Acta Horti Bot. Bucurest. **29**: 375 – 382
Bucureşti 2001

NON-PARAMETRIC METHODS FOR ESTIMATING SPECIES RICHNESS: A STUDY CASE ON HERBACEOUS VEGETATION IN THE DANUBE FLOODPLAIN

ŢOPA Sorana, GHEORGHE Iuliana Florentina,
COGĂLNICEANU Dan, GHIOCA Dana, VĂDINEANU Angheluţă*

Abstract: From the different statistical approaches used to estimate species richness, this paper discusses non-parametric techniques. A comparative analysis of species richness in the herbaceous layer from the lower part of the Danube floodplain among sites was done using different estimators. We tested five estimators based on incidence data (presence/absence): *Chao 2, Jackknife 1, Jackknife 2, Bootstrap, ICE* and *species accumulation curve.* We compared the precision of the estimators, their dependence on sample size or on the distribution model and the number of samples required to reach stable values. *ICE* produced the best estimates and is recommended for this type of studies.

Keywords: species richness, Danube floodplain, herbaceous vegetation, species accumulation curve, species richness estimators.

Introduction

Estimating species diversity is extremely important for management programs of the natural capital. Knowing the difficulties of making complete inventories, reliable methods for estimating this parameter are needed as a fast and low cost alternative. The main limitation of the proposed estimators is their dependence on the sample size. Considering the three main measures of species diversity: species richness, models of species abundance and estimators/indices based on the relative abundance of species, this paper discusses the first category.

The difficulty of inventories based on repeated sampling is that no matter how many sampling units and what method is used the result will consist of a small number of species with many individuals and a large number of species with few individuals. Since a substantially incomplete study resembles very much with a substantially complete study, the question that rises is how can we appreciate the further effort and time needs to achieve a statistically defined accuracy of species richness. The methods of estimation of species richness through extrapolation can offer a possible answer to this question.

Local species richness can be estimated through extrapolation using the following three statistical approaches (Chazdon et al. 1998):

1. Extrapolation of species-accumulation curves – these curves are extrapolated either to an asymptotic value or to sample sizes or areas larger than the observed. This method requires only qualitative data (presence/absence);
2. Fitting parametric distributions of relative abundance – this class of methods fits data of the relative abundance of species in a single sample to a theoretical parametric distribution;

* University of Bucharest, Department of Systems Ecology, Splaiul Independentei 91-95, RO - 76 201 - Bucureşti, e-mail: sorana@bio.bio.unibuc.ro, tel. 4112310

3. Non-parametric methods – these estimators are based either on species incidence (presence/absence) data or on relative abundance data.

In this paper we attempt to validate the applicability of different non-parametric methods of estimation of species richness, insufficiently tested on real data series, and to test for the precision with which we estimated the species richness of the herbaceous layer in the Danube floodplain.

We focused on the following aspects:

1. The usefulness of the non-parametric approach for the estimation of species richness;
2. Which estimators best fulfil the requirements of not being sensitive to the sample size (cumulative number of quadrates) or to the spatial distribution model;
3. To estimate the number of species and the number of quadrates where the estimators reach stable values.

Materials and methods

Sampling sites and methods

The data used in this paper was obtained during a comparative study done in 1998 of the structure of the herbaceous layer, (***, 1998). The study area, is located in the lower Danube floodplain under a natural regime of flooding and is characterised by an average annual temperature of 11.2°C, and an average annual rainfall of 600-700 mm.

Two transects, along a hydric gradient perpendicular to the Danube were chosen. The first transect was on the Island of Fundu Mare (F transect with three sites: F2, F3, F4) and the other was on Popa Islet (I transect with three sites: I2, I4, I6). We begun the numbering of the sites in the vicinity of the Danube and continued inside the islands.

In each site we did floristic lists and the inventory of species using the quadrates method (0.25 m^2 area) on a monthly basis (from April to October, excepting August) for the F transect, and a seasonal frequency (April, July, October) for the I transect. A number of 10 quadrates for each site was used for estimating the species richness, except the F4 site where this number varied according to the periods of flooding. The total number of quadrates on which the species inventory relied was 60 for F2 and F3 sites and 17 for F4 at the end of the year. For the I transect sites, the number of quadrates was 30 at the end of the year.

