

Estimating the richness of species with variable mobility

Ulrich Brose and Neo D. Martinez

Brose, U. and Martinez, N. D. 2004. Estimating the richness of species with variable mobility. – *Oikos* 105: 292–300.

The vast majority of species are animals that, unlike most plants and fungi, are variably and often highly mobile. While species' mobility affects species' probabilities of being sampled, effects of movement on the estimation of species richness have yet to be systematically investigated. Information-rich abundance-based estimators may be able to address variably mobile species but the accuracy of these estimators has also yet to be investigated. Here, we address both issues by variably sampling simulated landscapes with up to 250 species and evaluating the performance of ten non-parametric estimators and one species accumulation curve. Our results show that some abundance-based estimators are as accurate as better known and tested incidence-based estimators. Increased movement heterogeneity between the species reduced estimator performance by reducing the sample coverage, which systematically determined which estimator was most accurate. Based on these findings, we present the first decision framework for choosing the most accurate of many available abundance-based species-richness estimators. These decisions, based on data coverage, can significantly improve investigators' ability to estimate faunal species richness.

U. Brose, Romberg Tiburon Center, Dept of Biology, San Francisco State Univ., 3152 Paradise Drive, Tiburon, CA 94920, USA (brose@sfsu.edu). – N. D. Martinez, Pacific Ecoinformatics and Computational Ecology Lab, Rocky Mountain Biological Laboratory, POB 519, Gothic, CO 81224, USA.

Species richness is often regarded as the fundamental unit of biodiversity, and is the most frequently applied measure in community ecology (Gaston 1996, Martinez 1996, Williams and Martinez 2000). As animals in general and arthropods in particular contribute most to overall diversity (May 1988), they should play a significant role in the development of ecological theory. However, because arthropods are typically small, express a wide range of mobility, and require enormous sampling intensity to count all species in diverse communities, they have been largely ignored in favour of larger, less mobile organisms such as trees and other plants. This study systematically investigates problems inherent to estimating the species richness of arthropods and any other variably mobile group of organisms that generally prohibit complete enumeration of all species in a reasonably sized area or habitat. Addressing these problems

should help the largest fraction of biodiversity to contribute more to biodiversity theory and management.

Several aspects of estimating species richness are well known. Species accumulation curves and the non-parametric methods appear to be the most promising estimators for a variety of conditions (Colwell and Coddington 1994, Gaston 1996). Alternative methods, such as fitting parametric models of the species relative abundances to data sets, appear problematic (Colwell and Coddington 1994) and perform poorly in case studies (Palmer 1990). Among the better performing estimators in plant communities, most extrapolations perform substantially better than the observed number of species (S_{obs}) and non-parametric estimators are usually more accurate than species accumulation curves (Palmer 1990, 1991, Colwell and Coddington 1994, Chazdon et al. 1998, Brose et al. 2003). Such findings

Accepted 25 September 2003

Copyright © OIKOS 2004
ISSN 0030-1299

are challenged by the fact that in empirical studies the true species richness, S_{true} , is rarely known for any reasonably large and diverse area or habitat. Brose et al. (2003) avoided this problem by using simulated landscapes to show that uneven abundance distributions violate the estimators' intrinsic assumption of equal detection probabilities for all species. This causes systematic underestimations of S_{true} , especially by the species accumulation curves, but also by non-parametric estimators. Such underestimation was unaffected by spatial gradients and autocorrelation. Here, we move beyond Brose et al.'s (2003) exclusive focus on sessile species and incidence-based estimators by extending the analysis to include variably mobile species and abundance-based estimators.

Comparative studies of estimating animal species richness including parasites (Walther and Morand 1998), birds (Boulinier et al. 1998) and ground dwelling arthropods (Brose 2002, Longino et al. 2002) have not evaluated the effects of movement heterogeneity on detection probabilities. To our knowledge, we are the first to systematically evaluate how movement heterogeneity among species alters the performance of species richness estimators. Other aspects of incidence-based non-parametric estimators have been evaluated (Palmer 1990, 1991, Baltanás 1992, Colwell and Coddington 1994, Chazdon et al. 1998) but these estimators require a system to be sampled several, preferably many, times using multiple sample locations. Most faunistic data-sets lack such sub-samples and instead quantify species' abundances in one total sample. Abundance-based forms of the non-parametric estimators for these single sample situations have been developed, but this study appears to be the first to systematically evaluate their accuracy.

Methods

Simulations

We simulated lattices of 100 by 50 cells (x, y), each representing one square meter. The species abundances were modelled according to the stochastic random-fraction model (Tokeshi 1993) that yields similar abundance distributions as Prestons (1948) deterministic lognormal model (Tokeshi 1993, Fig. 1). The individuals were randomly placed on the lattice. We simulated individuals randomly walking across the lattice for a total of 60 time steps. Their movement was coded independently for every species as the number of steps per time step. To suppress boundary effects, we have chosen wrap-around margins, in which every individual leaving the lattice in one direction re-enters at the opposite margin. Sampling cells were randomly placed on the lattice mimicking traps. Each individual stepping

on a sampling cell was added to the sampling list of the respective cell and did not move on.

