

## ESTIMATING PARAMETERS OF NEUTRAL COMMUNITIES: FROM ONE SINGLE LARGE TO SEVERAL SMALL SAMPLES

FRANÇOIS MUNOZ,<sup>1,4</sup> PIERRE COUTERON,<sup>1,2</sup> B. R. RAMESH,<sup>1</sup> AND RAMPAL S. ETIENNE<sup>3</sup>

<sup>1</sup>French Institute of Pondicherry (IFP), 11 Saint Louis Street, Pondicherry 605001, India

<sup>2</sup>IRD, UMR AMAP, Boulevard de la Lironde, TA A51/PS2, F-34398 Montpellier Cedex 5, France

<sup>3</sup>Community and Conservation Ecology Group, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

**Abstract.** The neutral theory of S. P. Hubbell postulates a two-scale hierarchical framework consisting of a metacommunity following the speciation–drift equilibrium characterized by the “biodiversity number”  $\theta$ , and local communities following the migration–drift equilibrium characterized by the “migration rate”  $m$  (or the “fundamental dispersal number”  $I$ ). While Etienne’s sampling formula allows simultaneous estimation of  $\theta$  and  $m$  from a single sample of a local community, its applicability to a network of (rather small) samples is questionable. We define here an alternative two-stage approach estimating  $\theta$  from an adequate subset of the individuals sampled in the field (using Ewens’ sampling formula) and  $m$  from community samples (using Etienne’s sampling formula). We compare its results with the simultaneous estimation of  $\theta$  and  $m$  (one-stage estimation), for simulated neutral samples and for 50 1-ha plots of evergreen forest in South India. The one-stage approach exhibits problems of bias and of poor differentiability between high- $\theta$ , low- $m$  and low- $\theta$ , high- $m$  solution domains. Conversely, the two-stage approach yielded reasonable estimates and is to be preferred when several small, scattered plots are available instead of a single large one.

**Key words:** community ecology; dispersal limitation; Etienne’s sampling formula; Ewens’ sampling formula; India; neutral theory; sampling; wet evergreen tropical forest.

### INTRODUCTION

The “unified neutral theory” of Hubbell (2001) has become a classic theory by translating concepts of neutral stochastic models from population genetics into the field of community ecology, thereby providing a long-awaited baseline against which diversity features of real communities can be compared (Holyoak et al. 2006). Conceptually, Hubbell (2001) developed the neutral theory in agreement with the principle of ecological hierarchy (Allen and Starr 1982), by distinguishing two major, nested, spatial and temporal scales: the metacommunity and the local community.

The metacommunity is assumed to feature a very large, say billion- to trillion-scale, number of individuals, and to experience biogeographical processes. In neutral theory, its composition is supposed to reflect an equilibrium between speciation and random extinction, analogous to the classical mutation–drift equilibrium of neutral alleles in population genetics, so that the corresponding theoretical framework provided by Ewens (1972) and Karlin and McGregor (1972) proved directly applicable (Hubbell 2001, Etienne and Olff 2004). In particular, both equilibria are characterized by a single parameter, the “fundamental biodiversity

number”  $\theta$ , which is approximately, in community ecology, the product of the metacommunity size times the rate of speciation events (Hubbell 2001, reassessed by Vallade and Houchmandzadeh 2003) and is closely related to Fisher’s  $\alpha$ .

At the lower level of the local community, speciation events are supposed negligible while ecological processes, such as dispersal limitation, are essential for community structure. Under the neutral theory, all species are assumed of equivalent fitness (Hubbell 2001, 2006) and specific composition of a local community is shaped by a lottery, replacing dead individuals by either migrants from the metacommunity or offspring of resident individuals. Thereby, species abundances in a given local community are ruled by the “fundamental dispersal number”  $I$  (Etienne 2005, Etienne and Alonso 2005, 2007), which is the number of migrants available for competition at each mortality event (which can be chosen as the reference time step; Hubbell 2001). This concept basically refers to the island model in population genetics (Wright 1931, Slatkin 1985), with overlapping generations (Moran 1958). Since migrants are drawn from the metacommunity, local species abundances are under the joint control of  $\theta$  and  $I$ , and recent theoretical efforts have achieved analytical formulas for their probability distribution (Vallade and Houchmandzadeh 2003, Alonso and McKane 2004, Etienne and Olff 2004, Etienne 2005). This paved the way to an inference method (Etienne 2005), based on an exact, analytical

Manuscript received 10 January 2007; revised 23 March 2007; accepted 26 March 2007; final version received 19 April 2007. Corresponding Editor: N. J. Gotelli.