Species accumulation curve and Coleman curve

Species accumulation curve (*SAC*) is a plot of the cumulative number of species found ($S_{(n)}$) within a defined area, as a function of some measure n of the effort to find them. In the vegetation studies, for estimating both the local species richness and the regional richness, species accumulation curves are referred to as *species-area curves* (Coleman et al., 1982). If the samples vary in species richness, then the shape of the species accumulation curve will depend strongly on the order in which samples are added. Variation in the curve shape due to accumulation order arises from real spatially and temporally heterogeneity among the samples, as well as from sampling errors. In order to eliminate these errors, the estimation of a *SAC* will be improved by repeated randomisation of sample order and calculation of a mean curve. The minimum number of randomisation needed to obtain a stable mean curve should be evaluated for individual data sets. Randomisation of the sample order makes sense even when samples have some intrinsic ordering (e.g. time or space series along a transect), as long as the samples themselves are reasonably homogeneous (Colwell & Coddington 1994). For our samples we used 50 randomisation.

One way to examine the level of heterogeneity is to compare the empirical mean of the randomised species accumulation curve with the mean of the theoretical curve expected if the individuals in all samples were pooled in a single (parental) sample and randomly assigned to the new (theoretical) samples. This theoretical curve is the *random placement curve* or *Coleman Curve*. Coleman's random placement method assumes that all $n*m$ individuals are randomly assigned to n new samples. As a note, *Coleman Curve* is not an estimator of species richness in the same sense as the other discussed estimators, as it estimates sample species richness from the observed species richness (Colwell & Coddington 1994; Coleman et al. 1982).

Non-parametric methods for estimating species richness

We selected five non-parametric estimators of species richness, based on incidence data (presence/absence), (Chazdon et al. 1998; Colwell 1997):

Chao 2: an estimator of species richness based on the number of species found in exactly one quadrate and the number of species found in exactly two quadrates:

$$S_{Chao2} = S_{obs} + \frac{Q_1^{\ 2}}{2Q_2}$$

Jackknife 1: *The first-order jackknife* estimator of species richness is based both on the number of species found in exactly one quadrate and on the total number of quadrates:

$$S_{jack1} = S_{obs} + Q_1\left(\frac{m-1}{m}\right)$$

Jackknife 2: The *second-order jackknife* estimator of species richness is based both on the number of species found in exactly one quadrate, respectively two quadrates, and on the total number of quadrates:

$$S_{jack2} = S_{obs} + \left[\frac{Q_1(2m-3)}{m} - \frac{Q_2(m-2)^2}{m(m-1)}\right]$$

Bootstrap: The *Bootstrap* estimator of species richness incidence-based takes into account the proportion of quadrates which contain each species:

$$S_{boot} = S_{obs} + \sum_{k=1}^{S_{obs}} (1-p_k)^m$$

ICE: *Incidence-based Coverage Estimator* of species richness takes into account the species found in 10 or fewer samples:

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

Where:

- S_{est} - estimated species richness, where *est* is replaced in the formula by the name of the estimator;
- S_{obs} - total number of species observed in the all pooled samples;
- S_{infr} - number of infrequent species (each found in 10 or fewer samples);
- S_{freq} - number of frequent species (each found in more than 10 samples);
- M - total number of samples;
- Q_j - number of species that have exactly j samples (Q_1 – frequency of uniques, Q_2 – frequency of duplicates);

- p_k - proportion of samples that contain species *k*;
- C_{ice} - sample incidence coverage estimator;
- γ_{ice}^2 - estimated coefficient of variation of the Q_j's for infrequent species

The software used to compute the various estimators was *EstimateS* (Colwell 1997).

Results and discussion

The estimated species richness was graphically analysed on the representation of the estimator as a function of the observed species richness and of the cumulative number of quadrates (Fig. 1). The SAC is similar to Coleman's curve for all stations, without reaching a plateau. The similarity of the two curves indicates that the samples have a homogenous species composition, and emphasizes the fact that spatial distribution model is important in species richness estimation. If the species had a patchy distribution, then the initial rate of species accumulation would be lower as new quadrates were collected. For this particular situation, we can conclude that the species richness estimation based on the extrapolation of the species accumulation curve is more sensitive to the distribution model than other estimators based on incidence.