Two different simulations were carried out. First, we compared estimator performance in communities with normal, log-normal, and bimodal-based on two overlapping normal distributions-frequency distributions of movement among species. These communities had the same movement heterogeneity: mean movement of 8 steps per time step and standard deviation of 4. Second, we studied the impact of varying movement heterogeneity on the estimators. All communities were simulated with a log-normal distribution of movement among species, but different heterogeneity: while the mean was constantly kept at a value of 8, the standard deviations varied between 2, 8 and 16.

Monte Carlo simulations with 250 replications per community type were repeated for communities with a true number of species (S_{true}) equal to 50 and 250 species. Our previous analyses (Brose et al. 2003) show that results obtained for these levels of S_{true} successfully represent results from a larger range (25–500) of S_{true} . Each of the simulations of 50 and 250 species communities was sampled at three levels of intensity of the 5000-cell communities: low (5 cells), intermediate (10 cells) and high intensity (20 cells). Two hundred and fifty replications of 50- and 250-species communities repeated for three movement patterns and sampled at low, intermediate and high levels generated 4500 landscapes each with their own sample used to estimate species richness in each of the two simulations.

Estimators

We compared the extrapolations of one species accumulation curve and ten non-parametric estimators (Appendix 1). This included the most frequently used species accumulation curve, the Michaelis-Menten model (Colwell and Coddington 1994). To avoid arbitrary effects, cell sample order was randomized 100 times in each sample. The mean species number $S(n)$ was calculated for each value of n cells sampled. Following Keating and Quinn (1998), we used least squares nonlinear regressions with the quasi-Newton algorithm to fit the Michaelis-Menten model to the mean $S(n)$ data. Note that the often discussed difference between number of individuals and number of samples as measurements of sampling intensity (Gotelli and Colwell 2001) does not apply to our simulations, because the individuals are evenly distributed across the lattice.

The ten non-parametric estimators used include the first to third order jackknife estimators and their respective abundance-based forms (Jack1–Jack3, Jack_a1–Jack_a3, Burnham and Overton 1979), the abundance-based estimator Chao1 (Chao 1984), the incidence-based estimator Chao2 (Chao 1987), the abundance-based

coverage estimator ACE (Chao and Lee 1992, Chao et al. 1993), and the incidence-based coverage estimator ICE (Lee and Chao 1994). To avoid complications by undefined cases, we used the corrected versions of the estimators Chao1, Chao2, ACE and ICE given by Colwell (1997, see Appendix 1).

Estimator performance

For every estimator, we calculated bias, imprecision and inaccuracy. Bias was defined as the mean proportional deviation of the estimates S_{est} from S_{true} :

$$\text{Bias} = \frac{S_{est} - S_{true}}{S_{true}}$$

Positive and negative bias indicates overestimation and underestimation, respectively. The mean bias over many replicates indicates systematic over- or underestimation, but a low mean bias does not necessarily imply that individual estimates are close to the true value as the average of alternate over- and underestimations may be unbiased (Sokal and Rohlf 1995). As precision is the closeness of repeated measurements of the same quantity to each other (Sokal and Rohlf 1995), we measured imprecision by the standard deviation of the bias. Precision measures the similarity of independent estimates and, therefore, it indicates the repeatability of an estimate, whether or not it is biased. Accuracy is the closeness of a measured or computed value to its true value (Sokal and Rohlf 1995). We measured inaccuracy using the mean square proportional deviation of S_{est} from S_{true} :

$$\text{Inaccuracy} = \left(\frac{S_{est} - S_{true}}{S_{true}} \right)^2$$

Bias, accuracy and precision were measured as percent of S_{true} . A high mean accuracy of an estimator indicates

that individual estimates are on average close to the true value, which implies a low mean bias and a high mean precision. In a previous study, we have shown that accuracy emphasizes bias-reduction under low sampling intensity when all estimators have a substantial negative bias and precision under high sampling intensity when all estimators are less biased (Brose et al. 2003). Therefore, we used bias and imprecision for the general estimator evaluation and inaccuracy to demonstrate the influence of movement on estimator performance.

Results

We first describe results on the general performance of the estimators concerning bias and precision, and compare the relative performance of incidence- and abundance-based estimators. Then, we analyse the impact of movement patterns on estimator accuracy. Finally, we present a decision framework based on sample coverage for choosing among the abundance-based estimators.

General performance of the estimators

All estimators were substantially less biased than S_{obs} , irrespective of the sampling intensity and S_{true} in the simulations of movement distributions (Table 1). All estimators underestimated S_{true} under low to intermediate sampling intensity. The incidence- and abundance-based second and third order jackknife estimators, Chao1, Chao2 and ACE were positively biased under high sampling intensity. In most data-sets, the abundance-based forms of the jackknife estimators were marginally more biased than the respective incidence-based forms, but the differences were below 1.5% of S_{true} . In contrast, the abundance-based estimator ACE was substantially less biased than the incidence-based ICE

Table 1. General bias and imprecision of the estimators.