<sup>4</sup> E-mail: francois.m@ifindia.org

expression of the likelihood of the species abundances, which allows simultaneous estimation of  $\theta$  and  $I$  from a single sample of a local community.

Etienne (2005) thus estimated neutral parameters for tree species in a single, large, 50-ha forest plot in Barro Colorado Island (Central America). Etienne's method has also been applied to a set of distinct small plots in South Africa fynbos and Amazonia (Latimer et al. 2005, Etienne et al. 2006), by lumping and treating distant plots as a single sample of a community. Such lumping may not be warranted if the distance between plots is much larger than the average dispersal distance. In fact, the different ways through which recent theoretical breakthroughs can be applied to multi-plot sampling designs have not yet been thoroughly considered and compared (but see Etienne 2007). This is all the more unsatisfactory because such a sampling scheme is genuinely consistent with the two-scale neutral theory of Hubbell, who furthermore advocated (2001:144, in reference to a data set from Belize) a kind of two-stage approach where the metacommunity parameter  $\theta$  is estimated from lumped plots and the migration parameter,  $I$ , is subsequently assessed at plot scale.

The present paper is the first attempt to investigate the main options that are conceivable for estimating the neutral community parameters,  $\theta$  and  $I$ , from a multi-plot sampling design, and to explore and discuss their relative pertinence and reliability. Its objective is twofold. First, we aim to provide caveats against a direct, simplistic application of Etienne's sampling formula to data sets that do not meet the fundamental assumptions of the method. Second, we provide a new, two-stage method, where  $\theta$  is estimated using a subset of individuals from all community samples, while  $I$  is subsequently estimated from each complete community sample. For both objectives, we shall first refer to simulations of neutral local communities before applying alternative approaches to a complex, unpublished data set (featuring 50 1-ha plots) from the tropical evergreen forests of the Western Ghats mountains, in southern India.

#### METHODOLOGICAL BACKGROUND

##### *Simplistic applications of Etienne's sampling formula*

Hubbell (2001) presented a first sampling formula, formerly formulated by W. J. Ewens (1972) in a population genetics context, to estimate  $\theta$  when there is no dispersal limitation. Etienne (2005) further provided the exact likelihood  $P(\mathbf{d}/\theta, I, J)$  of a community sample, hereafter called Etienne's sampling formula, i.e., the probability of observing its vector of species abundances  $\mathbf{d}$  according to parameters  $I$  and  $\theta$  and the number of individuals in the sample,  $J$ . The underlying reasoning assumes that, in a local community, the probability of having an offspring colonizing an empty space is the same for all established individuals ("homogeneity"). This means that dispersal limitation

occurs between local communities but not within (Etienne and Alonso 2005).

When data are available for a network of rather small plots, which may belong to several distinct local communities, two direct, yet questionable, ways of applying Etienne's formula suggest themselves: (1) estimating  $\theta$  and  $I$  for each individual sample  $i$  featuring the species abundance distribution  $\mathbf{d}_i$ , using the likelihood  $P(\mathbf{d}_i/\theta_i, I_i, J_i)$ , then comparing results across samples; (2) lumping individual samples to obtain a single sample, which is treated as a single local community sample (Latimer et al. 2005, Etienne et al. 2006).

The first approach calls for no specific assumption on possible relationships between samples, but the robustness of parameter estimation can be questioned since the estimation process does not integrate the information that the samples belong to the same metacommunity. With the second approach, the lumped sample may not be homogenous if individual samples come from distinct local communities, which is likely to be the case if samples are located far away from each other and if dispersal limitation is strong. For this reason, we shall not consider it.

Very recently, Etienne (2007) provided an analytical formula that generalizes Etienne's (2005) sampling formula for multiple samples from the same metacommunity, but a tractable algorithm to compute this formula has so far only been found for the case where all samples are assumed to have the same  $I$  value.

