



Generalized Logarithmic Species–Area Relationship Resolves the Arrhenius–Gleason Debate

Mark Carey¹ · John Boland¹ · Gunnar Keppel²

Received: 15 March 2022 / Accepted: 28 December 2022 / Published online: 22 February 2023
© The Author(s) 2023, corrected publication 2023

Abstract

The species–area relationship (SAR) is widely applied in ecology. Mathematically, it is usually expressed as either a semi-log or power-law relationship, with the former being introduced by Gleason and the latter by Arrhenius. We here resolve the dispute about which form of the SAR to prefer by introducing a novel model that smoothly transforms between the Gleason semi-log (GSL) and Arrhenius power law (APL) forms. The model introduced has the form of $\ln_q(S) = a + z \ln A$, with \ln_q being a generalized logarithmic function, which is a linear map ($y = x$) for $q = 0$ and a logarithmic map ($y = \ln x$) for $q = 1$ and q can take any intermediate value between 0 and 1. We applied this model to 100 datasets (mostly islands), linking species richness to island area. The APL was the preferred model in 68% of head-to-head comparisons with the GSL. Both models were supported in 40% of cases. In just under half (44%) of the cases, an intermediate model best explained the data. The results demonstrate the utility of a simple intermediate SAR model. Visualizing the profile of the range of model fits for all $q \in [0, 1]$ (a q chart) allows us to gain extra insight into SARs not yielded by head-to-head comparisons of GSL and APL. The mathematics related to the generalized logarithmic function introduced here should have applications to other areas of ecological modelling.

Keywords Biogeography · Generalized logarithm · Power law · Species–area relationships

1 Introduction

One of the earliest and persistently robust observations in the field of biogeography and ecology is that the diversity of species of a taxon increases in a predictable way with the area surveyed or the total area available (i.e., as on an oceanic island). This species–area relationship (SAR) has been described as the closest thing to a rule in ecology [1]. SARs have become fundamental to the understanding of patterns of biodiversity and a critical tool for predicting

biodiversity loss [2–4]. The form and parameterization of SARs are known to be significantly affected by sampling scheme, spatial scale, and the types of organisms or habitats involved [4–7].

First observed as a qualitative phenomenon by naturalists such as Forster [8], it was originally presented in mathematical form by Arrhenius [9] as a power law:

$$S = cA^z, \quad (1)$$

where S is species richness, A is area, and c and z are parameters which are determined empirically. This was almost immediately challenged by Gleason [10] who counter-proposed a SAR with a semi-log form (hereafter GSL):

$$S = a + b \ln A. \quad (2)$$

One of Gleason’s main critiques of the Arrhenius’ power law SAR (hereafter APL) was that it gave “impossibly high estimates” for large areas. In practice, this is not a such problem for oceanic islands, since island size is naturally bounded and usually relatively small (one of the largest oceanic islands is the large island of Hawaii which is approximately 10,457 km², but more typical are the islands of the Cape Verde archipelago which range

✉ Mark Carey
Mark.Carey@unisa.edu.au

John Boland
John.Boland@unisa.edu.au

Gunnar Keppel
Gunnar.Keppel@unisa.edu.au

¹ UniSA STEM, University of South Australia, Mawson Lakes Campus, GPO Box 2471, Adelaide, SA 5001, Australia

² UniSA STEM & Future Industries Institute, Mawson Lakes Campus, University of South Australia, GPO Box 2471, Adelaide, South Australia 5001, Australia

from around 2 to 1000 km²). This consideration has, however, led to later scale-dependant variants of the SAR, for instance, the persistence model of Plotkin et al. [11] for tropical forest plots (i.e., [11]) and the triphasic SAR model applicable across many orders of magnitude of scale (see [12]. At the scales we are interested in for this study, this consideration is largely irrelevant. In contrast, however, of particular relevance for us here is the observation that, for many field studies, the reported data often fall somewhere between the power law and semi-log models (i.e., [13, 14]). This suggests that a simple head to head comparison between the APL and GSL models may be inadequate.

Tjørve [15] proposed a hybrid model SAR to better fit datasets which did not cleanly fit an APL or GSL pattern. This model involved simply multiplying the two models together while introducing parameters to “slide” between mixed states of the GSL and APL, i.e.:

$$S = (c_1 + b \log A)^{\frac{dA}{A+n}} \times (c_2 A^z)^{1 - \frac{dA}{A+n}}, \tag{3}$$

where $c_1, c_2, b, d,$ and n are parameters.

This model has one immediate major problem, however, which is the number of free parameters it now contains. Tjørve resolves this problem by fixing four of the six parameters, $b, z, c_1,$ and $c_2,$ so that only two parameters, d and $n,$ remain to be fitted by regression. The justification for this scheme of parameter fixing relies on data not being noisy, which is a fair assumption for nested sample areas, such as those used for species accumulation curves. However, data from non-nested areas, such as islands, do not conform to this assumption. This is why Tjørve [15] states that this hybrid model is not intended for, nor should not be fitted to, island SARs. The model that we propose has no such limitation and is, therefore, applicable to any SAR, including the island SAR and species accumulation curves [16].

In this study, we present a novel method of constructing SARs, intermediate between the APL and GSL forms, without the limitations of the Tjørve model. Our approach utilizes the generalized logarithmic function, which has not previously been part of the mathematical ecologist’s toolkit and is likely to be useful in other contexts (see Section 4.2). The resulting model contains only one additional free parameter and is identical to the regular APL and GSL in the limit cases ($q = 1$ and $q = 0$). We test the new approach on 100 datasets gathered from the literature, including the original datasets of Arrhenius and Gleason.

