APPLICATION

Program SimAssem: software for simulating species assemblages and estimating species richness

Gordon C. Reese[1]*, Kenneth R. Wilson[1] and Curtis H. Flather[2]

[1]*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA; and* [2]*Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO 80526, USA*

Summary

1. Species richness, the number of species in a defined area, is the most frequently used biodiversity measure. Despite its intuitive appeal and conceptual simplicity, species richness is often difficult to quantify, even in well-surveyed areas, because of sampling limitations such as survey effort and species detection probability. Nonparametric estimators have generally performed better than other options, but no particular estimator has consistently performed best across variation in assemblage and survey parameters.

2. In order to evaluate estimator performances, we developed the program SimAssem. SimAssem can: (i) simulate assemblages and surveys with user-specified parameters, (ii) process existing species encounter history files, (iii) generate species richness estimates not available in other programs and (iv) format encounter history data for several other programs.

3. SimAssem can help elucidate relationships between assemblage and survey parameters and the performance of species richness estimators, thereby increasing our understanding of estimator sensitivity, improving estimator development and defining the bounds for appropriate application.

Key-words: biodiversity, community ecology, landscape ecology, relative abundance distribution, simulation, spatial aggregation

Introduction

Species richness, the number of unique species in a defined area, is the most commonly used measure of biological diversity (Gaston 1996; Moreno *et al.* 2006). Species richness (SR) can be used to delineate protected areas, monitor biological systems and investigate environmental relationships. Surveys rarely encounter all of the species in an area; therefore, numerous estimators have been proposed to improve upon the negative bias of raw counts.

Three categories are regularly used to classify SR estimators (Colwell & Coddington 1994). The first category includes extrapolative methods applied to species accumulation curves or species–area curves. The Michaelis–Menten equation (Michaelis & Menten 1913), negative exponential model (Holdridge *et al.* 1971) and power model (Arrhenius 1921; Tjørve 2009) are commonly used to extrapolate to an estimate of SR at some large sample or large area.

A second category includes parametric estimators that make assumptions about the underlying species-abundance distribution or species detection probabilities (*p*). One type of parametric estimator uses a fitted distribution, often either a log-normal or log-series. For this category, required steps such as estimating total abundance and selecting the discrete abundance classes to which a continuous distribution is fit are often prohibitive (see Colwell & Coddington 1994; Magurran 2004). There are also parametric estimators based on the assumption that *p* is constant across species.

A third category includes nonparametric estimators, which are those that are neutral on the probability distribution from which parameters are drawn. Many of the nonparametric SR estimators were originally derived from methods to estimate the number of individuals in a closed population (e.g. Burnham & Overton 1978; Chao 1984; Pledger 2000).

The search for a single best estimator has not yet been resolved. However, general comparisons of the three estimator categories favour the nonparametric methods (see table 1 in Cao, Larsen & White 2004; table 3 in Walther & Moore 2005). Nonparametric estimators are therefore the focus of this project.

The performance of nonparametric SR estimators can be affected by species- and assemblage-level attributes as well as by survey design parameters, hereafter collectively referred to as factors (Keating & Quinn 1998; Brose, Martinez & Williams 2003). Several studies have indicated that bias decreases as species-abundance distributions become more even (Wagner & Wildi 2002; O'Dea, Whittaker & Ugland 2006). One assumption of the closed population estimators, translated for species data, holds that species are equally detectable across space. Spatial aggregation regularly challenges this assumption

*Correspondence author. E-mail: greese126@gmail.com

(Schmit, Murphy & Mueller 1999). Other factors found to affect SR estimator performance include the number of species (Keating & Quinn 1998; Poulin 1998), total abundance or density of individuals (Baltanás 1992; Walther & Morand 1998) and species detection probability, *p* (Boulinier *et al.* 1998).

Raw sample data and consequently, SR estimates, are also affected by survey design parameters such as effort (Burnham & Overton 1979; Brose, Martinez & Williams 2003). Additionally, survey configuration has been important to other estimation issues (Reese *et al.* 2005). Selecting survey locations randomly is unbiased and therefore preferable; however, survey locations are often selected based on accessibility and previous results (Beck & Kitching 2007). The above factors can all affect sample coverage (*sc*), which is the proportion of a species pool represented in a sample and the single most important factor with respect to estimator performance (Baltanás 1992; Brose, Martinez & Williams 2003). Unfortunately, one needs to know the true number of species to calculate *sc* and, if this information were available, estimation would be unnecessary. It is therefore important to understand how the aforementioned factors affect performance.