##### *A new two-stage approach*

To circumvent the problems that are likely to arise with such simplistic applications of Etienne's sampling formula to networks of small samples, we have designed a new two-stage approach, consistent with the hierarchical principle of the Hubbell's neutral model. Apart from neutrality itself, the only hypothesis involved in this approach is that an acceptable random sample of the metacommunity may be obtained from the set of available community samples. To ensure this, taking all the individuals in a lumped sample (lumping procedure) may not be a satisfactory approach, since spatially aggregated species abundances at the community level would make the lumped sample significantly different from a random metacommunity sample. To limit this problem and prevent bias in the estimation process, it would be preferable to be closer to an ideal metacommunity sample by randomly selecting only one individual from each of the community samples (resampling procedure), in order to be insensitive to local spatial aggregation on the basis of samples of a far smaller size.

The Ewens' sampling formula (1972), which expresses the probability conditional to  $\theta$  of observing  $k$  species in a random metacommunity sample of  $N$  individuals, and is a special case of Etienne's sampling formula (namely for infinite  $I$ ), is then used to get an estimate of  $\theta$  from the total number of species found in a metacommunity

sample. The first-stage maximum likelihood estimate,  $\hat{\theta}_{\text{meta}}$ , is thus the value satisfying

$$k = \sum_{j=1}^N \frac{\hat{\theta}_{\text{meta}}}{\hat{\theta}_{\text{meta}} + j - 1}$$

(Ewens 1972). Using the resampling procedure, we randomly draw one individual from each of the  $N$  community samples belonging to the same metacommunity, so as to apply the Ewens' sampling formula to the  $N$ -individual metacommunity sample and to estimate  $\hat{\theta}_{\text{meta}}$ . We repeat this procedure many times and average the results in order to refine the estimate  $E(\hat{\theta}_{\text{meta}})$  and to gain insights on its standard error. This repetition thus allows gaining precision on the estimation, despite the small size of each metacommunity sample. On the contrary, when using the lumping procedure, a single estimate value  $\hat{\theta}_{\text{meta}}$  is deduced from the lumped data set with no information on its standard error.

The second stage is the estimation of  $I_i$  in each local community sample  $i$ , using Etienne's sampling formula with  $\theta$  set to  $E(\hat{\theta}_{\text{meta}})$ . We used Mathematica 5.2 (Wolfram Research, Champaign, Illinois, USA) and Matlab 7.0 (Mathworks, Natick, Massachusetts, USA) to perform maximum likelihood estimation.

In the sequel, we will thus investigate the performance of two alternative methods for estimating neutral parameters from a network of small community samples: (1) naively using the Etienne's sampling formula in its basic sense, that is simultaneously estimating  $\theta$  and  $I$  in each community sample ("one-stage" approach); (2) using the new two-stage, two-scale approach, to first estimate  $\theta$  from a metacommunity sample, then  $I$  from each community sample.

#### MATERIAL FOR COMPARISON

##### *Sampling tree communities in the Western Ghats*

We analyzed 50, 1-ha, tropical forest plots established in the north part of the Western Ghats mountain range, belonging to the Karnataka State, south India (Ramesh and Swaminath 1999). Plots are laid within a study area of  $51 \times 174$  km, and the average interplot distance is large (68 km), compared to the decametric to hectometric range of propagule dispersal, which can be expected for tropical wet forest trees. For analyses, we selected trees whose diameter at breast height (dbh) was above 10 cm, in agreement with Hubbell (2001), i.e., a total of 19 555 trees belonging to 304 species (45.2 species and 391.1 individuals per plot, on average). A table featuring species abundances in the forest plots is provided in the Supplement. Plots are all situated on the windward slope of the Ghats exposed to the southwest monsoon (from June to November). Average annual rainfall (4112 mm) and length of the dry season (five to seven months) are within the boundaries used to define the wet evergreen type (Pascal 1988). Indeed, in any of the plots, more than 50% of the individuals belong to

evergreen tree species (e.g., an emblematic species, *Saraca asoca* (Roxb.) Wild. [see Plate 1]).

