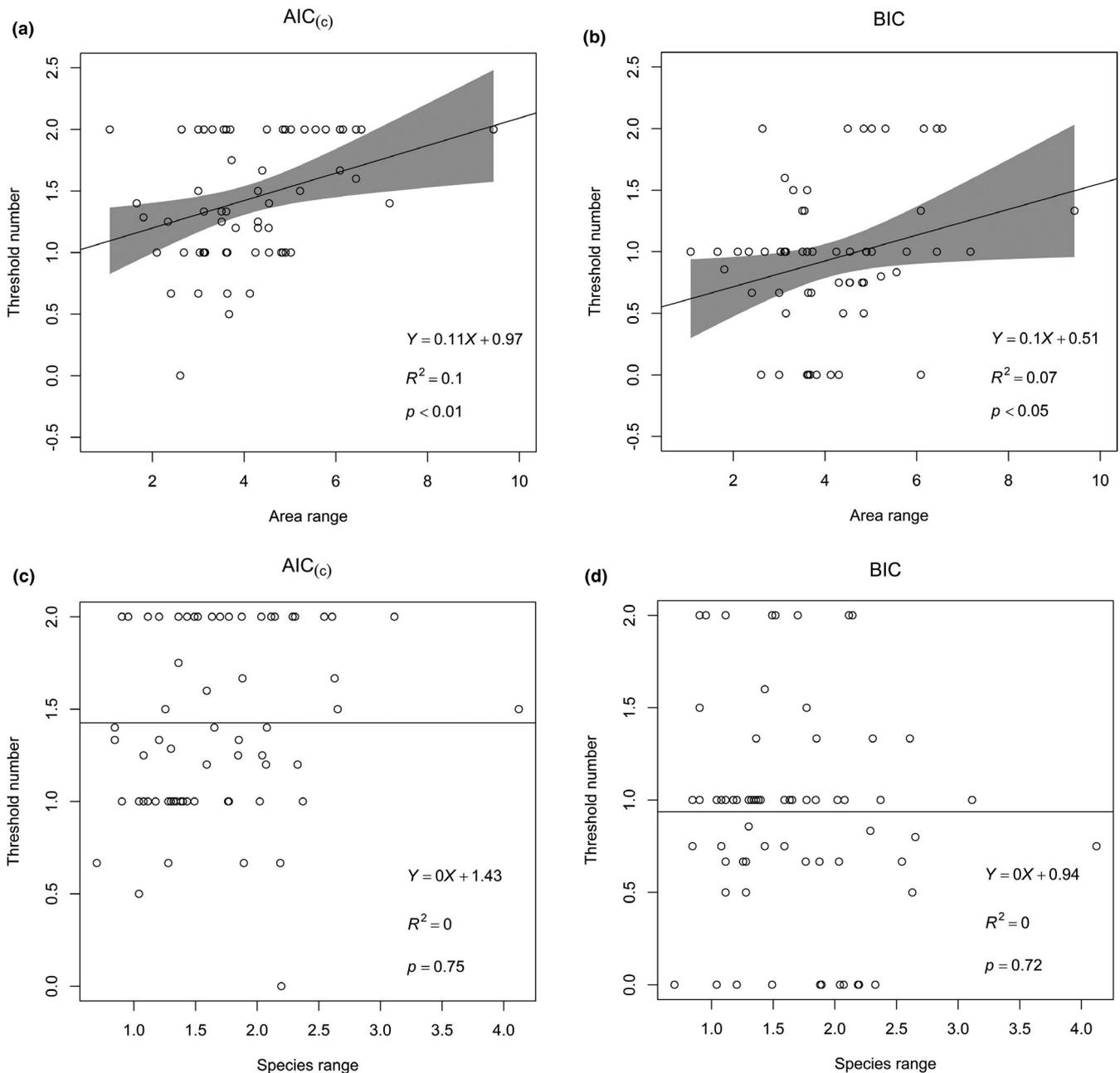


FIGURE 3 Linear regression, showing the relationship between threshold number and \log_{10} -transformed area range (a and b) and species range (c and d) of a studied system, using $AIC_{(c)}$ and BIC for model selection, respectively