Evaluating SR estimators across a wide range of factors in the field is difficult because of temporal, financial and logistical constraints as well as uncertainty about species- and assemblage-level parameters. Despite the simplifications, simulations are advantageous because they can be systematically varied and randomly surveyed, and most important, the true number of species is known. Our objective therefore was to develop a program in which specified parameters are used to simulate and survey species assemblages, thereby revealing the behaviour of SR estimators in a controlled setting. Most, if not all, of the programs currently available for estimating SR, for example, EstimateS (Colwell 2006), SPADE (Chao & Shen 2010) and ws2m (Turner, Leitner & Rosenzweig 2003), process existing encounter history data (information indicating whether a species was encountered during a particular survey occasion), but include little or no simulation capability. In addition, SimAssem includes a more comprehensive suite of SR and variance estimators than other programs.

Program SimAssem

SimAssem is application software developed in Visual Basic 6.0 for 32-bit versions of Microsoft Windows and includes a graphical user interface (Fig. 1) and internal dialogue with R software (R Development Core Team 2009). SimAssem can
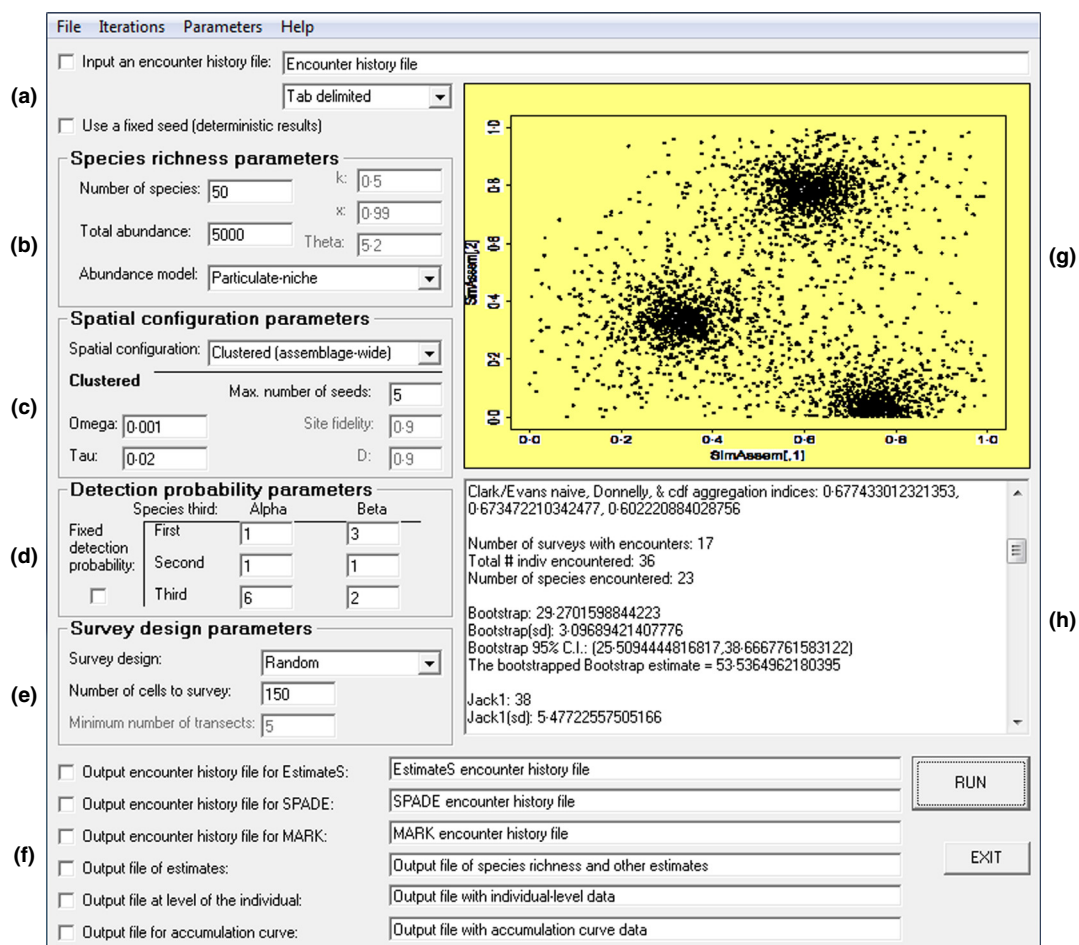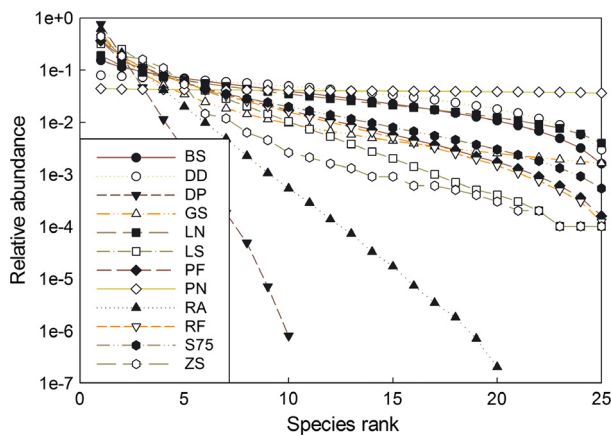