##### *Simulating neutral community samples*

We also simulated neutral community samples, each featuring  $J$  individuals, to investigate the reliability of the two alternative methods. In the following, in addition to  $I$ , we shall also express dispersal via the migration rate:

$$m = \frac{I}{I + J - 1}$$

(Etienne and Olff 2004, Etienne 2005), which is widely used and more convenient, though it also depends on the sample size. It goes to one when  $I$  goes to infinity (infinite migrant pool, no migration limitation), i.e., when local communities are nothing but random samples from the metacommunity.

In agreement with the hierarchical vision of the neutral theory, we first generated an ancestor pool as a large metacommunity sample of  $50 \times J$  individuals, with fixed  $\theta$  and  $m = 1$ , i.e.,  $I = \infty$ , using Etienne's algorithm (2005: Appendix S2). Then we wrote, using the Matlab language and environment, a modified implementation of Etienne's algorithm to simulate  $N$  local community samples of size  $J$ , with known migration parameters ( $m < 1$ ), and for which all ancestors belong to the predefined ancestor pool. The principle of the algorithm is detailed in the Appendix. This simulation approach is one of the simplest ways to introduce dependence between simulated samples, via a common migrant pool originating in a given metacommunity.

We simulated small community samples, each featuring  $J = 400$  individuals, for which we varied the migration parameter,  $I$ , from 10 to 900 with an increment of 10, hence letting  $m$  vary from 0.02 to 0.69. The biodiversity number was set to  $\theta = 50$ , a value comparable with the diversity level one can expect for semi-evergreen or evergreen tropical forests (Hubbell 2001:147) and also for our real case study in the Western Ghats of India (see *Results: Estimation on forest plots*). For each value of  $I$ , we simulated 50 replicate community samples, which all originate in a common pool of migrants from a given metacommunity. Therefore we dealt with an overall set of 4500 community samples.

#### RESULTS

##### *Estimation on small, simulated samples*

Using the one-stage Etienne's sampling formula, we estimated  $\hat{m}_i$  and  $\hat{\theta}_i$  for each of the 4500 simulated neutral community samples. This obviously does not integrate the relationship existing between the 50 replicate community samples, generated for each value of  $I$ , through the common origin of their migrant ancestors. We compared estimates  $\hat{m}_i$  and  $\hat{\theta}_i$  to expected values, i.e., varying  $m_i$  and constant  $\theta_i = 50$ . We specifically exemplified on Fig. 1 the results for two

expected  $I_i/m_i$  values, namely  $I_i = 50/m_i = 0.11$  and  $I_i = 400/m_i = 0.5$ , and noticed a high variation in both estimates  $\hat{m}_i$  and  $\hat{\theta}_i$ , according to some hyperbolic trend in the  $(\hat{\theta}_i, \hat{m}_i)$  space (Fig. 1), which proved to be fairly consistent with the theoretical relationship linking  $\theta_i$  and  $m_i$  (Etienne 2005) through the Simpson diversity,  $d$  (continuous line, Fig. 1). We furthermore highlighted an overall bias with underestimation of  $\hat{m}_i$  and overestimation of  $\hat{\theta}_i$ , when varying  $m_i$  (Fig. 2a). In fact, although  $\theta_i$  was kept equal to 50 in all the simulations, we found an average  $\hat{\theta}_i$  value of 675.6 over all simulations, with a large standard deviation of 1113, which suggested a high bias and a high instability in the estimation process.

The two-stage approach then yielded far more consistent results than the one-stage estimation of  $\hat{\theta}_i$  and  $\hat{m}_i$  in each community sample. In fact, for any expected  $I_i/m_i$  value, 50 community samples have been generated as belonging to the same metacommunity, in a way which is fully consistent with our hierarchical approach of parameter estimation. Using the resampling procedure, the first-stage estimation provided  $\hat{\theta}_{\text{meta}}$  values averaging to  $E(\hat{\theta}_{\text{meta}}) = 50.77$ . (Averaging was done over the 90 expected  $I_i/m_i$  values and, for each of them, over 100 resampled, 50-individual metacommunity samples, hence globally over 9000 estimated values.) Such an estimate was close to the expected value of  $\theta_i = 50$  (Fig. 2b, right), with a standard error of 0.75. The lumping procedure yielded a comparable value of  $\hat{\theta}_{\text{meta}} = 49.80$ , on average over the expected  $I_i/m_i$  values. The very small discrepancy between results obtained through resampling and lumping (50.77 vs. 49.80) was to be expected because there is no inter-individual spatial dependence in our simulation model, contrary to what is likely to occur in most of the real-world data sets. At the second stage,  $\hat{m}_i$  estimates were also far more consistent than the ones given by the one-stage approach (Fig. 2b, left).