S_{true}	Bias						Imprecision					
	50			250			50			250		
	5	10	20	5	10	20	5	10	20	5	10	20
Traps												
S_{obs}	-0.397	-0.291	-0.176	-0.492	-0.377	-0.239	0.093	0.064	0.032	0.073	0.053	0.027
Jack1	-0.262	-0.138	-0.003	-0.353	-0.205	-0.026	0.073	0.036	0.011	0.059	0.030	0.005
Jack2	-0.202	-0.060	0.099	-0.285	-0.108	0.106	0.065	0.030	0.042	0.051	0.019	0.018
Jack3	-0.170	-0.008	0.178	-0.248	-0.042	0.208	0.065	0.045	0.106	0.047	0.014	0.052
Jack _{a1}	-0.275	-0.145	-0.006	-0.366	-0.212	-0.030	0.075	0.037	0.011	0.060	0.031	0.005
Jack _{a2}	-0.216	-0.066	0.097	-0.299	-0.116	0.102	0.068	0.030	0.042	0.052	0.020	0.017
Jack _{a3}	-0.184	-0.012	0.178	-0.263	-0.049	0.203	0.067	0.042	0.104	0.048	0.016	0.050
Chao1	-0.179	-0.052	0.151	-0.285	-0.124	0.094	0.573	0.572	1.805	0.058	0.040	0.098
Chao2	-0.233	-0.102	0.066	-0.297	-0.137	0.082	0.131	0.196	0.425	0.058	0.030	0.089
ICE	-0.331	-0.253	-0.105	-0.420	-0.330	-0.151	0.080	0.054	0.050	0.062	0.043	0.014
ACE	-0.229	-0.110	0.039	-0.306	-0.161	0.038	0.071	0.058	0.132	0.052	0.025	0.043
M-M	-0.281	-0.209	-0.119	-0.373	-0.282	-0.159	0.076	0.048	0.022	0.059	0.037	0.015

Notes: Average values across all community types are given. M-M = Michaelis-Menten model.

(Table 1). The abundance-based estimator Chao1 was less biased than the incidence-based Chao2 under low to intermediate sampling intensity, whereas Chao1 overestimated S_{true} more than Chao2 under high sampling intensity. In all of the 50-species communities, Chao1 and Chao2 were markedly imprecise (Table 1). In consequence of their low precision, the estimators Chao1 and Chao2 were excluded from the following analyses. Furthermore, due to the minor differences compared to the incidence-based forms, we focused on the abundance-based forms of the jackknife estimators.

Movement

In most data-sets, there were no differences in mean estimator accuracy between the different movement distributions (Fig. 1a). Significant differences only occurred in the highly sampled data sets and in the intermediately sampled 250-species communities (Kruskal-Wallis test with a Bonferroni corrected α' of 0.007). The magnitude of these differences, however, was always below 1.4% of S_{true} and usually below 1% of S_{true} . This indicated minor differences in estimator accuracy between the movement distributions simulated.