Fig. 1. The graphical user interface for SimAssem which includes sections for: inputting a file (a); setting simulation parameters for: species and abundance (b), spatial configuration (c), species detection probability (d), and survey design (e); outputting files (f); displaying a simulated assemblage (g); and displaying estimates (h).

process both existing encounter history data and data from assemblages simulated with user-specified parameters. Other than for some R functions, the Mersenne twister pseudorandom number generator is used for randomizations (Matsumoto & Nishimura 1998). The SimAssem program and source code are available at http://sourceforge.net/projects/simassem/. We hope that the source code provides a valuable foundation for the quick evaluation and development of new estimators.

SIMULATING AN ASSEMBLAGE

In SimAssem, assemblages are simulated by specifying the number of species (S), total abundance across species (N) and a distribution to which abundances conform (Fig. 2, Table 1). SimAssem will run only when $N \geq S$.

The fundamental attributes of an assemblage include its species richness and species-abundance distribution. Two established theories about species abundance are: (i) abundances are generally unequal amongst species and (ii) most species are relatively rare (Fisher, Corbet & Williams 1943). The geometric-series (Motomura 1932), log-normal (Preston 1948) and log-series distributions (Fisher, Corbet & Williams 1943) have been successfully fit to biological datasets; however, the representation of abundance distributions with purely mathematical models has been criticized for not explaining the patterns. Some of the earliest alternatives focusing instead on process include the broken-stick and particulate-niche models (MacArthur 1957). More recent work with abundance distributions continued to emphasize the methodological steps required to create a distribution and, by way of analogy, the ecological processes that result in real abundance distributions (Tokeshi 1990, 1993, 1996). These models are assumed to approximate the interactions and subsequent patterns of small groups of taxonomically related species, that is, species vying

Table 1. Species-abundance distributions in SimAssem. See the referenced publications and Appendix S1 for details

Abundance distributions	Descriptive literature
Broken-stick (BS)[1]	Tokeshi (1990)
Dominance-decay (DD)[1]	Tokeshi (1990)
Dominance-preemption (DP)[1]	Tokeshi (1990)
Geometric-series (GS)	Tokeshi (1990)
Log-normal (LN)[1]	Appendix S1
Log-series (LS)	Magurran (2004)
Power-fraction (PF)[1]	Tokeshi (1996)
Particulate-niche (PN)[1]	Tokeshi (1993)
Random-assortment (RA)[1]	Tokeshi (1990)
Random-fraction (RF)[1]	Tokeshi (1990)
Sugihara's sequential model (S75)[1]	Sugihara (1980), Tokeshi (1993)
Zero-sum (ZS)[1]	Hubbell (2001)

[1]The allocation of abundances to species is stochastic. Final abundances are the integer portion of averages across the specified number of iterations (*Iterations* menu item).