2 Methods

2.1 The Generalized Logarithmic Transform

A logarithmic function can be conceptualized as a compression of space (or time) along some dimension. A generalized logarithmic function is where we have some

parameter (here q) which controls, continuously, the degree of compression along that dimension. Mathematically, such a function (in base e) can be defined as:

$$\ln_q(x) = \int_q^x \frac{dt}{t^q}. \tag{4}$$

Equation (4) is a modified form of the generalized logarithmic function as given in Tsallis [17] with the bottom limit changed from 1 to q . This makes the transformation exact rather than approximate (see also Martinez [18]). This then evaluates as

$$\ln_q x = \begin{cases} \frac{x^p - q^p}{p} & \text{if } q \neq 1 \\ \ln x & \text{if } q = 1 \end{cases}, \tag{5}$$

where $p = 1 - q$ and $q \in [0, 1]$ (derivation of this given in Appendix 2). This function smoothly transforms between a null transform ($\ln_0 x = x$) and a natural logarithmic transform ($\ln_1 x = \ln x$). Equation 5 is a modification of the Box Cox transform which is already approximately what we seek (the Box Cox transform goes from $y = \ln x$ to $y = x + 1$ instead of to $y = x$ so it just slightly “misses the mark”). This is fixed by the introduction of the q^p term. The relation of Eq. 4 to the Box Cox transformation is given in detail in Appendix 1 in the Supplementary Materials, along with a discussion of the additional mathematical properties of this function.

2.2 A New Hybrid SAR Model

The GSL model can be expressed as a log-transformed version of the APL, i.e.,

$$s = \ln(cA^z) = \ln c + z \ln A. \tag{6}$$

Therefore, a function such as that defined in Eq. 5 can smoothly transition from a linear map to a logarithmic map, producing a very simple hybrid SAR model, i.e.,

$$S = \ln_q(cA^z). \tag{7}$$

Although the form of Eq. 7 is the most convenient to conceptually understand the basis of our approach, a direct non-linear regression on data using this model immediately runs into some difficulties. If we expand Eq. 6 using Eq. 4, we get

$$S = \frac{(cA^z)^p - q^p}{p} = \frac{C^p A^{pz} - q^p}{p}. \tag{8}$$

If we fit this model in this form, then $p (= 1 - q)$ appears with both c and z as well as twice independently (counting q^p as a single instance). This has the potential to introduce statistical artefacts into the parameter estimates since the estimate

of the optimal q value will be weighted more heavily over the over the estimates of c and z .

If instead we fit

$$\ln_q S = z \ln A + d, \tag{9}$$

where $d = \ln c$ (i.e., similarly to how the APL is regularly fit as a linear regression in log-log space), then this cleanly separates the optimization of the \ln_q transformation of the response variable from the estimation of the parameters on the RHS (z and c). Henceforth, we refer to Eq. 9 as the SqA model, and when we refer to the q value associated with this model, we are referring to the q in Eq. 9. In addition, the iterative optimization of the RHS of Eq. 9 over a range of different q values gives us a graph of an ensemble of models, which we call a q chart, which furnishes us more information about the relationship between species and area than would a single, non-linear, model fit.

There are some algebraic subtleties involved in the relationship between the expressions of the overall model given in Eq. 7 versus Eq. 9. Specifically, if we designate the q of Eq. 7 as q^- and the q of Eq. 9 as q (or vice versa), then the relationship between them is given by

$$\ln_{q^-} x = \ln_q(\exp_q(\ln_q(x))), \tag{10}$$

where $\exp_q x$ is the inverse of the \ln_q function ($\exp_q x = (x^p + q^p)^{1/p}$). Thus, the model of Eq. 9 is properly expressed, in linear space, as

$$S = \exp_q(\ln(cA^z)). \tag{11}$$

It is cleaner, for reasons already discussed, to estimate the parameters of this model in the form given in Eq. 9. The only reason that we introduced this model in the form of Eq. 7 is that, so expressed, the motivation for such a model is most immediately obvious. These mathematical details and the derivations they rely upon are documented more fully in Appendix 1 in the Supplementary Materials. For our current purposes, it is enough to note that as long as the path in function space between the GSL and APL is monotonic, and a full set of intermediate models is represented, then we will have meaningful and interpretable results.

2.3 Data

We analysed 100 datasets for which at least one of either the APL or GSL was statistically significant. Data was collated from GIFT database and from datasets previously published in the literature [19]. A full list of the sources of the data used is provided in Appendix 3 in the Supplementary Materials. For Arrhenius' and Gleason's original datasets [9, 10], we present the analysis in detail, since these were the data that were the historical context for the original APL vs GSL

debate. Arrhenius used species counts for plant associations of different types lying in the islands of Stockholm, sampled areas increasing by square decimetres up to 100, except for weed association species where the maximum area was 300 dm². Gleason, by contrast, used species counts of a series of scattered and contiguous quadrats for aspen associations in North Michigan. In Arrhenius's data set, the *Pinus* woody species were aggregated, similar to Tjørve [15], and the weed association species considered separately.