thresholds without a left-horizontal segment; Table S2 in Appendix S1). For Australian Islands mammals, if we compare Equation 4 with Equation 5 or 6, no slope for the SIE will be the conclusion; however, if we compare Equation 2 or 3 with Equation 5, 6, or 7, it is ambiguous to say whether there is a slope (Table S2 in Appendix S1). Thereby, the traditional method of detecting the SIE is not robust, and different choices of model comparison can lead to totally different conclusions.

From the result of linear regressions conducted by using 68 island datasets, a significant positive relationship was found between threshold number and area range under both AIC_c ($p < 0.01$) criterion and BIC criterion ($p < 0.05$; Figure 3). This result suggests more species–area patterns could be detected as the area distributions of sampled islands get broader. However, no obvious relationship was found between threshold number and species range (Figure 3).

4 | DISCUSSION

Our results imply both SIE thresholds (58 cases, 85% and 46 cases, 68% under AIC_c and BIC criteria, respectively) and steep–shallow thresholds (31 cases, 46% and 19 cases, 28% under AIC_c and BIC criteria, respectively) are common in nature. Under AIC_c criterion, three-segment piecewise models are more prevalent (42 cases, 72% and 26 cases, 84% in detecting SIE and steep–shallow thresholds, respectively), whereas under BIC criterion, two-segment piecewise models are more prevalent (36 cases, 78% and 13 cases, 68% in detecting SIE and steep–shallow thresholds, respectively). Two-segment piecewise models have been widely used in previous studies (e.g., Lomolino & Weiser, 2001; Dengler, 2010; Wang et al., 2016); however, three-segment piecewise models, despite their prevalent existence, have long been neglected.

The shallow–steep–shallow relationship for Aegean Sea isopods (Figure 2d) is inconsistent with the result of Dengler (2010) who detected only two segments; however, the three-segment models have lower AIC_c values than those of two-segment models. The first two SAR patterns (shallow–steep) are similar to those of Australian Islands mammals, and two possible reasons may explain the last pattern (shallow). First, larger islands normally possess larger human population and richer natural resources, so that a more advanced economic development could be expected (Gao & Perry, 2016b; Helmus, Mahler, & Losos, 2014). Thus, human-mediated dispersal among those large islands could be facilitated and thereby lowering the slope of SAR. Second, species richness of isopods might be seriously underestimated for the largest island, because if the largest island is removed, a shallow–steep–steeper relationship is obtained (Figure 2e). The shallow–steep–steeper relationships of Aegean Sea isopods (the largest island removed; Figure 2e) and West Indies herpetofauna are in accordance with the proposal that put forward by Lomolino and Weiser (2001) and Rosenzweig (2004), which states three biological scales of species–area curve with three corresponding dominant processes of species addition and slope ranges.

Besides, other factors such as regional biogeographical setting (Gerstner, Dormann, Václavík, Kreft, & Seppelt, 2014), historical events (Hewitt, 2000; Kisel, McInnes, Toomey, & Orme, 2011), climate change (Lewis, 2006), human activities (Corlett, 2015; Gao & Perry, 2016b), and fire and other ecological disturbances (Blondel & Vigne, 1993) may also be responsible for the variation among organisms in their response to a particular physical setting, sensitivity to barriers to dispersal, and habitat extent requirement for proliferation and diversification (Blondel & Vigne, 1993; Reyjol et al., 2007). All these factors, acting either additively or synergistically, may influence the proposed species–area patterns across portion or the whole scale. Thereby, the larger size range with a larger number of sampled islands, the more detailed factors shaping the patterns could be depicted, and the more kind of SARs might be detected. Thus, the number of thresholds is positively correlated with the area range of a studied system, which is consistent with Matthews, Triantis, et al. (2016) and Wang et al. (2016). In contrast, species range received considerably less support in determining the number of thresholds. The main reason for the weak relationships is probably that species range is less effective in determining the limits of the SIE and evolutionary processes considering different taxonomic group, island age, and human influence among the studied systems and is less likely to correlate to area range.

