Environmental Skeptics and Critics, 2012, 1(2):25-29

Article

To be fractal or not to be fractal? Misuse of the species-area relationship

Alessandro Ferrarini
Department of Evolutionary and Functional Biology, University of Parma, Via G. Saragat 4, I-43100 Parma, Italy
E-mail: sgtpm@libero.it, alessandro.ferrarini@unipr.it

Received 1 May 2012; *Accepted 9 May 2012*; *Published online 5 June 2012*
IAEES

Abstract

Species-area relationship (SAR) is a tenet in ecology. Unfortunately, even in very recent works researchers take for granted what they shouldn't take for granted, that is to say the fractality of such relationship. This is a serious mistake, both from a theoretical and methodological viewpoint. Unvealing non-fractality or multi-fractality of species-area relationship could mean a lot for conservation purposes of rare and not rare species. Detection of discontinuities in SAR may indicate significant changes in the processes that generate and maintain biodiversity, in particular with regard to different degrees of density-dependent events or inter-specific interactions. Methodological tools are already at hand, so I invite scientists to employ them as soon as possible.

Keywords fractality; multi-fractality; non-fractality; piecewise regression; species-area relationship.

1 Introduction

Species-area relationship (SAR) is quite like a dogma in ecology, and is widely regarded as one of ecology's few laws (Dodds, 2009). SAR postulates that as sampling area increases so does the number of species recorded as well. Quantification of this relationship dates back to the 19th century (Watson, 1835, 1859), and now embraces thousands of studies about a wide variety of taxa and scales (e.g., Connor and McCoy, 1979; Rosenzweig, 1995; Lomolino and Weiser, 2001; Drakare et al., 2006).

The shape taken by SAR can be estimated by numerous functions. The most commonly used models are the power model (Arrhenius, 1920, 1921)

$$S= cA^{z} \qquad (1)$$

and the exponential one (Gleason, 1922)

$$S= c+z\log A \qquad (2)$$

with c and z acting as parameters. Further three-parameters SAR models are the sigmoid extended power

$$S= cA^{z-(d/A)} \qquad (3)$$

and the sigmoid Gompertz model

$$S=d\exp[-\exp(-z(A-c))] \quad (4)$$

The power model remains the most frequently used model, both for fitting curves to species–area data, and as a basis for explanatory theories of species diversity (MacArthur and Wilson, 1967; Hubbell, 2001).

Using recent papers on important journals as a scientific benchmark, in this paper I emphasize that SAR is always, and sometimes mistakenly, *a priori* supposed to be fractal, i.e. all equal along the interval of areal values, even when evidences suggest that at a certain point of the areal domain "something happens" that changes such relationship. In addition, I suggest which methodologies researchers should apply in order to test whether SAR is fractal or not. Above all, I suggest to *a priori* doubt about SAR fractality in order to derive *a posteriori* correct implications. Unvealing non-fractality of SAR could mean a lot for conservation purposes of rare and not rare species.

2 Misuse of SAR

Researchers take for granted that SAR is a fractal relationship; in other words, they grantedly accept that it's a scale-invariant relationship. As a consequence, when SAR is log-log transformed

$$\log S=\log c+z\log A \quad (5)$$

just a straight line should be enough to determine the expected log number of species once given the log areal coverage. Even very recent scientific papers use this *a priori* supposition.

Triantis et al. (2012) conducted an extensive quantitative analysis of the form taken by SAR among 20 models to determine the best-fit model. They concluded that over most scales of space, SAR is best represented by the power model, $\log S=\log c+z\log A$, and that more complex, sigmoid models may be applicable when the spatial range exceeds three orders of magnitude. They neither consider the chance to use piecewise SAR models in their analyses nor that, when the spatial range exceeds three orders of magnitude, the solution is not a change of the SAR model but just a piecewise version of the same power model. Why, in fact, sigmoid models should be fitter for wide areal domains, while power models for narrow intervals? Which ecological explanation for this outcome? Is not more reasonable that the same SAR model holds for the whole areal interval, despite with different values of the z parameter due to density-dependent events or different inter-specific interactions? Even from an interpretative viewpoint, shifting from power to sigmoid models means quite nothing, while detecting statistically-proven breakpoints in SAR means a lot.