2.4 Analysis

A model of the form of Eq. 7 was fitted iteratively by simple multiple regression while changing the value of q in the \ln_q transformation of the response variable over the interval of [0, 1] by increments of 0.01. Any models which are statistically significant for any q value are kept. We then compared the statistically significant models using the second-order bias correction to the Akaike information criterion (*AICc*) recommended for small samples [20].

Comparing *AICc* values for model fits with different transformations of the response variable (such as models defined by Eq. 9 which have a different q value) would mean comparing residuals in spaces where they are scaled differently, and so the errors are not directly comparable. There is, fortunately, a simple way to correct for this and recover useful *AICc* comparisons. This is by employing a Jacobian term as advised by Akaike [21]. A Jacobian transformation, the same one as used when changing variables in calculus, can be multiplied by the likelihood function when comparing models in different coordinate systems by likelihood to make such comparisons statistically meaningful. Since *AIC* is constructed using a log-likelihood, this turns out to be a simple additive correction:

$$AIC_C Q = AIC_C + 2q \left(\sum_i \ln s_i \right). \tag{12}$$

where s_i is the number of species in the i th sample or island and q is the q in Eq. 9. The novel part here is how the q term is incorporated, which, fortunately, turns out to be as simple as one could want: the usual Jacobian correction is multiplied by q . When using *AICc*, the model with the lowest value is considered to be the best, that is closest to the unknown "true" model. Models within an absolute difference ($\Delta AICc$) of one or two units are usually considered to be indistinguishable in statistical power [22, 23]. Thus, we consider a model to be equally well supported by the data as another model if it had $\Delta AIC < 2$ following Triantis et al. [23].

Regressions were run for Eq. 9 for 101 q values spanning the range $q \in [0, 1]$. This was done using *fitlm* in MATLAB R2019b (code used given in Appendix 2 in the Supplementary Materials). For each regression, we extracted the *AICc*

and the p -values of the slope and the intercept from model structure outputs and plotted the q values for models against the AICc values for those values of q which generated a statistically significant model (a “ q chart”). A model was deemed statistically significant if both the slope and the intercept were statistically significant at the < 0.05 level.

In Section 3, we give overall statistics for all model fits. The q charts are given there for the Arrhenius and Gleason datasets. Those for the remainder of the model fits are given in Appendix 4 in the Supplementary Materials.

3 Results

3.1 Arrhenius Stockholm Dataset

For the (aggregated) *Pinus* wood species in Arrhenius’ study, we have the optimal SqA model at $0.5 < q < 0.8$. All SqA models are statistically significant for all q . For the weed association species, the optimal SqA model was in the range $0.75 < q < 0.95$. There was a statistical significance cutoff at around $q = 0.4$; that is, no SqA models with $q < 0.4$ were statistically significant. If only the APL and GSL were compared, the APL would be preferred for both datasets (see Fig. 1).

3.2 Gleason North Michigan Dataset

The scattered quadrats clearly show a semi-log SAR ($q = 0$). For the contiguous quadrats, the power law is actually preferred on a head to head comparison, although the optimum model was $q = 0.79$ which is better supported ($\Delta AICc > 2$) statistically than the APL model (see Fig. 2).

3.3 Example Model Fit Graphs

Examples of the model fits against the data are shown in Fig. 3, where we see how the APL, the GSL, and the optimum SqA SAR fit the real data. In the case of Arrhenius’ *Pinus* woody species (Fig. 3A), the raw data more strongly supported the APL. For Gleason’s contiguous quadrats of aspens, the data more strongly supported an intermediate model, although the GSL would still have registered as statistically significant if it was the only model tested. In both cases, the optimal SqA model was a better fit for the data than either the APL or the GSL.

3.4 Statistics for Model Fits Over the 100 Datasets

The APL was statistically significant for 90% of the datasets, and the GSL was statistically significant for 50% of the datasets. On head to head comparisons (i.e., if the intermediate models are not considered), the APL is preferred in the majority (68%) of cases. The GSL was preferred in 26% of cases. In 6% of cases, the APL and GSL models were equally well supported by AICc ($\Delta AICc < 2$). Both models were at least statistically significant (if not preferred) in 40% of cases.

The ranges of q values for statistically equally well supported models are shown in Fig. 4. We see here that there are a majority of datasets where an intermediate model would be preferred but also that the APL would be preferred in the majority of cases in a head to head comparison with the GSL. There are a number of cases for which $q = 1$ for the optimal model, but hardly any with $q = 0$ for the optimal model. Overall, higher q values predominated (mean = 0.7, median = 0.76). The range of q values representing SqA models of