The currently widely used method for detecting the SIE is not robust. First, although a two-segment left-horizontal function performs better than a two-slope function, it may be worse than a three-slope function; and/or although a two-slope function performs better than a two-segment left-horizontal function, it may be worse than a three-segment left-horizontal function (e.g., Aegean Sea isopods; West Indies herpetofauna). Second, even if under the two-segment circumstances, different choices of comparison between two models with and without a left-horizontal part may lead to totally different conclusions (e.g., Australian Islands mammals). It implies that the currently available but incomprehensive piecewise models may provide a biased result in detecting the SIE. Therefore, we suggest previous SIE detection works conducted using an incomprehensive set of piecewise models, while ignoring the possibility of three segments need to be reanalyzed. It is very important to do so because the threshold value, where the slope changes, may be important for a successful application of island theory to conservation biogeography (Triantis & Bhagwat, 2011). However, the incomprehensive set of piecewise models may provide a misleading result on the existence of an SIE; moreover, the traditional two-segment method may cause poor estimations for both slope and threshold value of an SIE. In order to improve the method for detecting the SIE, we suggest that it is essential to determine the number of area thresholds and apply a comprehensive set of piecewise models rather than just as few as two. We agree with the suggestions that a large sample size and small or even empty islands should be included in analyzing the SAR of a system (Matthews et al., 2014; Wang et al., 2016). For instance, if some more large islands (e.g., Australia) were included into Australian Islands mammals, a third species–area pattern with a higher slope might appear; and if some more small

empty islands were included into Austrian semi-natural grassland Orthoptera species, an SIE might appear. Besides, more precise threshold values could be obtained if a complete sampling effort was taken and a large sample size was included.

Although the three-segment piecewise regressions could display two area thresholds proposed by Lomolino and Weiser (2001) and Rosenzweig (2004) and may provide better estimations for slope and threshold value within an SIE, they incorporate more parameters into the models, which in turn will decrease the statistical power for the robustness as compared with the two-segment regressions using the same data size. Therefore, a larger number of surveyed islands in the dataset are required for the applicability of three-segment approach. To ensure an adequate island number incorporated in the dataset, we suggest, first, empty islands from a smaller spatial scale are welcomed to be collected because the effect of empty islands in generating SIEs is prevalent (Wang et al., 2016). Second, an incomplete sampling effort is usually taken, leaving many other islands in the studied system uninvestigated (e.g., Davies & Smith, 1998; Panitsa, Tzanoudakis, Triantis, & Sfenthourakis, 2006; Wang, Bao, Yu, Xu, & Ding, 2010). However, the applicability of complicated models declines from relatively complete island lists to extremely incomplete ones, so we suggest island collection is ought to be complete at the current studied spatial scale to allow more complex models tested. Third, sometimes a surveyed location is involved in a larger-scale system, for instance, Bahamas, where Morrison (2014) studied the SIE of plants, is actually a portion of the whole West Indies region. The regional SAR patterns could give insight into the causes of local patterns (Schrader, Moeljono, Keppel, & Kreft, 2019), so islands collected from a larger spatial scale are also suggested.