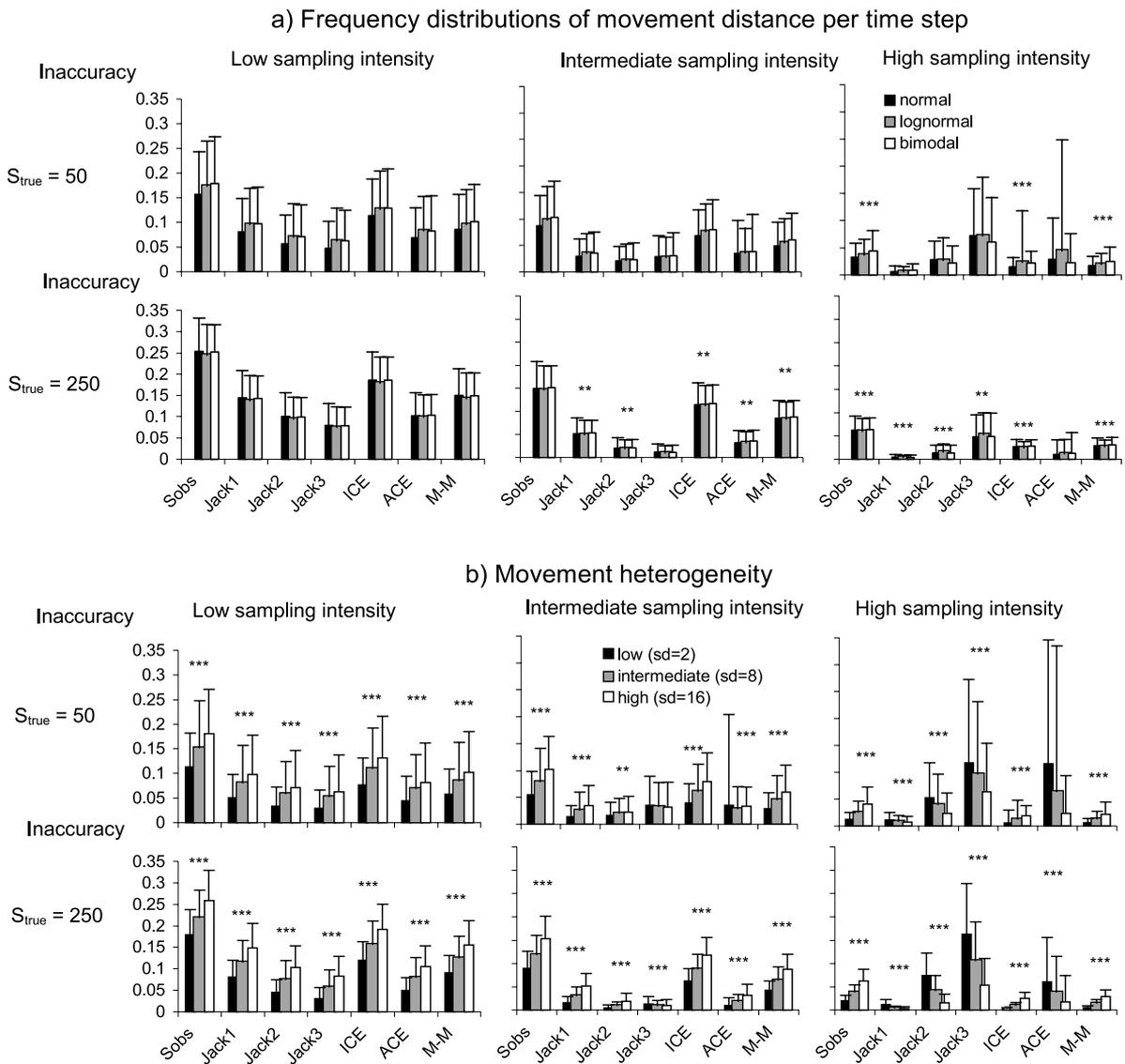


Fig. 1. Estimator inaccuracy (mean + 1 stdev.) at different levels of S_{true} and sampling intensity in dependence on a) the frequency distribution of movement among species and b) movement heterogeneity defined by standard deviations. Mean movement = 8 steps per time step. M-M = Michaelis-Menten model; Kruskal-Wallis test: *** = $p < 0.001$; ** = $p < 0.01$.

The second simulation based on communities with a log-normal distribution of movement revealed large differences in estimator accuracy caused by differences in movement heterogeneity (Fig. 1b). Most of these differences were highly significant (Kruskal-Wallis test with a Bonferroni corrected α' of 0.007). In all data-sets, the accuracy of S_{obs} and the Michaelis-Menten model decreased with increasing movement heterogeneity (Fig. 1b). The non-parametric estimators showed a more complex response: while their accuracy generally decreased with increasing movement heterogeneity under low sampling intensity, the opposite pattern was detected for most of them in the highly sampled communities. While S_{obs} was most sensitive to differences in movement heterogeneity under low and intermediate sampling intensity, Jack_{a3} and ACE were most sensitive at high sampling intensity (Fig. 1b).

Estimator performance did not only depend on movement heterogeneity but also on sampling intensity (Fig. 1b). While Jack_{a2}, Jack_{a3} and ACE were most accurate under low sampling intensity, Jack_{a1}, ICE and the Michaelis-Menten model performed best in the highly sampled communities (Fig. 1b). In contrast to this, Jack_{a2}, Jack_{a3} and ACE were less accurate than S_{obs} under high sampling intensity.

Sample coverage

We analysed the relative performance of the estimators with respect to sample coverage, i.e. the fraction of S_{true} present in the sample. As one might expect, sample coverage was negatively correlated with S_{true} (Spearman rank correlation, $r = -0.269^{***}$, $p < 0.001$) and positively correlated with sampling intensity ($r = 0.76^{***}$). Furthermore, sample coverage was negatively correlated with movement heterogeneity ($r = -0.246^{***}$). The accuracy of most estimators was significantly correlated with all these independent variables and sample coverage yielded the highest correlation coefficients (Table 2). The only two exceptions were that ACE was only correlated

with sample coverage and Jack_{a3} was most highly correlated with sampling intensity. After removing the effect of sample coverage in partial correlation analyses, only minor correlations with the other independent variables remained (Table 2). We plotted mean inaccuracy in one-percent steps against sample coverage and the resulting curves demonstrated that for every level of sample coverage, there was one most accurate estimator (Fig. 2). The intersections of these curves indicated that the most accurate estimators were Jack_{a3} up to a sample coverage of 64%, Jack_{a2} up 73%, Jack_{a1} up 86%, ICE and the Michaelis-Menten model up to 96% and S_{obs} above 96% (Fig. 2, intersections are accurate $\pm 1\%$).