*Estimation on forest plots*

We used the Etienne’s sampling formula to estimate  $\hat{I}_i/\hat{m}_i$  and  $\hat{\theta}_i$  in each of the 50 forest plots (one-stage estimation). This provided small  $\hat{\theta}_i$  values (15.53 on average, with standard error of 1.32), while corresponding  $\hat{I}_i$  were very high, resulting in  $\hat{m}_i$  values close to 1 (0.921 on average, with standard error 0.03), which would mean the absence of dispersal limitation.

We alternatively used the two-stage approach to estimate neutral parameters for these forest plots, which can be considered as belonging to the same metacommunity, in the northern Western Ghats. Using the resampling procedure, we estimated  $\hat{\theta}_{\text{meta}}$  for each of 1000 metacommunity samples featuring 50 individuals, one randomly drawn from each forest plot, before computing the average  $E(\hat{\theta}_{\text{meta}}) = 62.3$ , while the alternative lumping procedure yielded a value substantially lower of  $\hat{\theta}_{\text{meta}} = 50.99$ . Using Etienne’s sampling formula, we subsequently estimated  $\hat{I}_i$  in each plot, given the overall fixed  $E(\hat{\theta}_i) = 62.3$  obtained from the

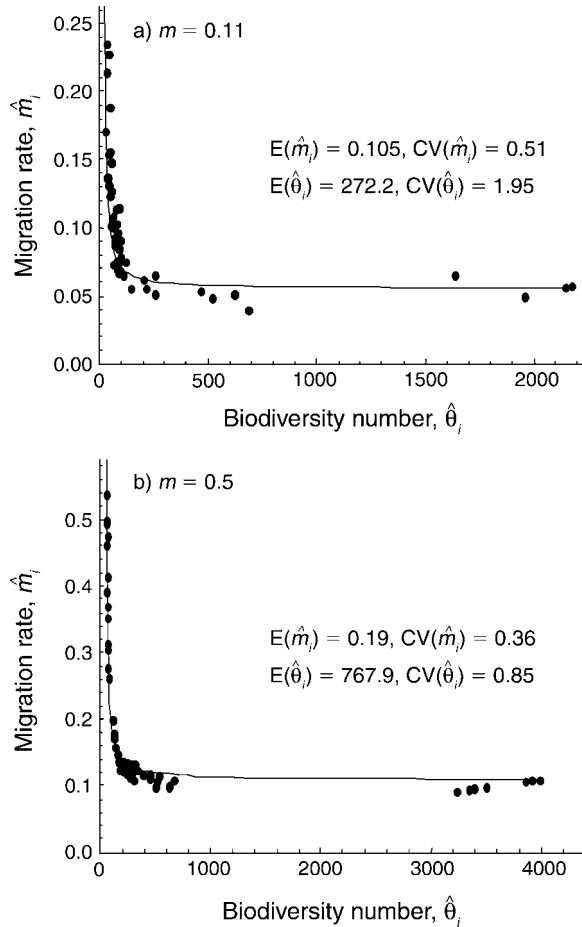


FIG. 1. One-stage estimation of  $\hat{\theta}_i$  (biodiversity number) and  $\hat{m}_i$  (migration rate) from the simulated community samples  $i$ , (a) with  $J = 400$  (number of individuals in the sample),  $\theta = 50$ , and  $I = 50$  (where  $I$  is the fundamental dispersal number) (i.e.,  $m \approx 0.11$ ), or (b) with  $J = 400$ ,  $\theta = 50$ , and  $I = 400$  ( $m \approx 0.5$ ). Each point represents the one-stage estimates  $(\hat{\theta}_i, \hat{m}_i)$ , using Etienne’s sampling formula for each of the 50 community samples  $i$  belonging to a same metacommunity, through a common pool of migrants. The mean values ( $E$ ) and coefficients of variation ( $CV$ ) over the 50 samples are given for  $\hat{m}_i$  and  $\hat{\theta}_i$ . The continuous line is the predicted relationship between  $\theta$  and  $m$ , when the Simpson diversity  $d$  is set to its mean over the 50 samples: (a)  $\bar{d} = 0.9587$  ( $SE = 0.00095$ ), (b)  $\bar{d} = 0.9798$  ( $SE = 0.00027$ ).

resampling procedure. Second-stage estimation yielded  $\hat{m}_i$  values averaging 0.055 (with standard error of 0.003).