Table 2. Spatial configuration algorithms in SimAssem. User-specified parameters are described in Appendix S1 and include distance (D) [0–1], fidelity (F) [0–1], maximum number of seeds (Sds), and the length of the shoulder (τ) [0–1] and rate of decline (ω) [0–1] for a distance-decay formula, $1 - (1 - \omega^{DistanceToSeed})^\tau$
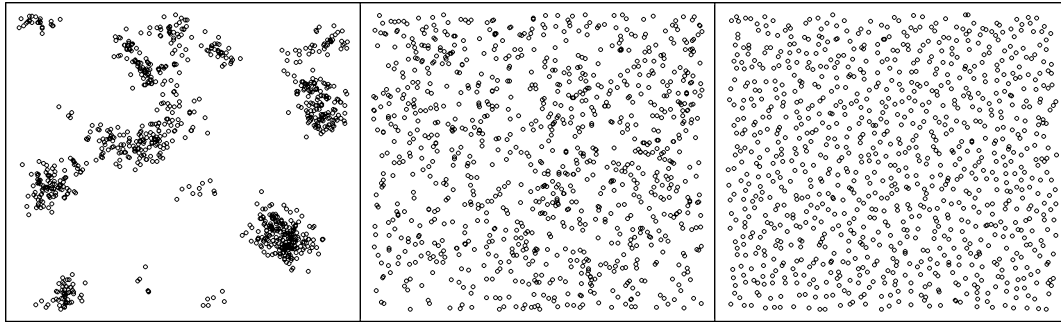
Configuration algorithms	Parameters
Aggregated (centres)	D, F
Aggregated (centres equal abun)	D, F
Aggregated (individuals)	D, F
Aggregated (individuals max dist)	D, F
Clustered (assemblage-wide)	Sds, τ, ω
Clustered (species-specific)	Sds, τ, ω
Hyper-dispersed	None
Random	None

for the same resources, and have therefore been termed niche-based models. A basic premise holds that niche apportionment can be modelled by a stick being broken, where the units of the stick represent individuals. Methods for generating the available species-abundance distributions and worked examples are given in Appendix S1.

Several different algorithms are included for distributing individuals across a square landscape (Table 2). Possible spatial patterns range from aggregated (species-specific or assemblage-wide), to random, to hyper-dispersed (more evenly spaced than random) (Fig. 3; see Appendix S1 for details).

CREATING ENCOUNTER DATA: DETECTION AND DESIGN

Before *ps* are assigned, species are grouped into thirds based on abundance, for example, one group is comprised of the least abundant species. A randomly selected group is increased by one for each species that remains when the true number of species (S_{true}) is not a factor of three. Within each group, species-specific *ps* can be randomly drawn from a beta distribution with specified α and β parameters (R function *rbeta*). Beta distributions are characterized by an expected value (mean), E

Fig. 2. Species-abundance distributions available in SimAssem (see Table 1 for abbreviations). In this example, we specified 25 species and 10 000 total individuals, and other than for the ZS distribution, each graph point was the average of 1000 iterations. Due to stochasticity, the line shown for ZS represents the first iteration with 25 species ($\theta = 4$). For the LS distribution, $x = 0.99969074$ (see Appendix S1 for details).

Fig. 3. Example configuration patterns with 1 species and 1000 individuals, including aggregated (left panel), random (middle panel) and hyper-dispersed (right panel).

$(X) = \alpha/(\alpha + \beta)$ and variance, $\text{var}(X) = (\alpha\beta)/[(\alpha + \beta)^2 (\alpha + \beta + 1)]$, where X is a random beta variate. Additionally, ps can be fixed for each abundance group.

Each simulated landscape is partitioned by a 100×100 grid for the purpose of conducting surveys. User-settings include the *Survey design*, that is, spatial configuration of surveyed grid cells, and *Number of cells to survey* (t), from 1 to 10 000. Cells are surveyed without replacement.

SimAssem includes two survey designs. Surveyed grid cells can be randomly selected (*Random*) or added to randomly oriented, horizontal or vertical linear transects that are each one grid cell wide (*Linear transect*). The *Linear transect* option requires a *Minimum number of transects* (m) across which t is divided. Due to landscape dimensions, the maximum transect length is the smaller of 100 or t/m. When transect length is truncated to 100 or when m is not a factor of t, additional transects are added until the number of surveyed cells equals t. One random uniform variate is drawn for every individual in a surveyed cell, and an individual is encountered when the random uniform variate $\leq p$.