Discussion

The evaluation of the estimators concerning their responses to differences in movement patterns is an important prerequisite for their application in faunistic studies. Furthermore, as most data-sets gathered from animal communities do not provide data on species presence-absence in sub-samples, the incidence-based forms of the non-parametric estimators often used and evaluated for plant communities are frequently not applicable. However, abundance-based versions of non-parametric estimators used in this study have not yet been evaluated. The present study appears to be the first that (1) tests the dependence of species richness estimators on movement heterogeneity and (2) compares the performance of various abundance-based estimators.

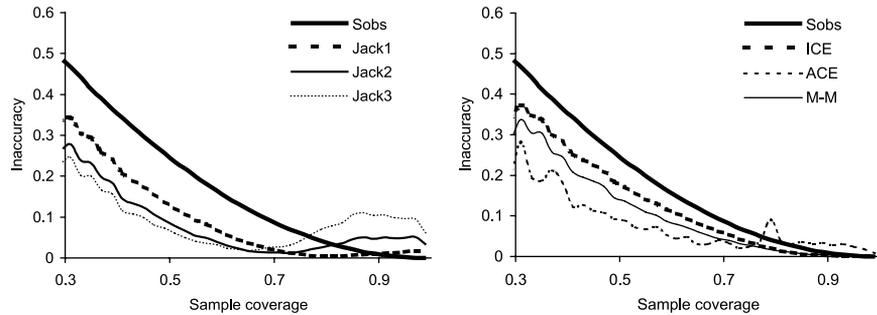
In our simulations, only the abundance-based estimator Chao1 and the incidence-based estimator Chao2 did not give precise estimates. Surprisingly, our results indicate that the incidence-based estimators that require less information than their abundance-based equivalents are in most cases less biased. Still, the differences between the abundance-based jackknife estimators and their incidence-based equivalents were only marginal.

Table 2. Simple Spearman rank and partial correlations among estimator inaccuracy and the explanatory variables – S_{true} , sampling intensity, movement heterogeneity and sample coverage.

	Simple correlations				Partial correlations		
	S_{true}	Sampling intensity	Movement heterog.	Sample coverage	S_{true}	Sampling intensity	Movement heterog.
Jack _{a1}	0.161***	-0.703***	-0.098***	-0.809***	-0.156***	0.068***	-0.045***
Jack _{a2}	0.112***	-0.139***	-0.020**	-0.238***	-0.021***	0.182***	-0.044***
Jack _{a3}	0.065 ns	0.202***	-0.018***	0.094***	0.054***	0.231***	-0.028***
ICE	0.199***	-0.777***	0.239***	-0.965***	-0.168***	0.042***	-0.005 ns
ACE	0.095 ns	-0.388 ns	-0.076 ns	-0.479***	-0.037***	0.050***	-0.024**
M-M	0.236***	-0.683***	-0.263***	-0.942***	-0.130***	0.148***	0.053***

Notes: *** $p < 0.001$, ** $p < 0.01$, ns $p > 0.05$; in partial correlations the effect of sample coverage was removed prior to further correlation analyses; M-M = Michaelis-Menten model.

Fig. 2. Mean estimator inaccuracy in dependence on sample coverage. M-M = Michaelis-Menten model.



We caution, however, that the present evaluation was carried out under the assumption of spatial homogeneity. In a previous study (Brose et al. 2003), we demonstrated that spatially heterogeneous species' distributions do not affect incidence-based estimators. In contrast, there is evidence that abundance-based estimators might be biased under such spatial heterogeneity (Chazdon et al. 1998). Accordingly, the performance of abundance-based estimators in spatially heterogeneous landscapes remains to be elucidated. Nevertheless, in our study, under low to intermediate sampling intensity all abundance-based estimators performed substantially better than S_{obs} . Accordingly, there is no reason to only use incidence-based estimators, especially since the present study illuminates the relative dependence abundance-based estimators on movement patterns, sampling intensity, S_{true} and sample coverage.

Movement

Although most animal species obviously move and also greatly vary in their movement, the distributions of movement within animal communities are rarely if ever quantified empirically. Therefore, we used a wide range of possible movement distributions for our simulations, but there are only minor differences in estimates among lognormal, normal and bimodal movement distributions. This consistency among a large range of different movement distributions suggests that the results of our simulation of movement heterogeneity based on lognormal distributions are representative for a range of natural communities, even if movement within the communities follows a different distribution. Furthermore, preliminary analyses show that results similar to those from different movement distributions can be obtained with different levels of movement intensity as simulated by the average number of steps moved per unit time.