DISCUSSION

We investigated two options that are easily conceivable to estimate neutral community parameters from a set of sampling plots, with special emphasis on the issue of their dependence through the metacommunity, which is the common source of migrants. The naive approach consisting of a direct application of Etienne’s sampling formula to small community samples, without accounting for such dependence, resulted in biased estimates of neutral parameters  $\theta$  and  $I$ .

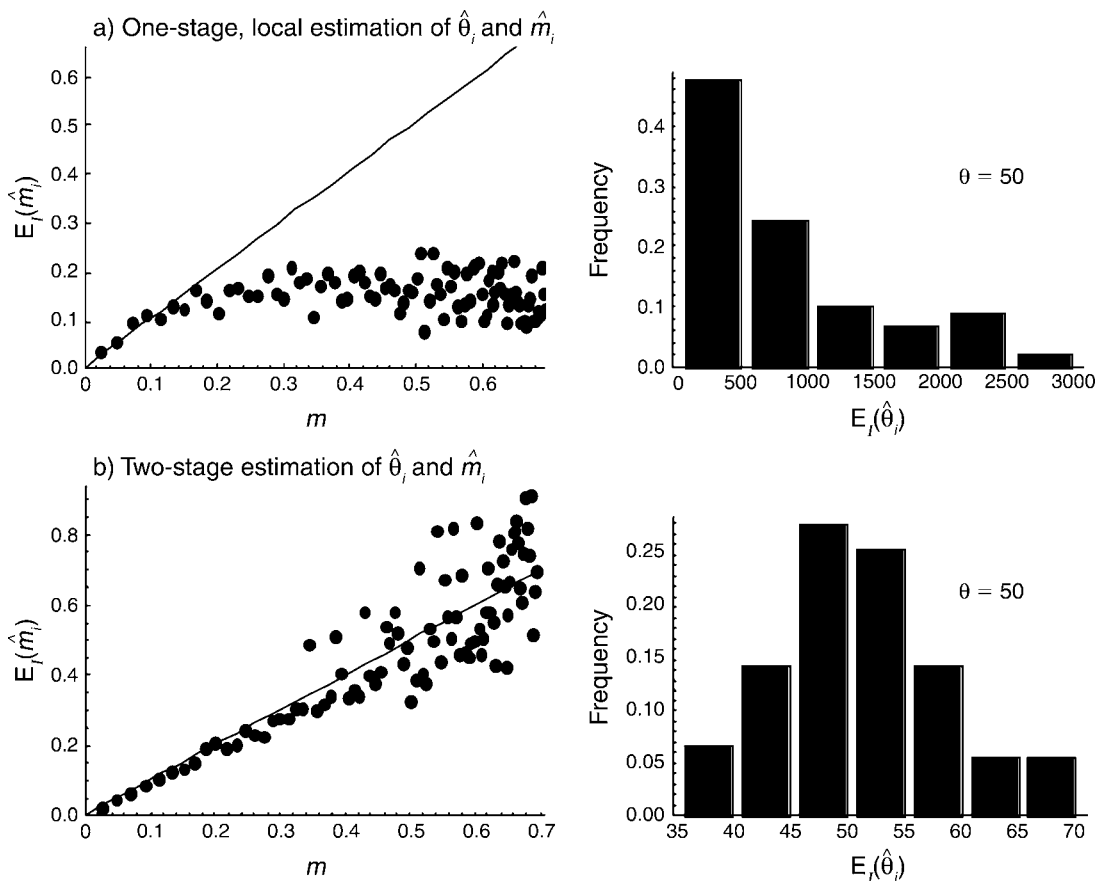


FIG. 2. Estimation of  $\hat{\theta}_i$  and  $\hat{m}_i$  from the simulated community samples, with  $J=400$ ,  $\theta=50$ , and  $I$  varying between 10 and 900, with increment 10, using (a) Etienne's sampling formula in each sample (one-stage), and (b) our new two-stage approach. Plots on the left compare expected values of  $m$  (abscissas) to  $E_I(\hat{m}_i)$  estimates (ordinates), where  $E_I(\hat{m}_i)$  is the average of estimates  $\hat{m}_i$  over 50 replicates of local community samples  $i$  belonging to the same metacommunity, and corresponding to a given  $I$ . The straight line is  $E_I(\hat{m}_i) = m$ . Histograms on the right display the relative frequency distribution of  $E_I(\hat{\theta}_i)$  values, which are calculated in panel (a) on average over 50 community samples using Etienne's sampling formula in each sample, and in panel (b) from 100 metacommunity samples, composed of one arbitrary individual from each elementary community sample originating in the same ancestor pool. Bins are referenced by their edge values.

In fact, for simulated samples, the one-stage, very variable  $\hat{\theta}_i$  and  $\hat{m}_i$  estimates showed a clear hyperbolic pattern (Fig. 1), fairly well fitted by the functional relationship existing between  $\theta$  and  $I$ , through the within-sample Simpson diversity (see Etienne 2005). Such a pattern allows understanding why there could be switches between high  $\hat{\theta}_i$ /low  $\hat{m}_i$  and low  $\hat{\theta}_i$ /high  $\hat{m}_i$  estimated values. Such switches, which have been evoked by Etienne et al. (2006), are likely when sample sizes are small, and basically do not have any biological meaning. On average, the one-stage estimates went towards inconsistently high  $\hat{\theta}_i$ /low  $\hat{m}_i$  values for simulated