Estimating species richness

SimAssem includes numerous SR estimators. Two, in particular, are computer intensive estimators, that is, CY-1 and CY-2 (see Table 3 for estimators and abbreviations), that performed relatively well in comparative studies (Reese 2012), but are unavailable elsewhere. Log-transformed variance estimates are used to restrict the lower bound of 95% confidence intervals to the number of species observed, S_{obs} (Burnham *et al.* 1987, part 3). Several estimators involve iterations that can be set under the *Iterations* menu item. For more details and estimator equations, see Appendix S1.

Additional output

Other reported values, some requiring simulated data, include: (i) the number of simulated species, (ii) the total number of simulated individuals, (iii) the number of species observed, (iv) the number of surveys with encounters, (v) the total number of individuals encountered, (vi) the true and estimated sample coverage (via CY-1), (vii) Shannon's evenness index (Shannon & Weaver 1949) and (viii) Clark and Evans aggregation index

(Clark & Evans 1954). Two diversity indices are also reported, Margalef's diversity index (Clifford & Stephenson 1975) and Menhininck's index (Whittaker 1977).

Biological surveys are generally expensive and often provide diminishing returns on investment, that is, effort. SimAssem includes two estimators of the additional effort needed to encounter a user-specified proportion (*Parameters* menu item) of certain SR estimators (Chao *et al.* 2009). One estimates the additional number of individuals needed to encounter the specified proportion of Chao1, thus requiring abundance data. An incidence-based version estimates the number of additional surveys, for example, quadrats, needed to encounter a user-specified proportion of Chao2.

Import and export options

SimAssem can import comma-, space- and tab-delimited encounter history data saved as a plain text file. The first line is disregarded and therefore useful for documentation, line two must contain two numbers, S_{obs} and the number of surveys conducted, and line three must begin the encounter history data, where each row represents a different species and each column a different survey result, either by abundance, that is, the actual number of individuals encountered, or by incidence, that is, a one indicates that one or more individuals were encountered and a zero that there were zero encounters.

There are several export options available in SimAssem. Estimates can be exported to a comma-delimited file, where the first line lists estimator names and the following lines list the estimates. Encounter history data can be formatted for programs EstimateS (Colwell 2006), MARK (White & Burnham 2009) and SPADE (Chao & Shen 2010). Also, individual-level data can be exported to a comma-delimited text file including (in the following order): a numerical species identifier, x-coordinate, y-coordinate, the grid cell in which it fell (beginning with 1 in the lower left corner and proceeding first across and then up), p, and whether the individual was encountered (1) or not (0). Accumulation curve data are also exportable where, beginning with the specified survey size (1-t) and increasing sequentially by that amount, surveys are randomly drawn without replacement, and estimates at each survey size are averaged over a user-specified number of replications.

Table 3. Species richness estimators and abbreviations used in SimAssem

Estimator[1]	Abbreviation	Reference
Abundance-based coverage (a)	ACE	Chao & Lee (1992)
Bootstrap[2] (i)	Boot	Smith & van Belle (1984)
Bootstrap; iterated[3] (i)	Boot-B	Smith & van Belle (1984)
Chao1[2] (a)	Chao1	Chao (1984)
Chao1 (bias-corrected)[2] (a)	Chao1BC	Chao (2005)
Chao2[2] (i)	Chao2	Chao (1987)
Chao2 (bias-corrected)[2] (i)	Chao2BC	Chao (2005)
Coverage-adjusted (i)	C1,C2,C3	Ashbridge & Goudie (2000)
CY-1[3] (i)	CY-1	Cao, Larsen & Hughes (2001)
CY-2 (i)	CY-2	Cao, Larsen & White (2004)
Darroch–Ratcliff (i)	DR	Darroch & Ratcliff (1980)
Incidence-based coverage (i)	ICE	Lee & Chao (1994)
1st-order jackknife[2] (i)	Jack1	Burnham & Overton (1978)
2nd-order jackknife[2] (i)	Jack2	Burnham & Overton (1978)
3rd-order jackknife[2] (i)	Jack3	Burnham & Overton (1978)
4th-order jackknife[2] (i)	Jack4	Burnham & Overton (1978)
5th-order jackknife[2] (i)	Jack5	Burnham & Overton (1978)
Mixture-model (i)	Mixture	Pledger (2000)
Observed species count	S_{obs}	

[1]A description of the estimator where (a) indicates that the estimator uses sample abundance data, that is, number of individuals, and (i) indicates that the estimator uses sample incidence data, that is, presence/absence in surveys.

