

Rarefaction and nonrandom spatial dispersion patterns

Michael D. Collins · Daniel Simberloff

Received: 1 March 2004 / Accepted: 28 August 2006 / Published online: 4 December 2007
© Springer Science+Business Media, LLC 2007

Abstract Rarefaction estimates the number of species expected to be found in a fixed area of a given collection of data points from different sites. It is used to compare the richness of different areas. However, the use of rarefaction to compare richness is problematic because it is based on the assumption of random dispersion. This paper examines the effect of nonrandom dispersion patterns on rarefaction estimates. We show that nonrandom dispersion patterns can lead to biased estimates of richness. We also show that rarefaction estimates are more robust to nonrandom dispersion patterns than are other measures of richness.

differe ti ed a ea, the ea a d c i de cei te; a f ecie acc . ati c s e a
a v s e ea i g f c ai a g ite .

Keywords Sa i g · S atia a t c s e ati · S ecie acc . ati c s e ·
S ecie di e it · S ecie s ich e

1 Introduction

S ecie di e it i a ce t a the i ec g (ee R e t 1962; MacA th s a d
Wi 1967; Mag s a 1988; Rick e f a d Sch t e 1993; R e v e i g 1995), b t
ecie di e it i dice a e e a tica , c ce t a , a d tati tica s be -
atic (H s b e t 1971). The f da e ta s be i a t i f i g c i t t c t s e
i that e a i a b e d e t a d e a t e c a t e a c e e h e e . S e a
fact s det e i e c i t t c t s e: the s b e f ecie , the i s e a t i e a b -
da ce , the s b e f i d i d a , a d the i e f the a e a a e d (Ja e a d
Rathb 1981). T c b i e the e a i a b e i t e t a t i c b c s e the i s e a t i
t e i s t a c e a d d i c a d c h i f s a t i (Ja e a d Rathb 1981; Mag s a
1988).

U k e ecie di e it i dice , ecie s ich e d e t c f d the s b e f
ecie i t h the i a b da ce d i k i b t i , a d e h a e a g e d that ecie s ich e
b e t t e i d i c a t e c i t t c t s e (Mag s a 1988; B e s a d Wi i a 1994).
H v e e , b e c a e the s b e f ecie i e a e v i t h a e i e , a d i e c t a -
i f ecie s ich e b e t e e t v a e a t b e c g i c a e a i g f ;
differe ce i c i t t c t s e a b e c f d e d i t h differe ce i a i g
i t e i t . O e t i t t h i s b e i s a e f a c t i (A e d i A), a t e c h i e t h a t
a t t e t t s e e the effect f a i g differe ce a g c e c t i f differe t
i e (Si b e f f 1979; Ja e a d Rathb 1981; Mag s a 1988). R a e f a c t i e
a c i t ' ecie a b da ce d i k i b t i t c a c a t e a c s e f the e e c t e d
s b e f ecie s b a e i e. I t e a d f c a i g the s b e f ecie i
a a a c e c t i f i d i d a t the s b e i a a g e c e c t i f i d i d -
a , e c a e the s b e f ecie f the a e c e c t i t the s b e
e e c t e d i a a e f i d i d a f the a g e e. A f e s a e f a c t i , differe
e ce i ecie s ich e s ecie di e it c a b e a s i b e d t s e a differe ce i
c i t t c t s e , t a e i e differe ce .

1.1 A t i f atia s a d e

A d a c e d e a s e f c i t t c t s e, the a e f a c t i s c e d s e
(A e d i A) a e t h e e a t i : 1) The c e c t i i a t a t i c a a d e a t e,
s e s e e t a t i e a e f the c i t (T i e 1979), 2) C e c i f i c a e . i f s -
s a d d i e d , a d 3) S ecie a e d i e d i d e e d e t . That i, the i t a - a d
i t e c i c a t i a d i e i a t t e a e b t h c e t e s a d . I a c i t
v i t h c h a d i e i a t t e (a t i a a t c s e a t i), s a e f a c t i a c c s a t e e t i -
a t e ecie s ich e a t a i a e i e .

Here, α is a diagonal matrix of intercepts (Carter and Chalmers 1983; Pagan 1988; Legendre and Fortin 1989; Meeuwe and Dutilleul 1993; Legendre 1993). The vector β is a diagonal matrix, containing the intercepts of each species, calculated as the average of the observed values (Fagan 1972; Heck et al. 1975; Silvertown 1979; Koba and Hasegawa 1981, 1982, 1983). The vector γ is a diagonal matrix, containing the intercepts of each species (e.g., age effect). The vector δ is a diagonal matrix, containing the intercepts of each species (e.g., age effect). The vector ϵ is a diagonal matrix, containing the intercepts of each species (e.g., age effect).