In contrast to the type of movement distribution, the degree of movement heterogeneity had a strong influence on estimator accuracy. In particular, accuracy is reduced by the increased heterogeneity of detection probabilities

that strongly reduces sample coverage at low or intermediate sampling intensity. The increase in accuracy of the non-parametric estimators with increasing movement heterogeneity in highly sampled communities is most likely caused by their dependence on parameters like the number of singletons (species found in only one sub-sample) and uniques (species represented by only one individual in one total sample). Even if sample coverage is reduced by movement heterogeneity, a simultaneous increase in singletons and uniques leads to a compensation for the loss in accuracy that normally accompanies low sample coverage. However, the non-parametric estimators that were most robust concerning movement heterogeneity under low sampling intensity were the least accurate under high sampling intensity, which was most likely caused by overcompensation. In contrast to the non-parametric estimators, the Michaelis-Menten estimates did not compensate for reduced sample coverage, and they were negatively correlated with movement heterogeneity, irrespective of sampling intensity. This is consistent with conclusions that the Michaelis-Menten model implicitly assumes a community structure with highly equal detection probabilities for all species (Keating and Quinn 1998).

Our results show that the effect of movement heterogeneity on estimator accuracy is mechanistically similar to recently reported effects of uneven abundances (Brose et al. 2003). Both movement and abundance heterogeneities yield unequal detection probabilities, which reduce sample coverage and consequently cause less accurate estimations. For a given number of sample cells – “traps” – a species' detection probability is affected by its abundance and its movement. More abundant species are more likely to be detected as are highly mobile species. More specifically, a species moving across ten cells of our simulated grid per time step has a ten times higher detection probability than a species occupying just one cell per time step. The heterogeneous movement as simulated in the present study is equivalent to varying home ranges of species in empirical studies (Kelt and Van Vuren 2001). Interestingly, it is even possible that heterogeneous movement might increase estimator

accuracy when there is a trade-off between species' abundances and their movement causing the least abundant species to move most. For instance, it might be reasonable that larger species have larger home ranges and lower abundances. Another possible trade-off, however, is that the least abundant species move least, which might be typical for species whose movement depends on their ecological amplitude on abiotic gradients. In this case, stenotopic species that are specialized on specific abiotic conditions might be less abundant and move less than eurytopic species. The second trade-off combines the negative effects of uneven abundances and heterogeneous movement on estimator performance yielding detection probabilities that are less equal than in plant communities with the same evenness of abundances. Given that there is no clear empirical corroboration for any of the above trade-offs between abundances and movement, we randomly assigned the movement abilities to the species. Our results indicate that the frequent use of S_{obs} or the Michaelis-Menten model in faunistic studies provides unreliable estimates of S_{true} under low or intermediate sampling intensity. Being less dependent on equal detection probabilities, some of the previously unevaluated, abundance-based non-parametric estimators may yield more accurate results.

Sample coverage

Our results suggest that the ranking of the estimators concerning their accuracy changes with sample coverage, which is consistent with a previous study on incidence-based estimators of immobile species (Brose et al. 2003). Although in both studies the jackknife estimators performed best along most of the sample coverage axis, the abundance-based estimators' curves yielded different intersections than their incidence-based equivalents. Therefore, the decision framework developed for incidence-based estimators (Brose et al. 2003) cannot be directly transferred to abundance-based estimators. Due to the slightly more biased performance of the abundance-based jackknife estimators, ICE and the Michaelis-Menten model were most accurate between 86 and 96% sample coverage. The similar performance of the Michaelis-Menten model and the estimator ICE corroborate the results of Longino et al. (2002) in their study of a tropical ant fauna. Our results, however, suggest that under lower sample coverage other estimators like $Jack_a2$ and $Jack_a1$ that were not included in their study might provide more accurate results. In contrast to the present study, the incidence-based first order jackknife estimator of sessile species was less biased and more accurate than ICE under sample coverage of 86 to 96% (Brose et al. 2003). We could not confirm this finding in the present study and the incidence-based first order jackknife estimator performed similar as its

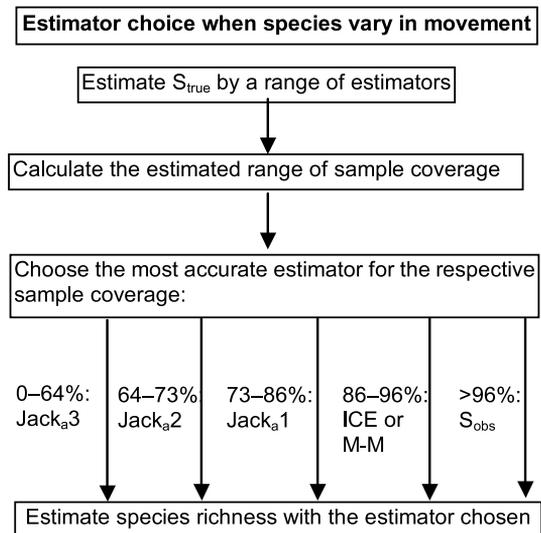


Fig. 3. Iterative process to choose the most appropriate estimator when species vary in movement. S_{obs} = observed species richness; M-M = Michaelis-Menten model.