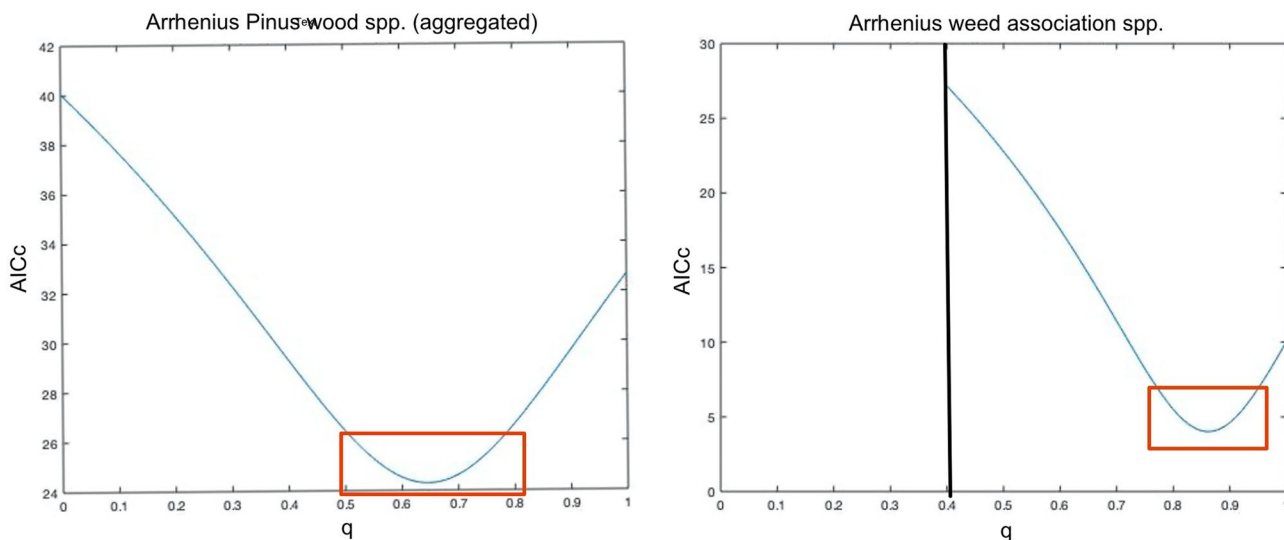


Fig. 1 SqA model fits for *Pinus* wood species and weed association species from [9] study of sites located in the islands of Stockholm. The red boxes indicate the range of SqA models which were within $\Delta AICc < 2$

of the optimal model. The black line on the right hand side graph of weed association species indicates the cutoff for statistically significant models, with no SqA models statistically significant for with $q < 0.4$

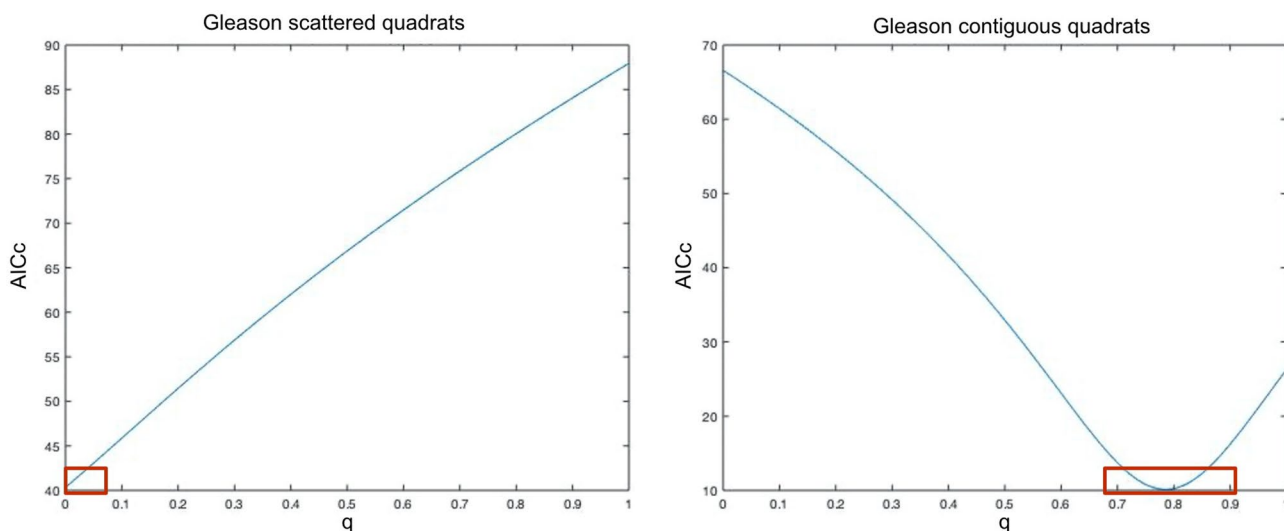


Fig. 2 SqA model fits for scattered and contiguous quadrats for aspen associations in North Michigan from [10] study. The red boxes indicate the range of models which were within $\Delta AICc < 2$ of the optimal model

$\Delta AICc < 2$ versus the optimum model was, however, rather large (mean = 0.4, median = 0.3) (Fig. 5). In 44% of cases, neither the APL or the GSL models were contained within the $\Delta AICc < 2$ interval around the optimal SqA model.

4 Discussion

Applying the generalized logarithmic function to analyse species-area relationships (SARs) revealed that models intermediate between the Arrhenius power law (APL) and the Gleason semi-log (GSL) models can often best describe

the data. If we take averages over the datasets, we do see broad support for the general practice of using the APL as the default SAR model. The mean of $q = 0.7$ together with a wide average range of $\Delta q \approx 0.4$ of models which were considered as good as the optimum model ($\Delta AICc < 2$) indicates that the APL should at least show no substantial lack of fit in most cases. This is consistent with the findings of Conner and McCoy [2] who found that of the 100 datasets, they examined the APL showed no substantial lack of fit for 75 of those datasets.

