Ecological Complexity 19 (2014) 84–95

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

Species–area relationship and a tentative interpretation of the function coefficients in an ecosystem simulation

Morteza Mashayekhi [a,*], Brian MacPherson [b], Robin Gras [c]

[a] School of Computer Science, University of Windsor, Canada N9B 3P4
[b] Department of Biological Science, University of Windsor, Canada N9B 3P4
[c] School of Computer Science, Department of Biological Science, Great Lakes Institute for Environmental Research, University of Windsor, Canada N9B 3P4

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 3 May 2014
Accepted 26 May 2014
Available online 24 June 2014

Keywords:
Coefficients interpretation
Individual-based modeling
Machine learning techniques
Nested and random sampling
SAR models
Beta diversity indices

ABSTRACT

In this paper, we identified the best species–area relationship (SAR) models from amongst 28 different models gathered from the literature, using an artificial predator–prey simulation (EcoSim), along with investigating how sampling approaches and sampling scales affect SARs. Further, we attempted to determine a plausible interpretation of SAR model coefficients for the best performing SAR models. This is the most extensive quantitatively based investigation of the species–area relationship so far undertaken in the literature.

We gathered 28 different models from the literature and fitted them to sampling data from EcoSim using non-linear regression and ΔAICc as the goodness-of-fit criterion. Afterwards, we proposed a machine-learning approach to find plausible relationships between the models' coefficients and the spatial information that likely affect SARs, as a basis for extracting rules that provide an interpretation of SAR coefficients.

We found the power function family to be a reasonable choice and in particular the Plotkin function based on ΔAICc ranking. The Plotkin function was consistently in the top three in terms of the best ranked SAR functions. Furthermore, the simple power function was the best-ranked model in nested sampling amongst models with two coefficients. We found that the Plotkin, quadratic power, Morgan–Mercer–Floid and the generalized cumulative Weibull functions are the best ranked models for small, intermediate, large, and very large scales, respectively, in nested sampling, while Plotkin (in small to intermediate scales) and Chapman–Richards (in large to very large scales) are the best ranked functions in random sampling. Finally, based on rule extractions using machine-learning techniques we were able to find interpretations of the coefficients for the simple and extended power functions. For instance, function coefficients corresponded to sampling scale size, patch number, fractal dimension, average patch size, and spatial complexity.

Our main conclusions are that SAR models are highly dependent on sampling scale and sampling approach and that the shape of the best ranked SAR model is convex without an asymptote for smaller scales (small, intermediate) and it is sigmoid for larger scales (large and very large). For some of the SAR model coefficients, there are clear correlations with spatial information, thereby providing an interpretation of these coefficients. In addition, the slope z measuring the rate of species increase for SAR models in the power function family was found to be directly proportional to beta diversity, which confirms the view that beta diversity and SAR models are to some extent both measures of species richness.

1. Introduction

The species–area relationship (SAR) is one of the most well-known and oldest patterns in ecological modeling (Lomolino, 2000; Tjørve, 2003; Dengler, 2009). SARs have a number of practical applications for managing natural communities. For example, SARs can be used for predicting the extinction rate of a

---

* Corresponding author. Tel.: +1 5192533000x3003.
E-mail addresses: mashaye@uwindsor.ca, mashayekhi.m@gmail.com (M. Mashayekhi), macphe4@uwindsor.ca (B. MacPherson), rgras@uwindsor.ca (R. Gras).

http://dx.doi.org/10.1016/j.ecocom.2014.05.011
1476-945X/

species based on habitat loss or reduction (Rompré et al., 2009; Rybicki and Hanski, 2013), for designing optimal reserve sizes (Desmet and Cowling, 2004), for identifying hotspots and geographical regions of high species richness (Fattorini, 2006), for assessing human impacts on biodiversity (Tittensor et al., 2007), for predicting the species richness of certain taxa based on richness of other species (Murakami and Hirao, 2010), and for estimating the species richness of larger regions (Plotkin et al., 2000).

The fundamental characteristic of SAR modeling is that species richness increases with the sampling area, with the increment rate decreasing for larger areas. Identifying the most biologically appropriate mathematical SAR model to characterize these behaviors has been one of the most important and controversial issues in biodiversity. Two of the earliest and most frequently applied mathematical models for the SAR, i.e., the power and logarithmic functions, were proposed by Arrhenius and Gleason in the 1920s (Lomolino, 2001; Tjørve, 2003). Subsequently, a number of researchers investigated how well these simple mathematical models fit the field data set obtained from different taxa (Lomolino, 2000; Plotkin et al., 2000; Ulrich and Buszko, 2003; Whittaker et al., 2007; Dengler and Boch, 2008; Fattorini, 2009). Others investigated a variety of practical applications of SAR models (Dolnik and Breuer, 2008; Murakami and Hirao, 2010; Azovsky, 2011).