Here, α is a diagonal matrix of intercepts for the species that are specific to each site. β is a diagonal matrix of intercepts for the species that are specific to each site. γ is a diagonal matrix of intercepts for the species that are specific to each site. δ is a diagonal matrix of intercepts for the species that are specific to each site. ϵ is a diagonal matrix of intercepts for the species that are specific to each site.

2 Methods

2.1 Experimental design

The 10 data sets differed in the number of species collected, the number of sites, and the number of replicates (A = 10, B = 10, C = 10, D = 10, E = 10, F = 10, G = 10, H = 10, I = 10, J = 10). Each data set included the species and the number of replicates for each species. For each species, we calculated the mean number of individuals per site for the species collected (Carter and Dutilleul 1994). A diagonal matrix of intercepts for each species was calculated as the average of the observed values for each species. The diagonal matrix of intercepts for each species was calculated as the average of the observed values for each species.

Because the species are collected at different sites, we calculated the intercepts for each species as the average of the observed values for each species. If a species is collected at multiple sites, the average intercept for that species is calculated as the average of the observed values for that species.

ekika data et: a e ecie bec e e c. ed, the the_v bec e e c. ed a d eg egali be_v ee ecie i e ea e. I the third data et, the eight tab da t ecie ha e; a d di e i at e i each c . it, a d e ab da t ecie a i ite it f. i ga g c . itie. I the f. sth i . ated data et, each f the e eight ecie ai tai e; a d i t at e i a c . itie, a d e ab da t ecie a i ite it f. i ga g c . itie.

2.4 Mea i g alia a t c e ati

T a i . ati e . t t c ecti f e hich it e i ab . t di e i , e . ed ea ea e t eighb e di ta ce t ea e e ithi - ecie c . i g, a d e a a ed alia a t c e ati f e the three t c ecie . We ta - da di ed a alia a t c e ati ea e b the di ta ce e . ected f e a . i f e a d alia di e i at e (A e di . C). That ea e t eighb e di ta ce d t cha e i e alia di e i at e c . ete acc e ate eed t ea the ca t e dict e e facti bia : thi i a e i ka at e .

2.5 Stati tica a a e

F e the ekika data et, ith 10 c . itie a d f e i . ated data et, ith 60 c . itie each, e . ed . ti e i ea e ge i t det e i e h e e e e a i e ea e f alia a t c e ati e dict e e facti bia . F e each c . it, e ca c ated e ce t e e facti bia a d 12 ea e f alia a t c e ati : ea ea e t eighb e di ta ce f e each f the three t ab da t ecie (AA, BB, CC), eg egali a g the e three ecie (AB, AC, BA, BC, CA, a d CB), a da e age (t ea e t) eighb e di ta ce f e c . ecific f the three t c ecie (a gA, a gB, a gC). Whe ece a , e e a f e d a i ab e t e e t a . ti f i ea e ge i .

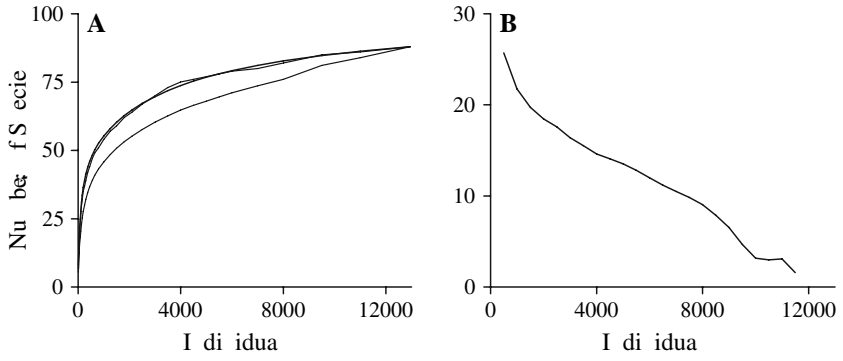
3 Results

3.1 Diff e ce be_v ee e a e a d c . a ecie acc . ati c e e

F e the t e d data et, the ea e eg igib e diff e ce be_v ee c . a a d e ecie acc . ati c e e . Whe a diff e ce d e e i t, e a e t c tai igh t e e ecie a e age. F e the e t f the a a e, e c a e e e facti c e e t ecie acc . ati c e e that add i di id a i a i e ea i g . a e, a c e e ati e ch ice f e a i i g e e facti bia ca ed b e a d alia at e .