Determining the number of area thresholds in a system is critical to estimating the precise species–area patterns. Thus, we suggest previous piecewise regression and SIE detection works need to be reanalyzed. Two thresholds rather than one shine new light on conservation biogeography: First, it offers opportunity to assess variables such as habitat diversity, productivity, island age, energy, and environmental heterogeneity (Anderson & Wait, 2001; Tjørve & Tjørve, 2011; Whitehead & Jones, 1969) within the limits of the first threshold value; second, speciation may be the dominant process adding to the species richness of assemblages beyond the limits of the second threshold value (Losos & Schluter, 2000); third, habitat diversity, carrying capacity, and extinction/immigration dynamics as envisioned by MacArthur and Wilson (1967) may govern species richness between the first and second threshold values; fourth, it appreciates other factors, such as historical events (Hewitt, 2000; Kisel et al., 2011), climate change (Lewis, 2006), human activities (Corlett, 2015; Gao & Perry, 2016b), and so on, to exert their influence across portion or the whole scale; and last, the rate of biodiversity loss from habitat loss varies among three area intervals, and thus, two thresholds can effectively facilitate the establishment of protection priority level among islands and maximize the diversity saved in a system with limited funds for conservation. We conclude that conservation biologists and applied ecologists should determine the number of

area thresholds when estimating the precise species–area patterns and making management strategies in fragmented landscapes.

Despite the widespread use of the AIC, some believe that it is too liberal and tends to select overly complex models (Kass & Raftery, 1995). As compared to the AIC, the BIC is more conservative, insisting on a greater improvement in fit before it will accept a more complex model (Link & Barker, 2006). That is the reason why in our results three-segment piecewise models are more prevalent under AIC_(c) criterion in finding SIE and steep–shallow thresholds, as well as the lower slope and less significant *p*-value of the relationship between threshold number and area range under BIC criterion. However, the merits and debits of AIC and BIC remain controversial, and a formal comparison in terms of performance between AIC and BIC is very difficult (cf. Burnham & Anderson, 2002, pp. 293–305, for a pro-AIC account, and Kass & Raftery, 1995, for a pro-BIC account), so we used both criteria rather than one. To date, although there are a number of species–area pattern studies, the AIC was dominantly applied. Our study is among the first attempt to apply both criteria for model selection in this field. Besides, previous studies concerning SARs and SIEs comprised a large portion of island-poor (<20 islands) archipelagoes in the analyses (e.g., Matthews, Guilhaumon, Triantis, Borregaard, & Whittaker, 2016; Triantis et al., 2012; Wang et al., 2016). Instead, to meet the complexity of the piecewise regression models, we selected island datasets with > 40 islands in each case. Our study on piecewise models and species–area patterns thus provide reliable results, contributing to the recent heated disputes on the SIE theory and expanding our horizons on SAR thresholds at different spatial scales.

ACKNOWLEDGMENTS

We kindly thank Yanping Wang, Kostas Triantis, Zhonglin Shang, and two anonymous referees for their constructive comments, which greatly improved this manuscript. This work was supported by the Doctoral Fund of Hebei Normal University (No. L2018B21), the Advanced Talents Foundation of Hebei Education Department (No. GCC2014037) and the Natural Science Foundation of Hebei Province (No. D2019205019).

CONFLICT OF INTEREST

The authors have no conflict of interest related to this work.

AUTHOR CONTRIBUTIONS

DG and ZC conceived the idea and collected island datasets; DG, ZC, and PX performed data analyses; DG and GP wrote the first version of the manuscript; and all authors contributed and approved the final version.

DATA AVAILABILITY

All datasets used in this study are sourced from the literature which can be found in Supporting information.