Guilhaumon et al. (2010) presented the application of 8 pre-built SAR models (power, exponential, logistic etc.) to the well-known dataset by Preston (1962) describing plants of the Galapagos Islands. Again, they missed to consider a non-fractal or multi-fractal relationship, and the fitting of such models (figure 2 of their article) is very weak in the central areal domain between 100 and 500 squared miles. Their models just hold when area is very limited, but it's clear by inspecting their figure that a new species-area rule happens at about 100 sq. mi., something that would require SAR model to bend.

Eckstut et al. (2010) analysed SAR and relative contribution of biotic expansion and *in situ* diversification events using power-law and linear regression analyses. In figure 2 of their paper, they find that power-law SAR doesn't fit well for the most important and widest island (Hawai'i) of their dataset (central part of figure 2) so they decide to leave out Hawai'i from their SAR model (figure 2 on the right). Instead, it's clear that their log-log SAR model would need a multi-fractal version with a sloping upward linear model up to about 3 on the

X-axis, and then a decreasing linear model up to 4. The same happens for figure 3, where the authors do the same for *in situ* species and biotic expansion species. Again, Hawai'i island does not obey a simple power-law SAR, hence the authors leave it out. In the discusion section, the authors declare that Hawai'i distorts the SAR because it is both the largest and youngest Hawaiian island, indicating that it may not have yet reached species equilibrium. I simply suggest that Hawai'i island distorts a simple fractal SAR, but not a piecewise one.

3 Fixing the Problem

From a theoretical viewpoint, we scientists should never take anything for granted. I understand that SAR in its fractal form is concise, elegant and operative, but is it true-to-life? Sometimes yes, sometimes no, thus we should test the hypothesis of fractality before employing it as a dogma. I suggest that fractality in SAR is the exception, not the rule.

From a methodological viewpoint, when modeling SAR relationship, we should test multi-fractality through the use of a piecewise regression model in distinct intervals of the predictor variable domain. In other words, instead of fitting a single straight line $\log S = \log c + z \log A$ to the data, we should allow SAR to bend.

Methodological models to do this are already at hand. Krummel et al. (1987) first proposed a piecewise regression which restricts the regression to subsets of the data and performs regression over a window containing a fixed number of data points, with the window ranging over the entire set. Ferrarini (2011) shared a software that implements Krummel's approach.

Friedman (1991) proposed another regression algorithm, called MARS, that approximates the nonlinearity of a model by searching over all possible shifts in the relationship between X and Y. MARS is a non-parametric tool that fits piecewise linear regression approximating the non-linearity of a model through the use of separate slopes in distinct intervals of the predictor variable range (Friedman, 1991; Steinberg et al., 1999). The idea behind local non-parametric modelling is to allow for a potentially non-linear relationship over different ranges of the predictor variable by searching over all possible shifts (called knots) in the relation between a dependant variable and a predictor one. When modelling such relationship, a MARS model takes the form

$$Y = a_0 + \Sigma\, a_i\, F_i(X) + \varepsilon \qquad (6)$$

where $F_i(X)$ is called the *i-th* basis function of X. $F_i(X)$ is expressed by

$$F_i(X) = \max(0, X-k) \qquad (7)$$

where k is called knot, $F_i(X)$ is equal to 0 for all values of X up to k and $F_i(X)$ is equal to X-k for all values of X greater than k. $F_i(X)$ can be built for any value of k.

At the beginning a constant model is fitted, then MARS starts searching for a knot that improves the model the most. Then, a subset of the data points are chosen as knots, basis functions are built around such knots, and regression is performed to determine the coefficients of each basis functions. At each step the procedure finds which potential basis functions can be added to the model. Knots are found through a very intensive search procedure that proceeds until an user-specified maximum model size is reached. As basis functions are added to the model, the residual sum of squares gets smaller, but it does not penalize for overfitting. Therefore, after implementing the forward stepwise selection of basis functions, a backward procedure is applied in which the model is pruned by removing those basis functions that are associated with the smallest increase in the least squares goodness-of-fit, called generalized cross-validation (GCV; Craven and Wahba, 1979). GCV is a measure of the goodness of fit that takes into account not only the residual error but also the model complexity. It is given by:

$$GCV = \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\left(1 - \frac{C}{N}\right)^2} \qquad (8)$$

where N is the number of observations in the data set. The numerator measures the lack of fit on the model. The quantity C is the penalty for adding a basis function. The optimal MARS model is the one with the lowest GCV value. When no shift-points are detected, MARS behaves like a simple linear regression. MARS has been also used by Ferrarini et al. (2005) to detect proper breakpoints in the area-perimeter relationship (APR) of landscape patches, showing that also APR is usually multi-fractal with deep implications for the ecological meaning of habitat shape.