When comparing the APL and GSL head to head, two results suggest that an intermediate model might perform

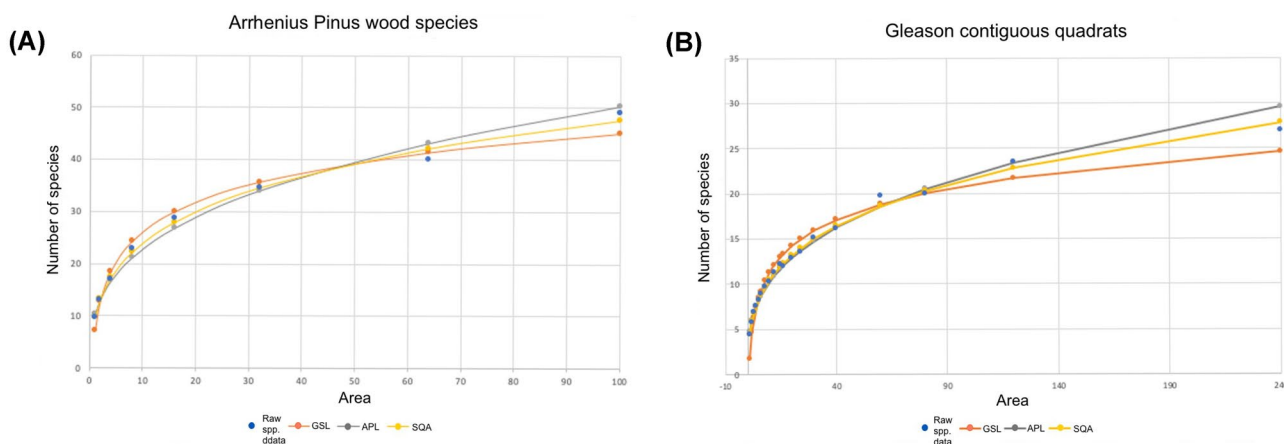


Fig. 3 Graphs illustrating the model fits for the Arrhenius power law (APL=grey), Gleason semi-log (GSL=orange), and optimal $\ln S = d + \ln A$ (SqA=yellow) SAR model fits against the raw data (blue dots) for A Arrhenius’s Pinus woody species data and B Gleason’s contiguous quadrats aspen association data. For the Arrhenius Pinus

woody species, in linear space, residual sum of squares (RSS) for the GSL = 31.411, APL = 18.627, and optimal SqA = 8.074. For the Gleason contiguous quadrats, RSS for GSL = 28.821, APL = 8.925, optimal SqA = 3.070

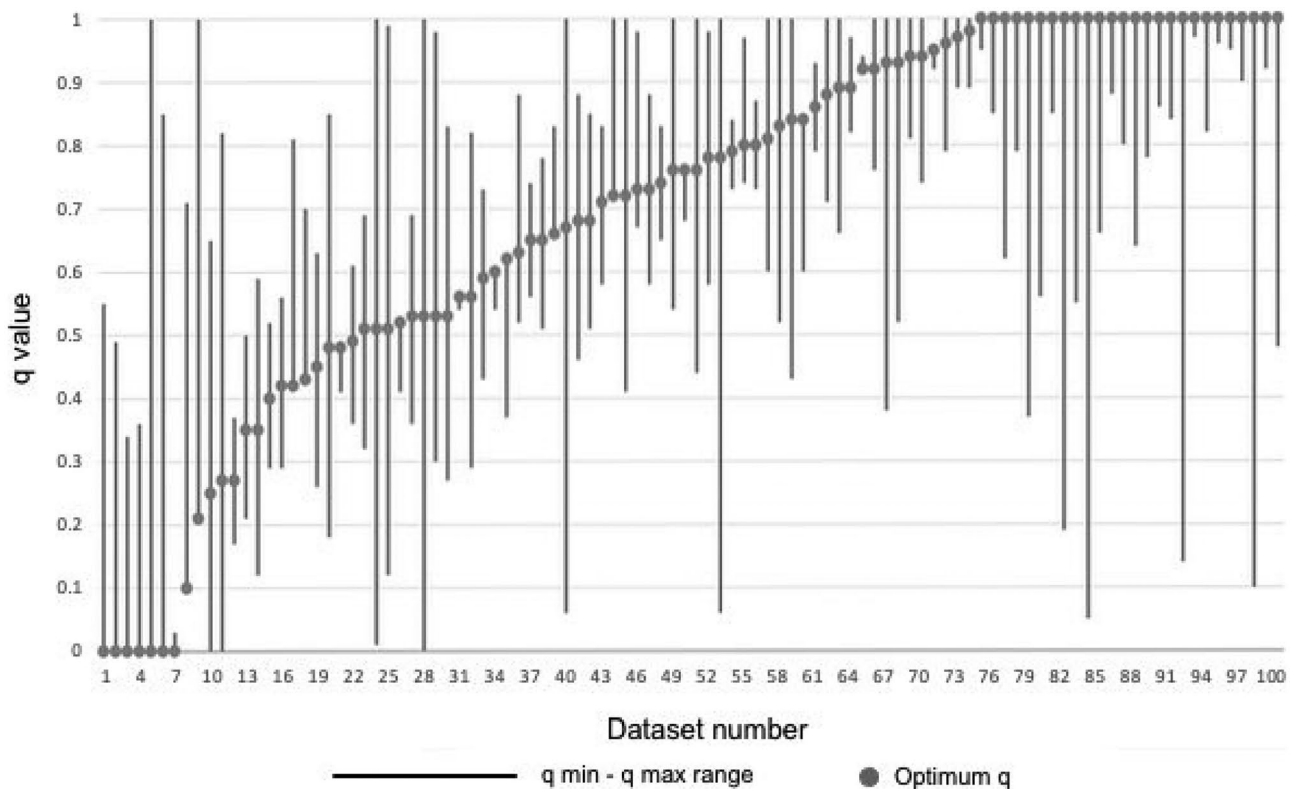


Fig. 4 Optimum q values and q ranges for models equally supported by AICc ($\Delta AICc < 2$ versus the optimal model) for models of the type $\ln q S = d + z \ln A$ for 100 datasets gathered from the literature. Datasets arranged in order of increasing optimum q value