ORCID

De Gao  <https://orcid.org/0000-0002-7055-5285>

REFERENCES

- Anderson, W. B., & Wait, D. A. (2001). Subsidized island biogeography: Another new twist on an old theory. *Ecology Letters*, 4, 289–291.
- Blondel, J., & Vigne, J. D. (1993). Space, time and man as determinants of diversity of birds and mammals in the Mediterranean Region. In R. E. Ricklefs, & D. Schluter (Eds.), *Species diversity in ecological communities: Historical and geographical perspectives* (pp. 135–146). Chicago, IL: University of Chicago Press.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed. New York, NY: Springer.
- Corlett, R. T. (2015). The Anthropocene concept in ecology and conservation. *Trends in Ecology and Evolution*, 30, 36–41. <https://doi.org/10.1016/j.tree.2014.10.007>
- Crowe, T. M. (1979). Lots of weeds: Insular phytogeography of vacant urban lots. *Journal of Biogeography*, 6, 169–181. <https://doi.org/10.2307/3038050>
- Davies, N., & Smith, D. S. (1998). Munroe revisited: A survey of West Indian butterfly faunas and their species–area relationship. *Global Ecology and Biogeography*, 7, 285–294. <https://doi.org/10.1046/j.1466-822X.1998.00288.x>
- Dengler, J. (2010). Robust methods for detecting a small island effect. *Diversity and Distributions*, 16, 256–266. <https://doi.org/10.1111/j.1472-4642.2010.00646.x>
- Drinnan, I. N. (2005). The search for fragmentation thresholds in a southern Sydney suburb. *Biological Conservation*, 124, 339–349. <https://doi.org/10.1016/j.biocon.2005.01.040>
- Fahrig, L. (2001). How much habitat is enough? *Biological Conservation*, 100, 65–74. [https://doi.org/10.1016/S0006-3207\(00\)00208-1](https://doi.org/10.1016/S0006-3207(00)00208-1)
- Fattorini, S. (2007). To fit or not to fit? A poorly fitting procedure produces inconsistent results when the species–area relationship is used to locate hotspots. *Biodiversity and Conservation*, 16, 2531–2538. <https://doi.org/10.1007/s10531-006-9029-6>
- Ficetola, G. F., & Denoël, M. (2009). Ecological thresholds: An assessment of methods to identify abrupt changes in species–habitat relationships. *Ecography*, 32, 1075–1084.
- Gao, D., & Perry, G. (2016a). Detecting the small island effect and nestedness of herpetofauna of the West Indies. *Ecology and Evolution*, 6, 5390–5403. <https://doi.org/10.1002/ece3.2289>
- Gao, D., & Perry, G. (2016b). Species–area relationships and additive partitioning of diversity of native and nonnative herpetofauna of the West Indies. *Ecology and Evolution*, 6, 7742–7762. <https://doi.org/10.1002/ece3.2511>
- Gentile, G., & Argano, R. (2005). Island biogeography of the Mediterranean Sea: The species–area relationship for terrestrial isopods. *Journal of Biogeography*, 32, 1715–1726. <https://doi.org/10.1111/j.1365-2699.2005.01329.x>
- Gerstner, K., Dormann, C. F., Václavík, T., Kreft, H., & Seppelt, R. (2014). Accounting for geographical variation in species–area relationships improves the prediction of plant species richness at the global scale. *Journal of Biogeography*, 41, 261–273. <https://doi.org/10.1111/jbi.12213>
- Harcourt, A. H., & Doherty, D. A. (2005). Species–area relationships of primates in tropical forest fragments: A global analysis. *Journal of Applied Ecology*, 42, 630–637. <https://doi.org/10.1111/j.1365-2664.2005.01037.x>
- Helmus, M. R., Mahler, D. L., & Losos, J. B. (2014). Island biogeography of the Anthropocene. *Nature*, 513, 543–546. <https://doi.org/10.1038/nature13739>
