Submission to: Methods in Ecology and Evolution

**Article Type: Applications**

**sars: an R package for fitting, comparing and evaluating species­­-area relationship [models]**

Thomas J. Matthews1, 2, Kostas Triantis, Robert J. Whittaker & Francois Guilhaumon

1GEES (School of Geography, Earth and Environmental Sciences), The University of Birmingham, Birmingham, B15 2TT

2CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências e Engenharia do Ambiente, PT-9700-042, Angra do Heroísmo, Açores, Portugal.

\*Correspondence: Thomas J. Matthews, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

Email: t.j.matthews@bham.ac.uk

Running header: sars R package

Word count: abstract: words; main text = words; Tables; Figures; references; appendices

**Keywords**

Diversity-area relationship, island biogeography, multi-model inference, power model, species accumulation curve, species-area relationship, random placement model

**Summary**

\*\*sar function in BAT package – but only fits 3 models (6 variants), and no mm averaging

\*add in random placement model and GDM – to introduction

\*previous version was specialist on mm, whereas this new version is general SAR package that provides mm inference, but also other SAR related functionality and is more general and programmed using standard S3 methods and thus people can easily add in their own models (e.g. new wave of countryside SAR models) and integrate these into the comparison and mm functions.

state that it is mmSAR but is for fitting multi-model diversity area relationships, a la Mazel.

**Introduction**

The species–area relationship (herein, ‘SAR’) describes the near universally observed pattern whereby the number of species increases with the area sampled, and it has been described as one of ecology’s few laws (Rosenzweig, 1995; Lawton, 1999). The SAR is a fundamental component of myriad ecological and biogeographical theories, such as the equilibrium theory of island biogeography (MacArthur & Wilson, 1967). In addition, SAR models have been widely used in applied ecology and conservation biogeography; for example, to predict the number of extinctions due to habitat loss (Whittaker et al., 2005). Numerous types of SAR have been described (see Scheiner, \*\*), and one primary dichotomy employed is the split of SAR into island SARs (ISARs) and species accumulation curves (SACs; Gray, Matthews et al.). ISARs characterise .. , whilst SACs describe … Whilst the remainder of the paper and the described R package are generally focused on ISARs, the models and the model fitting procedure can equally be applied to SACs (see Matthews et al.), although \*it should be noted that in SACs the data points are not independent.

Over twenty SAR models have been described in the literature (Tjorve, Triantis) \*\*and these models\*\* can be grouped into three \*\*SAR shapes\*: linear, convex and sigmoidal, \*\*and those with or without an asymptote (Triantis et al. 2012). However, despite this wide range of models, the majority of SAR studies are based exclusively on the power model (Arrhenius, other refs), which is given by the formula \*\*\*. In the past, due to \*\*, the log-log interpretation of the power model has generally been employed, as it can be fitted using standard linear regression; however, it has been found that the linear implementation does not produce identical parameter estimates to the non-linear implementation of the model (Triantis). Whilst the power model has been found to provide a good fit to a wide range of datasets (Triantis, Troussellier et al. 2017), it is not universally the best model, and a number of studies have reported other models, and perhaps more importantly, other model shapes, to provide better fits to empirical data than the convex power model (e.g. Benchimol & Peres, 2013; Matthews et al. 2015). For example, the scale dependency of the SAR has long been recognised, such that at smaller scales the form is predicted to be \*\*, whilst at larger spatial scales the form of the SAR is theorised to be sigmoidal (He & Legendre, \*\*). In sum, there is considerable model uncertainty with regards to the form of the SAR (Guilhaumon et al. 2008).

In response to the \*reported\* model uncertainty in SAR studies, a number of recent studies have employed a multi-model inference approach (Burnham & Anderson, 2002) in the analysis of SARs, whereby either (1) multiple SAR models are compared using various criteria (e.g. AIC) and a best model is chosen (e.g. Benchimol & Peres, 2013), or (2) multiple SAR models are fitted and a multi-model averaged curve is calculated using, for example, AIC weights (see Burnham & Anderson, 2002). For example, Guilhaumon et al. (2008) constructed multi-model averaged SAR curves using N models and used a bootstrap methodology to \*locate biodiversity hotspots\*. These N models were later incorporated into an R package (‘mmSAR’; Guilhaumon et al. 2010). However, as the mmSAR package only contains N models, it does not include a large number of models that have been found to provide the best fits to several empirical datasets (see Triantis et al. 2012; Matthews et al. 2015). In addition, the mmSAR package \*\*old mmSAR package no longer complies with recent programming good practices and is not on CRAN, is not user friendly (e.g. requires the user to load individual models prior to fitting), and since 2010 there have been many advances in R packages and functionality (see Wickham, 2015) that are lacking in mmSAR, such that there is the need for a new package to be designed from scratch. To address these concerns, we have developed the R package ‘mmSAR2’ (available on CRAN; version 1.\*), that provides functionality to fit twenty SAR model using non-linear regression, and to calculate multi-model averaged curves using various information criteria. The software also provides easy to use functionality to plot multi-model SAR curves and to generate confidence intervals using bootstrapping.

**Package description**

Following Guilhaumon et al. (2010), the sars package uses non-linear regression

*Initial parameters (custstart)*

*Theoretical and observed shape (latter is a novelty to this paper)*

*Class mmsar2 and nls*

*Additional functions*

In addition to the main functions used to fit and compare the twenty SAR models, the sars package provides additional functions for specific SAR based analyses.

Linear power (log version) and parameter comparisons

Coleman

*Examples*

**Conclusions**

Done in general way so people can add new models etc

**Authors’ contributions**

**Acknowledgements**

**Data accessibility**

**References**