better than either form of the SAR in some cases. Firstly, both models were statistically significant for 40% of the datasets, and these cases could therefore count as support for either model, if only the APL and the GSL were considered. Secondly, in only a few (6%) of the cases were the APL and GSL equally well supported by AICc. The first point is well illustrated by fits for the Aspen association data of Gleason. Although taken together the contiguous and scattered quadrats show statistical support for the GSL, the contiguous quadrats are actually better fit by the APL.

The results of the model fits for the 100 datasets fall into three main categories of interest:

1. where either the APL or GSL would be clearly preferred on a head-to-head comparison and one of them is the optimal model.
2. where both are equally well supported by AICc.
3. where an intermediate model is clearly preferred.

In just under half (44%) of cases, a hybrid model ($q \neq 1$, $q \neq 0$) was preferred and neither the GSL or the APL were equally well supported as the optimal model by AICc ($\Delta AICc < 2$). Thus, we find that we concur with Tjørve

[15] in that, for a large proportion of the datasets in the literature, the best fit is a model somewhere between the APL and GSL. Indeed, 3 of the 4 original datasets used to derive the APL and GSL models are better described by intermediate models. It is thus both important to consider the optimum SAR (and surrounding models) model prior to further analysis and interpretation. Investigating q charts, such as those shown in Figs. 1 and 2, is more informative, we believe, than fitting a single model.

4.1 A Note on Scattered Versus Contiguous Quadrats

Since the time of Arrhenius and Gleason, much research has been devoted to how sampling effort and sampling design can affect the observed shape of the species-area curve (see, for instance, [24–26] and references therein). What is particularly interesting when comparing Arrhenius' and Gleason's datasets, in the context of the original debate, is the use of both scattered and contiguous quadrats in Gleason's study. It is the scattered quadrats, rather than the contiguous quadrats, which furnish the clearest justification for the GSL in Gleason's dataset. In [27] reply to Gleason's original paper [27], he noted that Gleason's