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913. <https://doi.org/10.1038/35016000>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kisel, Y., McInnes, L., Toomey, N. H., & Orme, C. D. L. (2011). How diversification rates and diversity limits combine to create large-scale species–area relationships. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2514–2525. <https://doi.org/10.1098/rstb.2011.0022>
- Ladle, R. J., & Whittaker, R. J. (2011). *Conservation biogeography*. Oxford, UK: Wiley-Blackwell.
- Laurance, W. F. (2008). Theory meets reality: How habitat fragmentation research has transcended island biogeographic theory. *Biological Conservation*, 141, 1731–1744. <https://doi.org/10.1016/j.biocon.2008.05.011>
- Lavin, M., Wojciechowski, M. F., Richman, A., Rotella, J., Sanderson, M. J., & Matos, A. B. (2001). Identifying tertiary radiations of Fabaceae in the Greater Antilles: Alternatives to cladistic vicariance analysis. *International Journal of Plant Sciences*, 162, S53–S76. <https://doi.org/10.1086/323474>
- Levenson, J. B. (1981). Woodlots as biogeographic islands in southern Wisconsin. In R. L. Burgess, & D. M. Sharpe (Eds.), *Forest island dynamics in man-dominated landscapes* (pp. 13–39). New York, NY: Springer-Verlag.
- Lewis, O. T. (2006). Climate change, species–area curves and the extinction crisis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361, 163–171. <https://doi.org/10.1098/rstb.2005.1712>
- Link, W. A., & Barker, R. J. (2006). Model weights and the foundations of multi-model inference. *Ecology*, 87, 2626–2635. [https://doi.org/10.1890/0012-9658\(2006\)87\[2626:MWATFO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2626:MWATFO]2.0.CO;2)
- Lomolino, M. V. (2000). Ecology's most general, yet protean pattern: The species–area relationship. *Journal of Biogeography*, 27, 17–26.
- Lomolino, M. V., & Weiser, M. D. (2001). Towards a more general species–area relationship: Diversity on all islands, great and small. *Journal of Biogeography*, 28, 431–445. <https://doi.org/10.1046/j.1365-2699.2001.00550.x>
- Losos, J. B., & Schluter, D. (2000). Analysis of an evolutionary species–area relationship. *Nature*, 408, 847–850. <https://doi.org/10.1038/35048558>
- MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, 17, 373–387. <https://doi.org/10.1111/j.1558-5646.1963.tb03295.x>
- MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton, NJ: Princeton University Press.
- Maron, M., Bowen, M., Fuller, R. A., Smith, G. C., Eyre, T. J., Mathieson, M., ... McAlpine, C. A. (2012). Spurious thresholds in the relationship between species richness and vegetation cover. *Global Ecology and Biogeography*, 21, 682–692. <https://doi.org/10.1111/j.1466-8238.2011.00706.x>
- Matthews, T. J., Guilhaumon, F., Triantis, K. A., Borregaard, M. K., & Whittaker, R. J. (2016). On the form of species–area relationships in habitat islands and true islands. *Global Ecology and Biogeography*, 25, 847–858.
- Matthews, T. J., Steinbauer, M. J., Tzirkalli, E., Triantis, K. A., & Whittaker, R. J. (2014). Thresholds and the species–area relationship: A synthetic analysis of habitat island datasets. *Journal of Biogeography*, 41, 1018–1028. <https://doi.org/10.1111/jbi.12286>
- Matthews, T. J., Triantis, K. A., Rigal, F., Borregaard, M. K., Guilhaumon, F., & Whittaker, R. J. (2016). Island species–area relationships and species accumulation curves are not equivalent: An analysis of habitat island datasets. *Global Ecology and Biogeography*, 25, 607–618. <https://doi.org/10.1111/geb.12439>