$S_{Chao\ 2}$ depends strongly on the sample size and on the heterogeneity of the samples, and it reaches relatively stable values only at a large sample size. However, this estimator is more sensitive towards rare species.

S_{Boot}, estimated species richness in a very similar way to the observed one for all the studied sites. All the others estimators, overestimated the species richness, but without reaching the plateau.

A comparative analysis of the different estimators at each site (Fig. 2), allowed to test the influence of the sample size. We can conclude that the estimators tend to be stable only for F3 station for a value of 60 samples.

To test the accuracy of the estimators, in Table I we present the estimated species richness, the total species richness, in each site based on repeated inventories and the observed species richness in quadrates. S_{Boot} underestimates the species richness and S_{Jack1}, even if underestimating the species richness for some stations, is the nearest to the real species richness. S_{ICE} is a proper estimator of species richness, slightly overestimating the species richness.

For the F2 site we made a comparative analysis of the precision of the different estimators. We plotted the standard deviation of the various estimators (Fig. 3), and observed that S_{Jack1} has the best precision.

Conclusions

From the present study, we can conclude:

- the rate of species accumulation as a function of sample size emphasizes important characteristics of the species spatial distribution model and the degree of the association;
- species richness estimators are dependent on sample size, which limits their utility;
- rather than an attempt of standardizing the sample size (either number of samples or area covered), the estimators are an indicator of our data collection effort;
- S_{ICE} performed best and we recommended its use in vegetation studies;
- the precision of the estimators is difficult to test without having complete phytosociological data of vegetation communities.