Grossi et al. (2001) suggested a different approach based on an iterative procedure that looks at every possible break of the X-Y relationship to select the best model. Despite very effective, this approach presents two drawbacks. When the number of observations is larger than 300, the procedure becomes prohibitive even for the most recent computers; in addition, this method requires a large sample size to perform the iterative procedure.

4 Conclusions

Species-area relationship is a tenet in ecology. SAR is appealing because it reflects a simple law across a range of spatial scales. However, SAR is only meaningful if the logarithmic relation between number of species and area is linear over the full range of the areal domain. If it is not, then SAR should be computed separately for the areal range over which it is constant.

Unfortunately, so far researchers take for granted what they shouldn’t take for granted, that is to say the fractality of such relationship. This is a serious mistake, both from a theoretical and methodological viewpoint. Omitting the discussion whether multi-fractal means non-fractal or not, I claim that unvealing non-fractality or multi-fractality of SAR might have deep implications for conservation purposes of rare and not rare species, because detection of discontinuities in SAR may indicate significant changes in the processes that generate and maintain biodiversity, in particular with regard to different degrees of density-dependent events or inter-specific interactions. Methodological tools are already at hand, so I invite scientists to use them as soon as possible.

References

Arrhenius O. 1920. Distribution of the species over the area. Meddelanden fran Vetenskapsakadmiens Nobelinstitut, 4: 1-6

Arrhenius O. 1921. Species and area. Journal of Ecology, 9: 95-99

Connor EF, McCoy ED. 1979. The statistics and biology of the species–area relationship. The American Naturalist, 113: 791-833

Craven P, Wabha G. 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerical Mathematics, 31: 317-403

Dodds WK. 2009. Laws, Theories and Patterns in Ecology. University of California Press, Berkeley, CA, USA

Drakare S, Lennon JJ, Hillebrand H. 2006. The imprint of geographical, evolutionary and ecological context on species–area relationships. Ecology Letters, 9: 215-227

Eckstut ME, McMahan CD, Crother BI, et al. 2010. PACT in practice: comparative historical biogeographic patterns and species–area relationships of the Greater Antillean and Hawaiian Island terrestrial biotas. Global Ecology and Biogeography (DOI: 10.1111/j.1466-8238.2010.00626.x

Ferrarini A, Rossi P, Rossi O. 2005. Ascribing ecological meaning to habitat shape through a piecewise regression approach to fractal domains. Landscape Ecology, 20: 799-809

Ferrarini A. 2011. Detecting ecological breakpoints: a new tool for piecewise regression. Computational Ecology and Software, 1(2): 121-124

Friedman JH. 1991. Multivariate adaptive regression splines. Annals of Statistics, 19: 1-141

Gleason HA. 1922. On the relation between species and area. Ecology, 3: 158-162

Grossi L, Zurlini G, Rossi O. 2001. Statistical detection of multiscale landscape patterns. Environmental and Ecological Statistics, 8: 253-267

Guilhaumon F, Mouillot D, Gimenez O. 2010. mmSAR: an R-package for multimodel species-area relationship inference. Ecography, 33: 420-424

Hubbell SP. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ, USA

Krummel JR, Gardner RH, Sugihara G, et al. 1987. Landscape patterns in a disturbed environment. Oikos, 48: 321-324

Lomolino MV, Weiser MD. 2001. Towards a more general species–area relationship: diversity on all islands, great and small. Journal of Biogeography, 28: 431-445

MacArthur RH, Wilson EO. 1967. The theory of island biogeography. Princeton University Press, NJ, USA

Preston FW. 1962. The canonical distribution of commonness and rarity: Part I. Ecology, 43: 185-215

Rosenzweig ML. 1995. Species Diversity in Space and Time. Cambridge University Press, New York, USA

Steinberg D, Bernstein B, Colla P, et al. 1999. MARS User Guide. Salford Systems, San Diego, CA, USA

Triantis KA, Guilhaumon F, Whittaker RJ. 2012. The island species–area relationship: biology and statistics. Journal of Biogeography 39: 215-231

Watson HC. 1835. Remarks on the Geographical Distribution of British Plants. Longman, Rees, Orme, Brown, Green, and Longman, London, UK

Watson HC. 1859. Cybele Britannica, or British Plants and Their Geographical Relations. Longman and Company, London, UK