samples (Fig. 2a), while going toward low  $\hat{\theta}_i$ /high  $\hat{m}_i$  values for forest plots. It seems that the method displays a very high sensitivity to violations of neutrality and/or of independence between samples, which are likely to occur in most real-world data sets. Furthermore, the nature and importance of such violations probably also affect the direction of the switch.

Although the one-stage estimation at individual plot level is likely to be highly sensitive to deviations from neutral assumptions, no test is presently available to assess the magnitude of such deviations in local communities. An appropriate test would greatly help to state on the reliability of local estimation (McGill et al. 2006), and could be, at low migration rates, a simple extension of the test of Ewens (Slatkin 1982). A related problem, inherent to small samples, is that large stochastic variation in species abundances could make the respective signatures of  $\theta_i$  and  $I_i$  too ambiguous to be distinguished at local scale, resulting in meaningless values of  $\hat{m}_i = 1$ , as in our results for the Western Ghats, or of very high  $\hat{\theta}_i$  (see Fig. 1), with large confidence intervals of the estimated parameters (Etienne et al. 2006, Etienne and Alonso 2007).

On the other hand, the two-stage approach, apart from being conceptually conform to the two-scale theory of Hubbell (2001), displayed no notable bias



PLATE 1. The Ashoka tree (*Saraca asoca* (Roxb.) Wild., Caesalpinoideae, Fabaceae) naturally occurs in the wet evergreen forests of the Western Ghats of Karnataka. It is a sacred tree for Hindu and Buddhist religions and is used as an important component in Ayurveda traditional medicine. Photo credits: (left) F. Munoz, (right) B. R. Ramesh.

when applied to the simulated samples and provided realistic estimates for the forest plots, in agreement with the values found by Hubbell (2001:144–147) for semievergreen and evergreen forests in central America. This improved estimation of neutral parameters through a two-scale framework may represent a valuable contribution for a better understanding of the spatial dynamics of tree diversity in such forests. Noticeably, the variation of  $\hat{m}_i$  across samples could be further analyzed in relation to the ecological context of the different plots.