- Morrison, L. W. (2014). The small-island effect: Empty islands, temporal variability and the importance of species composition. *Journal of Biogeography*, 41, 1007–1017. <https://doi.org/10.1111/jbi.12264>
- Morrison, L. W., & Spiller, D. A. (2008). Patterns and processes in insular floras affected by hurricanes. *Journal of Biogeography*, 35, 1701–1710. <https://doi.org/10.1111/j.1365-2699.2008.01910.x>
- Neigel, J. E. (2003). Species–area relationships and marine conservation. *Ecological Applications*, 13, 138–145. [https://doi.org/10.1890/1051-0761\(2003\)013\[0138:SARAMC\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2003)013[0138:SARAMC]2.0.CO;2)
- Panitsa, M., Tzanoudakis, D., Triantis, K. A., & Sfenthourakis, S. (2006). Patterns of species richness on very small islands: The plants of the Aegean archipelago. *Journal of Biogeography*, 33, 1223–1234. <https://doi.org/10.1111/j.1365-2699.2006.01481.x>
- R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Reyjol, Y., Huguency, B., Pont, D., Bianco, P. G., Beier, U., Caiola, N., ... Virbickas, T. (2007). Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, 16, 65–75. <https://doi.org/10.1111/j.1466-8238.2006.00264.x>
- Rosenzweig, M. L. (2004). Applying species–area relationships to the conservation of diversity. In M. V. Lomolino, & L. R. Heaney (Eds.), *Frontiers of biogeography: New directions in the geography of Nature* (pp. 325–343). Sunderland, MA: Sinauer Associates.
- Schrader, J., Moeliono, S., Keppel, G., & Kreft, H. (2019). Plants on small islands revisited: The effects of spatial scale and habitat quality on the species–area relationship. *Ecography*, 42, 1–10. <https://doi.org/10.1111/ecog.04512>
- Sfenthourakis, S., & Triantis, K. A. (2009). Habitat diversity, ecological requirements of species and the Small Island Effect. *Diversity and Distributions*, 15, 131–140. <https://doi.org/10.1111/j.1472-4642.2008.00526.x>
- Suarez, A. V., Bolger, D. T., & Case, T. J. (1998). Effects of fragmentation and invasion on native ant communities in coastal southern California. *Ecology*, 79, 2041–2056. [https://doi.org/10.1890/0012-9658\(1998\)079\[2041:EOFAIO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[2041:EOFAIO]2.0.CO;2)
- Tjørve, E., & Tjørve, K. M. (2011). Subjecting the theory of the small-island effect to Ockham's razor. *Journal of Biogeography*, 38, 1836–1839. <https://doi.org/10.1111/j.1365-2699.2011.02534.x>
- Triantis, K. A., & Bhagwat, S. A. (2011). Applied island biogeography. In R. J. Ladle, & R. J. Whittaker (Eds.), *Conservation biogeography* (pp. 190–223). Oxford, UK: Wiley-Blackwell.
- Triantis, K. A., Guilhaumon, F., & Whittaker, R. J. (2012). The island species–area relationship: Biology and statistics. *Journal of Biogeography*, 39, 215–231. <https://doi.org/10.1111/j.1365-2699.2011.02652.x>
- Triantis, K. A., Vardinoyannis, K., Tsolaki, E. P., Botsaris, I., Lika, K., & Mylonas, M. (2006). Re-approaching the small island effect. *Journal of Biogeography*, 33, 914–923. <https://doi.org/10.1111/j.1365-2699.2006.01464.x>
- Wang, Y., Bao, Y., Yu, M., Xu, G., & Ding, P. (2010). Nestedness for different reasons: The distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, 16, 862–873.
- Wang, Y., Chen, C., & Millien, V. (2018). A global synthesis of the small-island effect in habitat islands. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20181868.
- Wang, Y., Millien, V., & Ding, P. (2016). On empty islands and the small-island effect. *Global Ecology and Biogeography*, 25, 1333–1345. <https://doi.org/10.1111/geb.12494>
- Whitehead, D. R., & Jones, C. E. (1969). Small islands and the equilibrium theory of insular biogeography. *Evolution*, 23, 171–179. <https://doi.org/10.1111/j.1558-5646.1969.tb03503.x>
- Williamson, M. H. (1981). *Island populations*. Oxford, UK: Oxford University Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gao D, Cao Z, Xu P, Perry G. On piecewise models and species–area patterns. *Ecol Evol*. 2019;9:8351–8361. <https://doi.org/10.1002/ece3.5417>