[2]Variance is estimated by an analytically derived estimator.

[3]Variance is estimated by the variance across iterations.

Utility

SimAssem allows users to quickly and easily evaluate numerous estimators across a wide range of assemblages. Such investigations are difficult, if not impossible, in the real world because of both sampling limitations and uncertainty regarding the true assemblage parameters. Simulating assemblages involves a considerable amount of stochasticity; therefore, SimAssem provides the option to set the number of runs with a specific set of parameters in order to facilitate comparisons.

We envision SimAssem being used to compare estimator performance under surveyed or expected field conditions, thereby improving estimator selection for a particular application. For example, suppose that data were collected from an assemblage with an apparent log-normal species-abundance distribution, species-specific spatial aggregation and where species detection probabilities varied around an average of 0.3. SimAssem could reveal that coverage-based estimators (see Chao & Lee 1992; Lee & Chao 1994) are less biased than other estimators in assemblages simulated with similar characteristics and robust to variation in the degree of spatial aggregation. Furthermore, SimAssem could provide an estimate of the amount of bias (e.g. −15%) given the level of effort expended.

References

Arrhenius, O.(1921) Species and area. *Journal of Ecology*, **9**, 95–99.

Ashbridge, J. & Goudie, I.B.J. (2000) Coverage-adjusted estimators for mark-recapture in heterogeneous populations. *Communications in Statistics: Simulation and Computation*, **29**, 1215–1237.

Baltanás, A. (1992) On the use of some methods for the estimation of species richness. *Oikos*, **65**, 484–492.

Beck, J. & Kitching, I.J. (2007) Estimating regional species richness of tropical insects from museum data: a comparison of a geography-based and sample-based methods. *Journal of Applied Ecology*, **44**, 672–681.

Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E. & Pollock, K.H. (1998) Estimating species richness: the importance of heterogeneity in species detectability. *Ecology*, **79**, 1018–1028.

Brose, U., Martinez, N.D. & Williams, R.J. (2003) Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*, **84**, 2364–2377.

Burnham, K.P. & Overton, W.S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**, 623–633.

Burnham, K.P. & Overton, W.S. (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology*, **60**, 927–936.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. & Pollock, K.H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph No. 5*. Bethesda, Maryland.

Cao, Y., Larsen, D.P. & Hughes, R.M. (2001) Estimating total species richness in fish assemblage surveys: a similarity based approach. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1782–1793.

Cao, Y., Larsen, D.P. & White, D. (2004) Estimating regional species richness using a limited number of survey units. *Ecoscience*, **11**, 23–35.

Chao, A. (1984) Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chao, A. (2005) Species estimation and applications. *Encyclopedia of Statistical Sciences*, 2nd edn., Vol. **12** (eds S. Kotz, N. Balakrishnan, C.B. Read & B. Vidakovic), pp. 7907–7916. Wiley Press, New York.

Chao, A., Colwell, R.K., Lin, C.W. & Gotelli, N.J. (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology*, **90**, 1125–1133.

Chao, A. & Lee, S.M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Chao, A. & Shen, T.-J. (2010) *Program SPADE (Species Prediction and Diversity Estimation)*. Available at: http://chao.stat.nthu.edu.tw/softwareCE.html [accessed 1 June 2013].

Clark, P.J. & Evans, F.C. (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35**, 445–453.

Clifford, H.T. & Stephenson, W. (1975) *An Introduction to Numerical Classification*. Academic Press, London.

Colwell, R.K. (2006). *EstimateS Statistical Estimation of Species Richness and Shared Species from Samples*, Version 8. Available at: http://viceroy.eeb.uconn.edu/EstimateS/ [accessed 1 June 2013].