abundance-based equivalent (unpubl.). These findings are consistent with Chazdon et al. (1998) who concluded that, with increasing sampling intensity, the number of uniques and singletons become more similar and, therefore, the difference between abundance- and incidence-based estimators vanishes. Still, our results indicate an additional influence of movement on the relative performance of the estimators. This influence is most likely mediated by effects on parameters like the number of uniques and singletons. This impact of movement suggests that estimating faunal species richness requires a distinct estimator choice framework that takes into account movement heterogeneity. In an approach similar to the one presented for sessile species (Brose et al. 2003), we suggest basing estimator choice on sample coverage. As sample coverage can only be estimated, the choice of an optimal estimator remains uncertain. Therefore, species richness can be iteratively estimated by using the more accurate estimators mentioned above to estimate sample coverage and then using the range of this estimated coverage to more specifically choose a single appropriate estimator (Fig. 3). While this process deserves empirical scrutiny, successful empirical tests of a very similar process suggests that the explicit framework in Fig. 3 will assist researchers and conservation ecologists to increase the accuracy of their comparative studies of mobile species richness.

Acknowledgements – We sincerely thank Jennifer Dunne, Richard J. Williams, Robert K. Colwell and Nicholas J. Gotelli for their suggestions. U.B. is supported by the German Academy of Naturalists Leopoldina by funds of the German Federal Ministry of Education and Science (BMBF-LPD 9901/8-44). N.D.M is supported by the Biocomplexity, Biological

References

- Baltanás, A. 1992. On the use of some methods for the estimation of species richness. – *Oikos* 65: 484–492.
- Boulinier, T., Nichols, J. D., Sauer, J. R. et al. 1998. Estimating species richness: the importance of heterogeneity in species detectability. – *Ecology* 79: 1018–1028.
- Brose, U. 2002. Estimating species richness of pitfall catches by non-parametric estimators. – *Pedobiologia* 46: 101–107.
- Brose, U., Martinez, N. D. and Williams, R. J. 2003. Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. – *Ecology* 84: 2364–2377.
- Burnham, K. P. and Overton, W. S. 1979. Robust estimation of population size when capture probabilities vary among animals. – *Ecology* 60: 927–936.
- Chao, A. 1984. Non-parametric estimation of the number of classes in a population. – *Scand. J. Stat.* 11: 265–270.
- Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. – *Biometrics* 43: 783–791.
- Chao, A. and Lee, S.-M. 1992. Estimating the number of classes via sample coverage. – *J. Am. Stat. Ass.* 87: 210–217.
- Chao, A., Ma, M.-C. and Yang, M. C. K. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. – *Biometrika* 80: 193–201.
- Chazdon, R. L., Colwell, R. K., Denslow, J. S. et al. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of northeast Costa Rica. – In: Dallmeier, F. and Comiskey, J. A. (eds), *Forest biodiversity research, monitoring and modeling*. The UNESCO and The Parthenon Publishing Group, pp. 285–309.
- Colwell, R. K. 1997. Estimates: statistical estimation of species richness and shared species from samples. Version 5. User's Guide and application. – Published at: <http://viceroy.eeb.uconn.edu/estimates>.
- Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – In: Hawksworth, D. L. (ed.), *Biodiversity: measurement and estimation*. R. Soc., Lond, pp. 101–118.
- Gaston, K. J. 1996. Species richness: measure and measurement. – In: Gaston, K. J. (ed.), *Biodiversity: a biology by numbers and difference*. Blackwell Science, pp. 77–113.
- Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – *Ecol. Lett.* 4: 379–391.
- Keating, K. A. and Quinn, J. F. 1998. Estimating species richness: the Michaelis-Menten model revisited. – *Oikos* 81: 411–416.
- Kelt, D. A. and Van Vuren, D. H. 2001. The ecology and macroecology of mammalian home range area. – *Am. Nat.* 157: 637–645.
- Lee, S.-M. and Chao, A. 1994. Estimating population size via sample coverage for closed capture-recapture models. – *Biometrics* 50: 88–97.
- Longino, J. T., Coddington, J. and Colwell, R. K. 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. – *Ecology* 83: 689–702.
- Martinez, N. D. 1996. Defining and measuring functional aspects of biodiversity. – In: Gaston, K. J. (ed.), *Biodiversity: a biology by numbers and difference*. Blackwell Science, pp. 114–148.
- May, R. M. 1988. How many species are there on earth? – *Science* 241: 1441–1449.
- Palmer, M. W. 1990. The estimation of species richness by extrapolation. – *Ecology* 71: 1195–1198.

- Palmer, M. W. 1991. Estimating species richness: the second-order jackknife reconsidered. – *Ecology* 72: 1512–1513.
- Prestons, F. W. 1948. The commonness and rarity of species. – *Ecology* 29: 254–283.
- Sokal, R. R. and Rohlf, F. J. 1995. *Biometry*, third ed. – Freeman and Company.
- Tokeshi, M. 1993. Species abundance patterns and community structure. – *Adv. Ecol. Res.* 24: 111–186.
- Walther, B. A. and Morand, S. 1998. Comparative performance of species richness estimation methods. – *Parasitology* 116: 395–405.
- Williams, R. J. and Martinez, N. D. 2000. Simple rules yield complex food web dynamics. – *Nature* 404: 180–183.