Still other researchers considered not only the simple mathematical models, but also in addition tested new kinds of models based on more complex mathematical functions. Some of these new models are an extension of simple SAR models, while others are completely new functions for this domain. For example, several authors have argued that there is no universal model to describe all data sets and that the best model should be discovered for each data set separately (Flather and Mountain, 1996; He and Legendre, 1996; Tjørve, 2003, 2009). Others have proposed various models for different spatial scales (He and Legendre, 1996; Connor and McCoy, 2001; Lomolino, 2001). Keeley and Fotheringham (2003) have argued for a re-adoption of the traditional exponential model for certain kinds of plant data sets while retaining the power model for other kinds of data sets, depending on the structure of the plant community.

However, there is support in the literature for the overall adequacy of the power function family in representing species–area relationships. Plotkin et al. (2000) proposed a generalization of the power function, whereas Dengler (2009) suggested using the simple power function as a general model for all kind of species–area data on any scale. Ulrich and Buszko (2003), Drakare et al. (2006), Surendra and Singh (2009), and Azovsky (2011), along with Merwe and Rooyen (2011) all advocate the power function as providing an adequate account of species–area relationships with respect to selected data sets. Finally, Triantis et al. (2012) reported that the power model along with other simple models best represent the island species relationship (ISAR).

There are a number of plausible explanations in the literature regarding the apparent variation of SARs at different scales, for different types of species, and for various geographic locations. For example, Connor and McCoy (2001) argue that the relative abundance distribution of the species or the range of sampling in one area can affect SARs. They also believe that different taxa within various spatial scales could generate a different functional form of SARs. He and Legendre (2002), Martín and Goldenfeld (2006), and Tjørve et al. (2008) have shown that SARs are affected by species abundance and spatial distribution factors like species dominance and the level of aggregation. Sampling methods may also change the SAR model as discussed in Scheiner (2003) and Dengler (2009). Drakare et al. (2006) have observed that SARs are affected significantly by sampling schemes, spatial scales, and types of taxa or habitat. In other experiments, the effect of spatial distribution and aggregation information, spatial scale, evenness or measure of distribution of relative abundances of different species in a community, species abundance model, latitude, self-similarity, and sampling effort have been investigated (He and Legendre, 2002; Olszewski, 2004; Drakare et al., 2006; Tjørve et al., 2008; Tjørve and Tjørve, 2008; Jost, 2010; Merwe and Rooyen, 2011; Pereira et al., 2012).

Generally speaking, there are a number of shortcomings in the papers published on SAR models. First, several studies assume the power function as the default SAR model without considering other possible models (Drakare et al., 2006; Surendra and Singh, 2009; Cencini et al., 2012). Second, sampling methods and sampling scales have been neglected in many studies when researchers search for the best SAR models with several exceptions such as Turner and Tjørve (2005), Dolnik and Breuer (2008) and He and Hubbell (2011). Finally, articles in the literature that address the issue of how to interpret the SAR coefficients tend to assume the simple power function as the default function without considering their meaning with respect to alternative functions (Connor and McCoy, 1979; Gould, 1979; Martin, 1981; Cencini et al., 2012; Franzén et al., 2012).

To help resolve the debate regarding the best SAR function, we employed species richness data sets from computer simulations (described below) in order to address the following questions.

1. Is the power function the best-suited SAR model overall?
2. How do nested sampling and random sampling affect the shape of the SAR curves?
3. Do different sampling scales affect the SAR models?
4. Is there any correlation between SAR model coefficients and spatial information?

To address these questions, we employed an individual-based modeling simulation, EcoSim, to investigate the SAR. This method helps with the investigation of the species area relationship by considering the abundance and the distribution of species from a finer-grained level of description in terms of the behavior of individual organisms. The number of species in a given region is the outcome of the evolutionary processes of speciation, extinction, and migration to that region which in turn are caused by processes operating at the level of individuals (Lawson and Jensen, 2006). Thus, it is useful to study the dynamics of the SAR at the level of individual organisms which form the species.

To answer the first three questions, we collected 28 different functions through literature searches and examined them for various sampling scales and sampling methods. For the last question, using potentially informative spatial information, i.e., spatial factors, gathered from previous studies and applying machine learning techniques (Mitchell, 1997; Duda et al., 2000; Crisci et al., 2012), we attempted to find important factors that aid in the interpretation of the models' coefficients.

2. Materials and methods

2.1. EcoSim

The outcome of the interaction between individuals in a given ecosystem gives rise to ecological structures such as the species area relationship. We employed an individual-based modeling system (IBM), EcoSim, to investigate the SAR. EcoSim is a generic platform useful for investigating a number of topics in ecology including speciation (Golestani et al., 2012; Mashayekhi and Gras, 2012), species extinction (Mashayekhi et al., 2014) and the SAR. It is a multipurpose individual-based evolving predator–prey