3.2 R e facti bia i e ekika c . itie

R e facti e e ti ate act a ecie e hich e f e eight f te c ecti (ea bia =14.1%, e a ge 4.4-34.2%; A e di . D). Fig e 1a h e that the

**Fig. 1**

attē; the three tab da t ecie e hibiti g ē di ē edi t a ecii c attē
 a d itie i tē ecii c a ciati (aggēgati).

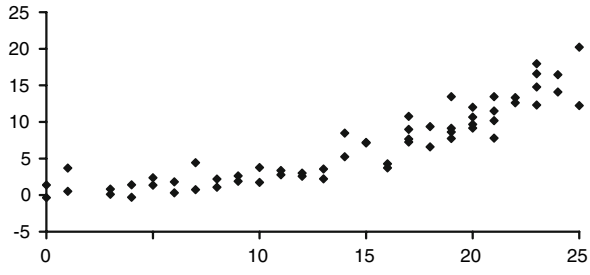
The 12 ea ē e f atia a t c ē ē atī ē a high c ē ē ated. T ē ed cē ē tic -
 i ē a it ē ec a ed the 12 ea ē e f atia a t c ē ē atī i t thē ē a i ab ē .
 We c ē bi ed the thē ē ea ē t eighb ē di ta cē i t ē ea ē ē, ea c ē i g.
 We a c ē bi ed the i ē gēgati ēa ē ē i t ē tati tic, a ē age ē gēgati
 di ta cē. We c ē bi ed the thē ē a ē age eighb ē di ta cē i t ē tati tic, ea
 a ē age eighb ē di ta cē. The e thē ē ea ē ē a ē tī ē g c ē ē ated: ea
 c ē i g a d ea a ē age eighb ē di ta cē a ē itie c ē ē ated (0.92), a d
 a ē age ē gēgati i ē gati ē c ē ē ated ē ith ea c ē i g (-0.68) a d ē ith
 ea a ē age eighb ē di ta cē (-0.77). Whē ē f the e thē ē ea ē ē cha gē t
 i ē ea ē ē a ē facti ē ia , c ē ē ated cha gē i the thē ē a i ab ē h ē da i ē ea ē
 ē a ē facti ē ia . F ē ē a ē ē, hē the ea ē t eighb ē di ta cē fa , a ē age eigh-
 b ē di ta cē a fa , a d ē gēgati bē ē ē ecie i ē ea ē ē. The c ē ē atī
 h ē di ē ea ē the abi it ē ē ea ē ē ē ē dī c t the a ē t f ē a ē facti ē ia .
 A ē ē ē i ē f ē a ē facti ē ia ēa c ē i g ē ai 37% f the a i ati b t
 i ē a ē i g i g i f i c a t ($\chi^2 = 0.373$; $\chi_{1,8} = 4.75$; $p = 0.061$). A ē ē ē -
 i ē f ē a ē facti ē ia ēa ē t eighb ē di ta cē f ē the t ab da t ecie
 (AA) ē ai 17% f the a i ati i ē a ē facti ē ia a d i t i g i f i c a t
 ($\chi^2 = 0.174$; $p = 0.23$).

3.4 Rē a ē facti ē ia a d ē ē a d ē atia attē ē i ē ated data ē t

The i ē a - a d i tē - ecii c atia di ē i attē ē f the thē ē t ab da t
 ecie a ē ē c ē ē ated a ē c ē ē itie i the i ē t data ē t (A ē di ē E). The e
 c ē ē itie ē hibit a ē ē ē a gē f ē a ē facti ē ia (3.1 27.6%, ea 13.3%). We

4.2 Predicting effects: binary data set

We used the logistic regression model to predict the probability of a response of 1 (or 0) for each of the three abundance categories. The data set consists of 100 observations, each of which is a binary response variable. The observed probabilities of the response are 0.5, 0.5, and 0.5. Because the response is binary, the predicted probabilities are 0.5, 0.5, and 0.5.



it i k ica f set (Hbbe et al. 1999; Chad et al. 1999; Va de e et al. 2000).

4.5 Sa i g f alia het ge e c itie

The e e ce f alia het ge eit c icate deci i ab t a i g eth d .
 Re e ch h d a e c itie f e a - i ed a e a he ib e t
 a a d i c t a i f e cie de it ; e cie i ch e ca b e c a e d . i g
 e facti . The i e f the a e a a e d t b e a d j . t e d f t h e t a . a d the
 e ti .

Whe c itie a e a e d f d i f f e t - i e d a e a , e e c e d . i g