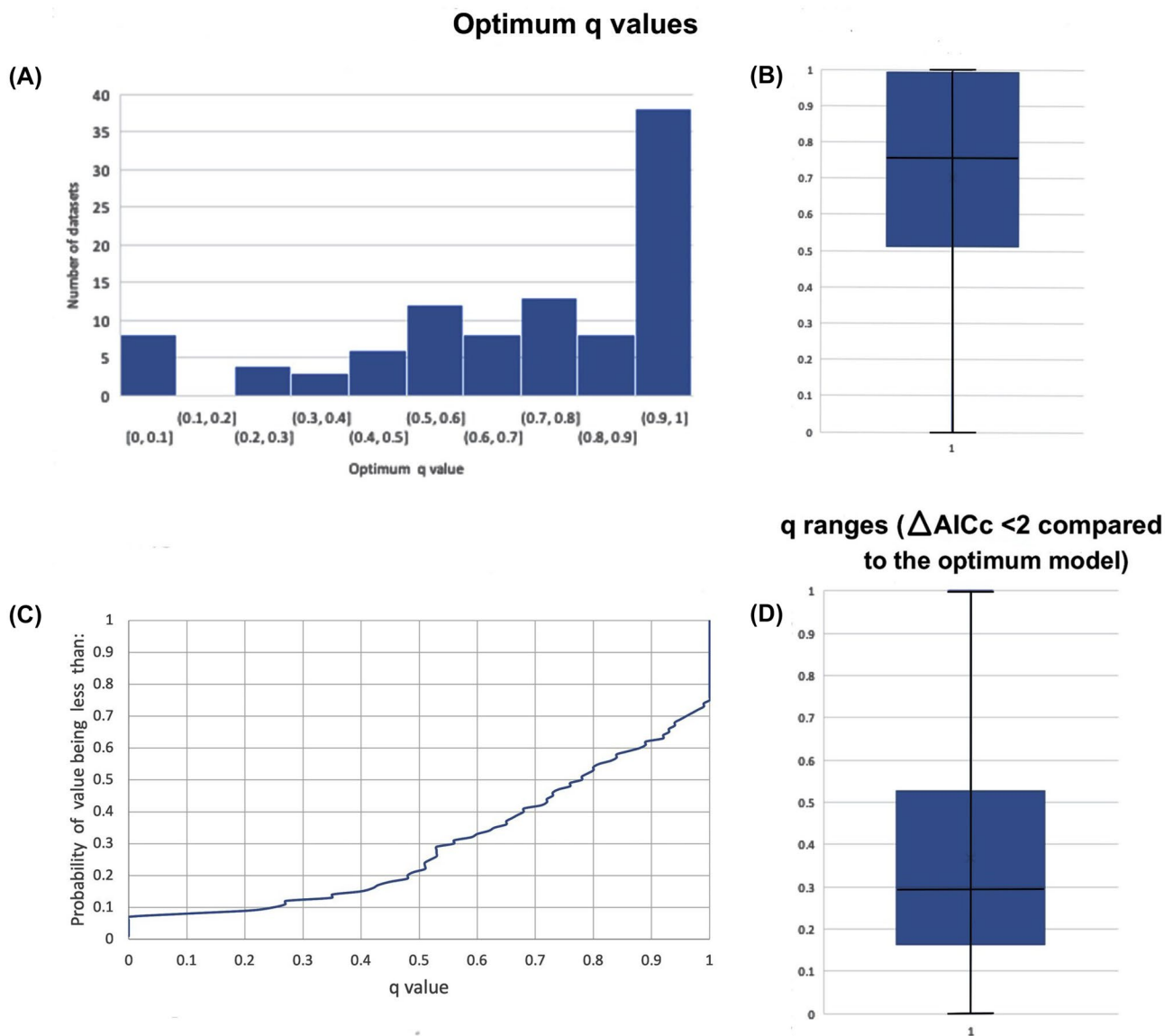


Fig. 5 Graphs of summary statistics for models of the type $\ln S = d + z \ln A$ for 100 datasets gathered from the literature. A and B show the distribution of q values for the optimal SqA models (lowest AICc), C shows the cumulative distribution function for these optimal q values, and D shows the distribution ranges of q values for SqA models which were statistically equivalent ($\Delta AICc < 2$ versus

the optimal model). For the optimal q values, we have mean=0.70, standard deviation=0.30, median=0.76, Q1=0.52, Q3=0.99, and interquartile range=0.47. The ranges of q values for “equally good” models ($\Delta AICc < 2$) around the optimum model had mean=0.37, standard deviation=0.27, median=0.30, Q1=0.17, Q3=0.52, and interquartile range=0.36

contiguous quadrats are well fitted by his own model, as we independently verified in this study. It is also relevant that Gleason [10] himself stated that he expected the accumulation of species with area to have a different character for scattered rather than contiguous quadrats. As nearby quadrats are, on average, more likely to be similar to each other than distant quadrats, adding species counts from distant quadrats sequentially should cause a much faster rate of

species accumulation with area than adding species counts from nearby quadrats [26]. These different accumulation rates will lead to different shapes of the species area curve as illustrated in Fig. 6. This is an interesting avenue for further study but our study was not designed to investigate this thoroughly, with most of our datasets being for islands. The ForestGEO network of large (4–120 ha) forest plots [28] would provide excellent datasets to investigate this further.

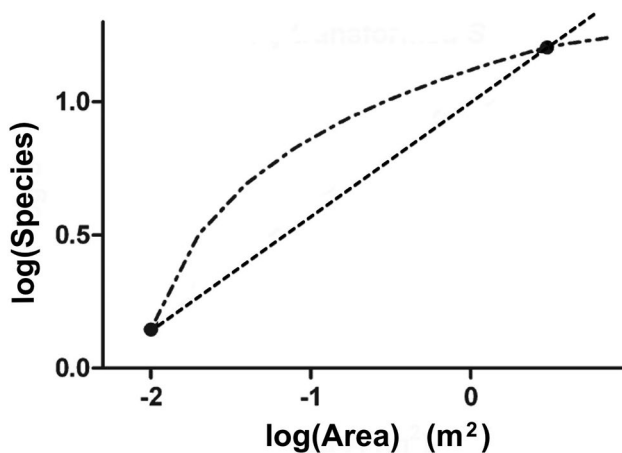


Fig. 6 A comparison of a power law (dashed line) and a semi-log SAR (dotted and dashed line) on a log–log plot. Note that the semi-log SAR initially accumulates more species than the power law and then later less. This is relevant to the interpretation of Gleason’s results using scattered versus contiguous quadrats (see main text). Figure adapted from Tjørve [15]

4.2 Other Applications of the General Logarithmic Function

The generalized logarithm has been applied to extending traditional approaches to statistical mechanics [29, 30], the theory of inter-temporal choices [31], and, even, to generalizing the fundamental formulas for calculus [32]. These applications, however, employ a version of the generalized logarithm which is defined as $\ln_q = (x^p - 1)/p$ rather than the version which we use here, defined in Eq. (5) $((x^p - q^p)/p)$. Because the -1 term does not vanish as the $-q^p$ term does as $q \rightarrow 0$, it causes problems when using this function as part of larger algebraic constructions, leading to more complex expressions than are strictly necessary and confusing the interpretation of the mathematical models used. In ecological modelling, the present authors have applied the generalized exponential function (see Section 2.2) to the unification of niche apportionment models [33]. This implementation employed a variably biased random variables defined by $X_q = (\exp q \lambda X)/(\exp q \lambda)$, where X is a uniformly distributed random variable and λ is a scale factor which sets the strength of the bias at $q = 1$. This kind of random variable is likely to be useful in other ecological modelling contexts.

5 Conclusion

The proposed approach of using generalized logarithmic functions provides deeper insights into SARs than the currently prevalent approaches of using the APL or head to head comparisons of the APL and GSL. It furnishes a deeper understanding of the relationship of the increase in species

richness with area. The capacity of generalized logarithmic and exponential functions to document ranges of modelling approaches that we have here demonstrated for SARs promises to have broader applications to other fields of ecological modelling.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10666-023-09873-6>.