Appendix 1

Formulas of the estimators compared (Colwell 1997, Burnham and Overton 1979)

S_{obs} : the number of species observed

Non-parametric estimators:

Chao1: The abundance-based estimator of species richness (Chao 1984)

$$S_{\text{est}} = S_{\text{obs}} + \frac{F_1^2}{2(F_2 + 1)} - \frac{F_1 F_2}{2(F_2 + 1)^2}$$

Chao2: The incidence-based estimator of species richness (Chao 1987)

$$S_{\text{est}} = S_{\text{obs}} + \frac{Q_1^2}{2(Q_2 + 1)} - \frac{Q_1 Q_2}{2(Q_2 + 1)^2}$$

Jack1: The first order jackknife estimator (Burnham and Overton 1979)

$$S_{\text{est}} = S_{\text{obs}} + Q_1 \left(\frac{m - 1}{m} \right)$$

Jack2: The second order jackknife estimator (Burnham and Overton 1979)

$$S_{\text{est}} = S_{\text{obs}} + \left(\frac{2m - 3}{m} \right) Q_1 - \left(\frac{(m - 2)^2}{m(m - 1)} \right) Q_2$$

Jack3: The third order jackknife estimator (Burnham and Overton 1979)

$$S_{\text{est}} = S_{\text{obs}} + \left(\frac{3m - 6}{m} \right) Q_1 - \left(\frac{3m^2 - 15m + 19}{m(m - 1)} \right) Q_2 + \left(\frac{(m - 3)^3}{m(m - 1)(m - 2)} \right) Q_3$$

Jack_a1–Jack_a3: the abundance based jackknife estimators (Burnham and Overton 1979) use the same formulas as the jackknife estimators with F_i instead of Q_i .

ACE: The abundance-based coverage estimator (Chao and Lee 1992, Chao et al. 1993)

$$S_{\text{est}} = S_{\text{abund}} + \frac{S_{\text{rare}}}{C_{\text{ace}}} + \frac{F_1}{C_{\text{ace}}} \gamma_{\text{ace}}^2$$

ICE: The incidence-based coverage estimator (Lee and Chao 1994)

$$S_{\text{est}} = S_{\text{freq}} + \frac{S_{\text{inf}}}{C_{\text{ice}}} + \frac{Q_1}{C_{\text{ice}}} \gamma_{\text{ice}}^2$$

where:

S_{est} = estimated species richness

Q_j = number of species that occur in exactly j samples

F_i = Number of species that have exactly i individuals when all samples are pooled

m = total number of samples

S_{freq} = number of frequent species (each found in more than 10 samples)

S_{abund} = Number of abundant species (each with more than 10 individuals) when all samples are pooled

S_{inf} = number of infrequent species (each found in 10 or fewer samples)

S_{rare} = Number of rare species (with 10 or fewer individuals) when all samples are pooled

C_{ice} = sample incidence coverage estimator:

$$C_{\text{ice}} = 1 - (Q_1/N_{\text{inf}})$$

C_{ace} = sample coverage estimate based on abundance data: $C_{\text{ace}} = 1 - (F_1/N_{\text{rare}})$

N_{inf} = total number of occurrences of infrequent species:

$$N_{\text{inf}} = \sum_{j=1}^{10} jQ_j$$

N_{rare} = total number of occurrences of rare species:

$$N_{\text{rare}} = \sum_{i=1}^{10} iF_i$$

γ_{ice}^2 estimates the coefficient of variation of the Q_j 's:

$$\gamma_{\text{ice}}^2 = \max \left\{ \frac{S_{\text{inf}}}{C_{\text{ice}}} \frac{m_{\text{inf}}}{(m_{\text{inf}} - 1)} \frac{\sum_{j=1}^{10} j(j-1)Q_j}{N_{\text{inf}}^2} - 1, 0 \right\}$$

γ_{ace}^2 estimates the coefficient of variation of the F_j 's:

$$\gamma_{\text{ace}}^2 = \max \left\{ \frac{S_{\text{rare}}}{C_{\text{ace}}} \frac{\sum_{i=1}^{10} i(i-1)F_i}{N_{\text{rare}}(N_{\text{rare}} - 1)} - 1, 0 \right\}$$

m_{inf} = number of samples that have at least one infrequent species

Species accumulation curves:

Michaelis-Menten model: (Colwell and Coddington 1994)

$$S(n) = \frac{S_{\text{max}}n}{B + n}$$

where:

n = units of sampling

$S(n)$ = S_{obs} after n units of sampling

S_{max} = estimated total species richness (asymptote)

B = fitted constant