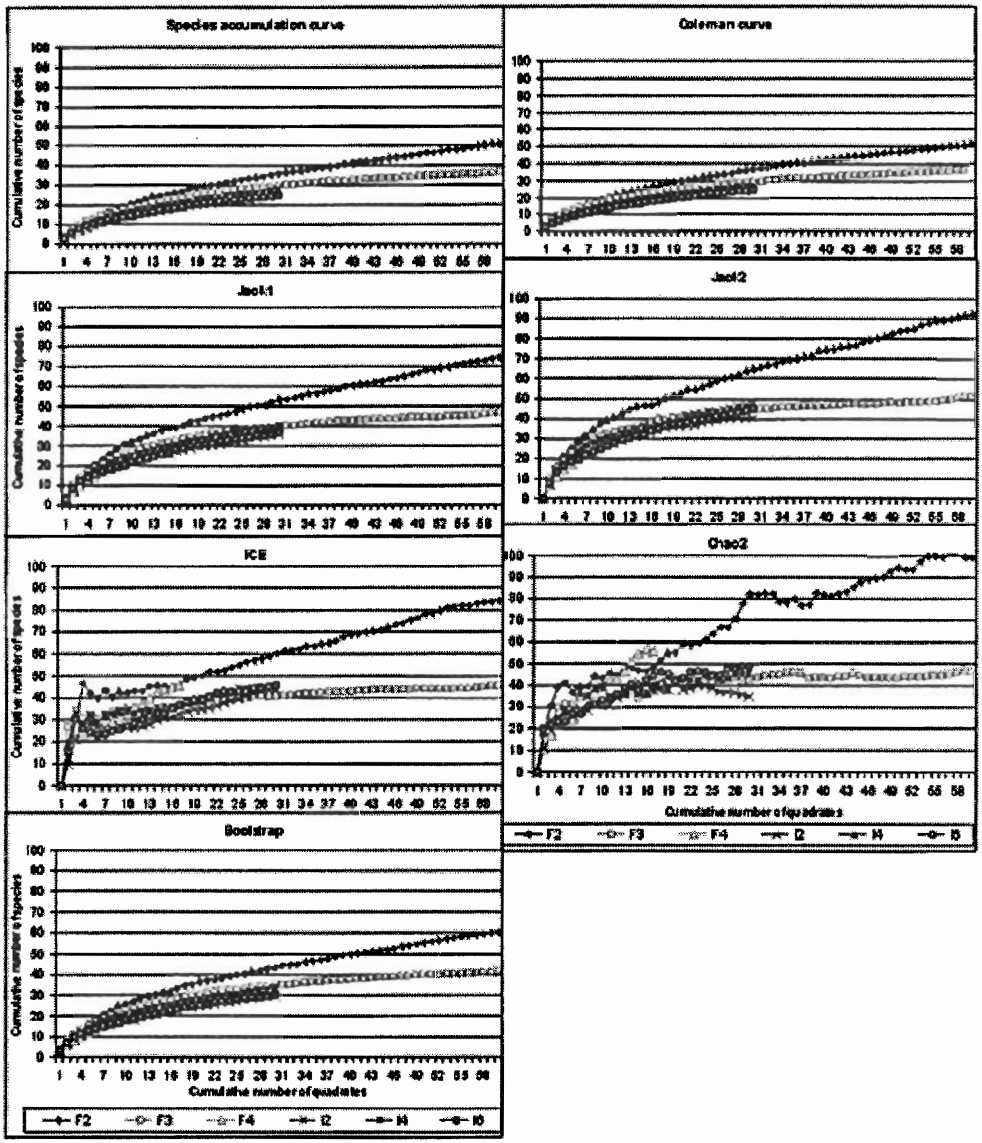

Fig. 1 Comparison of species richness of herbaceous layer among sites, using different estimators.

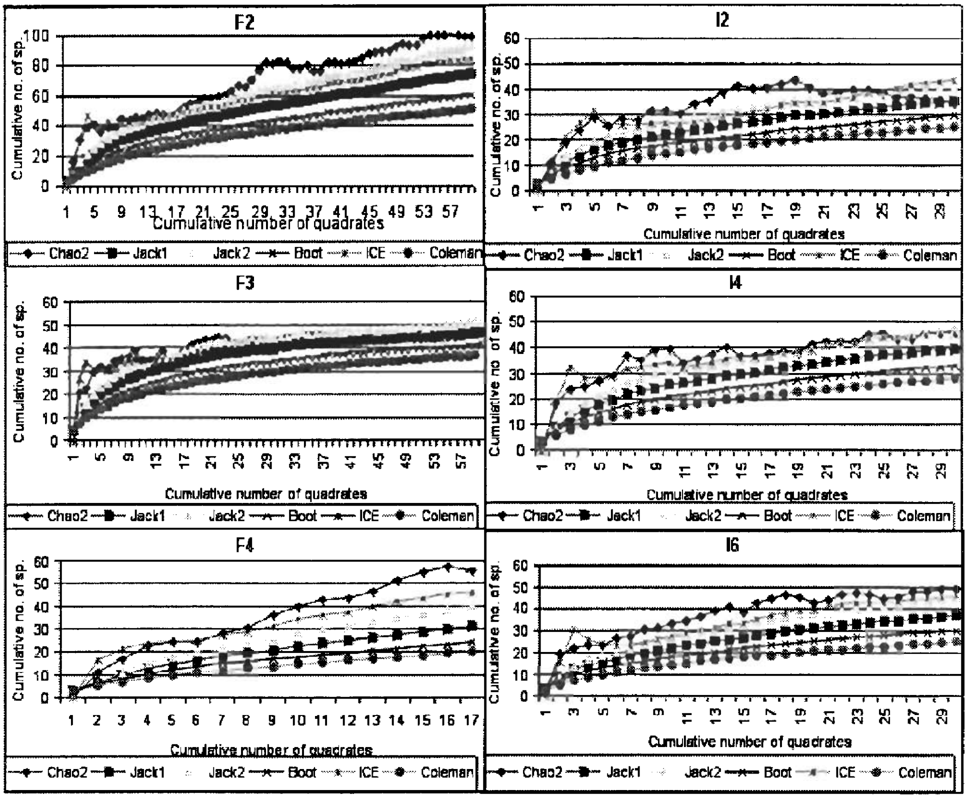

Fig. 2 Comparison of the performance of different species richness estimators for selected sites.