Author Contribution All authors contributed to the study design based on an initial concept by MC. Material preparation, data collection, and analysis were performed by MC. The first draft of the manuscript was written by MC. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. Sources for the data are given in Appendix 3 of the Supplementary Materials.

Code Availability The MATLAB code used to fit the model is included in Appendix 2 with data for a working example included. The code as presented here has been lightly edited to more clearly illustrate the structure of the algorithm and so that it might be easier for the reader to translate it into other languages if desired.

Declarations

Ethics Approval Not applicable

Consent to Participate Not applicable

Consent for Publication Not applicable

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Lomolino, M. V. (2000). Ecology’s most general, yet protean pattern: The species-area relationship. *Journal of Biogeography*, 27(1), 17–26.
2. Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist*, 113(6), 791–833.

3. Triantis, K. A., Guilhaumon, F., & Whittaker, R. J. (2012). The island species area relationship: Biology and statistics. *Journal of Biogeography*, 39(2), 215–231.
4. Drakare, S., Lennon, J. J., & Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, 9(2), 215–227.
5. Lomolino, M. V. (1989). Interpretations and comparisons of constants in the species–area relationship: An additional caution. *The American Naturalist*, 133(2), 277–280.
6. Fattorini, S., Borges, P. A., Dapporto, L., & Strona, G. (2017). What can the parameters of the species–area relationship (SAR) tell us? Insights from Mediterranean islands. *Journal of Biogeography*, 44(5), 1018–1028.
7. 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(8), 1405–1414.
8. Forster, J. R., Thomas, N., & Guest, H. (1778). *Observations made during a voyage round the world*. University of Hawaii Press.
9. Arrhenius, O. (1921). Species and area. *Journal of Ecology*, 9(1), 95–99.
10. Gleason, H. A. (1922). On the relation between species and area. *Ecology*, 3(2), 158–162.
11. Plotkin, J. B., Potts, M. D., Douglas, W. Y., Bunyavejchewin, S., Condit, R., Foster, R., Seng, L. H. (2000). Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences*, 97 (20), 10850–10854.
12. Hubbell, S. P. (2001). The unified neutral theory of species abundance and diversity. *Princeton University Press, Princeton, NJ. Hubbell, SP (2004) Quarterly Review of Biology* , 79, 96–97.
13. Gitay, H., Roxburgh, S. H., & Wilson, J. B. (1991). Species–area relations in a New Zealand tussock grassland, with implications for nature reserve design and for community structure. *Journal of Vegetation Science*, 2(1), 113–118.
14. Simberloff, D., & Gotelli, N. (1984). Effects of insularisation on plant species richness in the prairie–forest ecotone. *Biological Conservation*, 29(1), 27–46.
15. Tjørve, E. (2012). Arrhenius and Gleason revisited: New hybrid models resolve an old controversy. *Journal of Biogeography*, 39(4), 629–639.
16. 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(5), 607–618.
17. Tsallis, C. (2016). Inter-occurrence times and universal laws in finance, earthquakes and genomes. *Chaos, Solitons & Fractals*, 88, 254–266.
18. Martinez, A. S., Gonzalez, R. S., & Ter, cariol, C. A. S. (2008). Generalized exponential function and some of its applications to complex systems. *arXiv preprint arXiv:0812.3071*
19. Weigelt, P., König, C., & Kreft, H. (2020). Gift—A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography*, 47(1), 16–43.
20. Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.
21. Akaike, H. (1978). On the likelihood of a time series model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 27(3–4), 217–235.
22. Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* , 81.
23. Triantis, K., Mylonas, M., Lika, K., & Vardinoyannis, K. (2003). A model for the species area–habitat relationship. *Journal of Biogeography*, 30(1), 19–27.
24. Azovsky, A. I. (2011). Species–area and species–sampling effort relationships: Disentangling the effects. *Ecography*, 34(1), 18–30.
25. Matthews, T. J., Triantis, K. A., & Whittaker, R. J. (2021). *The species–area relationship: Theory and application*. Cambridge University Press.
26. Palmer, M. W., & White, P. S. (1994). Scale dependence and the species–area relationship. *The American Naturalist*, 144(5), 717–740.
27. Arrhenius, O. (1923). On the relation between species and area — A reply. *Ecology*, 4(1), 90–91.
28. Davies, S. J., Abiem, I., Salim, K. A., Aguilar, S., Allen, D., Alonso, A., . . . others (2021). Forestgeo: Understanding forest diversity and dynamics through a global observatory network. *Biological Conservation*, 253 , 108907.
29. Boghosian, B. M. (1996). Thermodynamic description of the relaxation of two-dimensional turbulence using Tsallis statistics. *Physical Review E*, 53(5), 4754.
30. Martinez, S., & Nicol´as, F., Pennini, F., & Plastino, A. (2000). Tsallis entropy maximization procedure revisited. *Physica A: Statistical Mechanics and its Applications*, 286(3–4), 489–502.
31. Cajueiro, D. O. (2006). A note on the relevance of the q-exponential function in the context of intertemporal choices. *Physica A: Statistical Mechanics and Its Applications*, 364, 385–388.
32. Borges, E. P. (2004). A possible deformed algebra and calculus inspired in nonextensive thermostatics. *Physica A: Statistical Mechanics and its Applications*, 340(1–3), 95–101.
33. Carey, M. (2021). *Statistical analysis of dynamic patterns in insular biogeography (Un-published doctoral dissertation)*. University of South Australia.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.