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London B*, **345**, 101–118.

Darroch, J.N. & Ratcliff, D. (1980) A note on capture-recapture estimation. *Biometrics*, **36**, 149–153.

Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.

Gaston, K.J. (1996) *Biodiversity: A Biology of Numbers and Difference*. Blackwell Science, Oxford.

Holdridge, L.R., Grenke, W.C., Hatheway, W.H., Liang, T. & Tosi, J.A. (1971) *Forest Environments in Tropical Life Zones*. Pergamon Press, Oxford.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.

Keating, K.A. & Quinn, J.F. (1998) Estimating species richness: the Michaelis-Menten model revisited. *Oikos*, **81**, 411–416.

Lee, S.M. & Chao, A. (1994) Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*, **50**, 88–97.

MacArthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Science USA*, **43**, 293–295.

Magurran, A.E. (2004) *Measuring Biological Diversity*. Blackwell Publishing, Malden.

Matsumoto, M. & Nishimura, T. (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**, 3–30.

Michaelis, M. & Menten, M.L. (1913) Der kinetic der invertinwirkung. *Biochemische Zeitschrift*, **49**, 333–369.

Moreno, C., Zuria, I., García-Zenteno, M., Sánchez-Rojas, G., Castellanos, I., Martínez-Morales, M. *et al.* (2006) Trends in the measurement of alpha diversity in the last two decades. *Interciencia*, **31**, 67–71.

Motomura, I. (1932) On the statistical treatment of communities. *Japanese Journal of Zoology*, **44**, 379–383 (in Japanese).

O'Dea, N., Whittaker, R.J. & Ugland, K.I. (2006) Using spatial heterogeneity to extrapolate species richness: a new method test on Ecuadorian cloud forest birds. *Journal of Applied Ecology*, **43**, 189–198.

Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, **56**, 434–442.

Poulin, R. (1998) Comparison of three estimators of species richness in parasite component communities. *Journal of Parasitology*, **84**, 485–490.

Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org [accessed 1 June 2013].

Reese, G.C. (2012) *Simulating species assemblages and evaluating species richness estimators*. PhD thesis, Colorado State University, Fort Collins. Available at: http://hdl.handle.net/10217/68193 [accessed 1 June 2013].

Reese, G.C., Wilson, K.R., Hoeting, J.A. & Flather, C.H. (2005) Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications*, **15**, 554–564.

Schmit, J.P., Murphy, J.F. & Mueller, G.M. (1999) Macrofungal diversity of a temperate oak forest: a test of species richness estimators. *Canadian Journal of Botany*, **77**, 1014–1027.

Shannon, C.E. & Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Smith, E.P. & van Belle, G. (1984) Nonparametric estimation of species richness. *Biometrics*, **40**, 119–129.

Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *The American Naturalist*, **116**, 770–787.

Tjørve, E. (2009) Shapes and functions of species-area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36**, 1435–1445.

Tokeshi, M. (1990) Niche apportionment or random assortment: species abundance patterns revisited. *Journal of Animal Ecology*, **59**, 1129–1146.

Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, **24**, 111–186.

Tokeshi, M. (1996) Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos*, **75**, 543–550.

Turner, W., Leitner, W. & Rosenzweig, M. (2003) *ws2m Software for the Measurement and Analysis of Species Diversity*. Version 3.2. Available at: http://www.mybiosoftware.com/population-genetics/9836 [accessed 1 June 2013].

Wagner, H.H. & Wildi, O. (2002) Realistic simulation of the effects of abundance distribution and spatial heterogeneity on non-parametric estimators of species richness. *Ecoscience*, **9**, 241–250.

Walther, B.A. & Moore, J.L. (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**, 815–829.

Walther, B.A. & Morand, S. (1998) Comparative performance of species richness estimation methods. *Parasitology*, **116**, 395–405.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46** (Supplement), 120–138.

Whittaker, R.H. (1977) Evolution of species diversity in land communities. *Evolutionary Biology*, Vol. 10 (eds M.K. Hecht, W.C. Steere & B. Wallace), pp. 1–67. Plenum Publishing Corporation, New York.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Details and worked examples of the options in program SimAssem.