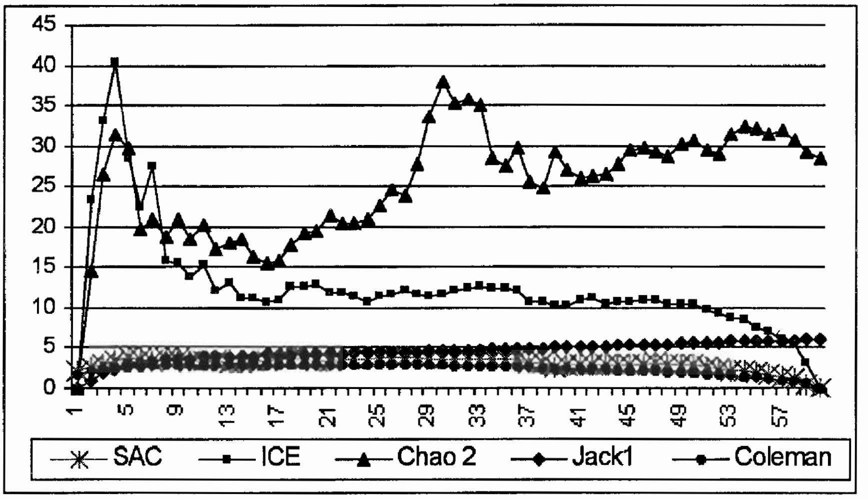

Fig. 3 Graphical representation of the standard deviations of the estimators as function of sample size (number of quadrates) for F2 site.

Table I Number of observed and estimated species and total number of species in the six sites investigated.

Site	Observed species in quadrates	Total recorded number of species	% observed species from total number of species	Estimated number of species				
				Chao 2	Jack 1	Jack 2	Boot	ICE
F2	51	81	62.96	99	75	92	61	84
F3	38	61	62.3	47	47	52	42	46
F4	20	31	64.52	56	31	40	25	46
I2	25	48	52.08	35	36	40	30	44
I4	28	36	77.78	46	40	47	33	46
I6	25	31	80.65	49	37	45	30	46

References

1. CHAZDON R. L., COLWELL R. K., DENSLOW J. S., GUARIGUATA M. R. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. Pp. 285-309. In: *Forest biodiversity research, monitoring and modeling: Conceptual background and Old World case studies.* Parthenon Publishing, Paris.
2. COLEMAN B. D., MARES M. A., WILLIG M.R. & HSIEH Y-H. 1982. Randomness, area, and species richness. *Ecology* **63**: 1121-1133.
3. 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.

4. COLWELL R. K. & CODDINGTON J. A. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B* **345**: 101-118.
5. *** 1998. Functional role of biodiversity. Scientific report. Department of Systems Ecology. Bucharest.

METODE NEPARAMETRICE PENTRU ESTIMAREA BOGĂŢIEI SPECIFICE: STUDIU DE CAZ – VEGETAŢIA IERBOASĂ DIN LUNCA INUNDABILĂ A DUNĂRII

Rezumat: Dintre abordările statistice prin care poate fi estimată bogaţia specifică, lucrarea discută tehnicile neparametrice. Pe un studiu de caz privind structura covorului vegetal ierbos din lunca inundabilă a Dunării, am realizat o analiză comparativă a bogaţiei specifice între staţii utilizând diferiţi estimatori şi o analiză comparativă a performanţelor diferiţilor estimatori. Pe lângă calcularea curbei de acumulare a speciilor, sunt prezentaţi cinci estimatori neparametrici ai bogăţiei specifice, bazaţi pe incidenţă (prezenţă/absenţă): *Chao 2, Jackknife 1, Jackknife 2, Bootstrap, ICE.* Am urmărit care dintre estimatorii prezentaţi satisface cel mai bine cerinţele unui estimator ideal al bogăţiei specifice, sensibilitatea acestora la mărimea probei sau la tipul de distribuţie, numărul de unităţi de probă la care aceştia ating valori stabile. Estimatorul *ICE* a aproximat cel mai bine bogăţia specifică pe setul nostru de date şi recomandăm utilizarea sa în astfel de studii.

Cuvinte cheie: bogăţie specifică,lunca Dunării, vegetaţie ierboasă, curba de acummulare a speciilor, estimatori specifici.