The two-stage method relies on the assumption that an acceptable random sample of the metacommunity can be extracted from the field plots by randomly drawing one individual in each plot, i.e., what we called the resampling procedure. Such a way of building the metacommunity sample greatly alleviates the risk of underestimating  $\hat{\theta}_{\text{meta}}$  due to the aggregation of individuals of the same species in some of the plots, a likely feature of many real-world data sets. For instance, estimating  $\hat{\theta}_{\text{meta}}$  by lumping all our Western Ghats field plots yielded a value of  $\hat{\theta}_{\text{meta}} = 50.99$ , substantially lower than the result obtained using resampling (i.e.,  $\hat{\theta}_{\text{meta}} = 62.3$ ), while there was no difference between lumping and resampling estimates of  $\hat{\theta}_{\text{meta}}$  for communities simulated without any form of spatial dependence. The resampling procedure yields far smaller metacommunity samples than the lumping procedure, but averaging estimated values over many iterations of the resampling/estimation process is likely to avoid any serious loss of field information, even though only a small fraction of the individual trees sampled in the field are resampled at each iteration. The resampling procedure is thus probably to be preferred to lumping. Regarding simulated neutral community samples (Fig. 2b), 50 independent plots (thus 50 resampled individuals) appeared to be enough to provide a good estimation of  $\hat{\theta}_{\text{meta}}$ , so that we assumed it should be also acceptable in our case study. But, importantly, field plots should be located far away from each other relative to the dispersal

distance, in order to allow the resampling process to provide acceptable samples of the metacommunity.

Solely using Etienne's sampling formula may be reliable when large community samples are available, as suggested from previous applications (Etienne 2005). Using an independent implementation of the one-stage method, we indeed recovered the reasonable estimates found by Etienne (2005) for the single forest plot of Barro Colorado Island (BCI), which features more than 20 000 trees (with diameter at breast height above 10 cm). Interestingly, estimates of  $\theta$  and  $m$  for area-based subsamples of the BCI plot, as computed by (see McGill et al. 2006: Fig. 2), proved systematically different (lower  $\theta$  and higher  $m$ ) from the estimates for the entire plot. In the light of our results, this may be re-interpreted as the consequence of a switch between high  $\hat{\theta}_i$ /low  $\hat{m}_i$  estimates, determining significant biases below a certain threshold of sample size. In any case, our data set of the Western Ghats forests seemed to have an average plot size ( $J = 391$ ) too low to permit reliable one-stage estimation.

As a conclusion, one single large sample can be suitable for a one-stage estimation, whereas a two-stage estimation is more reliable if numerous, fairly independent small samples are available. In sampling theory, there is a classical trade-off in allocating a given effort of data collection towards either more samples or larger ones. Further investigations are therefore needed to assess the efficiency of sampling strategies with respect to bias and variance of neutral parameter estimates, and to furnish suggestions or guidelines about adequate sampling schemes. This question is obviously not unrelated to the strength and the range of spatial autocorrelation in species distributions. For instance, the two-scale approach assumes independent samples and thus requires inter-sample distances to be larger than the range of autocorrelation. As a consequence, and although Hubbell's initial theory was spatially implicit, paying more attention to species spatial

patterns would be needed to progress toward improved inferences of neutral communities parameters.

#### ACKNOWLEDGMENTS

The Lavoisier program of the French Ministry of Foreign Affairs has funded the post-doctoral grant of F. Munoz, and this research is part of the OSDA project of the French Ministry of Ecology and Sustainable Development (program Tropical Ecosystems, convention 0000223). Field data collection is supported by the French Global Environment Fund (FGEF) under convention number 12-645-01-501-0 (1997-99, H/CIN 100301) and by the French Institute of Pondicherry (IFP). The contributions of S. V. Patil and M. H. Swaminath are kindly acknowledged along with the support of the Karnataka Forest Department. We also thank R. Pélissier and two anonymous referees for their useful comments on the manuscript.

#### LITERATURE CITED

- Allen, T. F. H., and T. B. Starr. 1982. Hierarchy: perspectives for ecological complexity. University of Chicago Press, Chicago, Illinois, USA.
- Alonso, D., and A. J. McKane. 2004. Sampling Hubbell's neutral theory of biodiversity. *Ecology Letters* 7:901–910.
- Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. *Ecology Letters* 8:253–260.
- Etienne, R. S. 2007. A neutral sampling formula for multiple samples and an “exact” test of neutrality. *Ecology Letters* 10: 608–618.
- Etienne, R. S., and D. Alonso. 2005. A dispersal-limited sampling theory for species and alleles. *Ecology Letters* 8: 1147–1156.
- Etienne, R. S., and D. Alonso. 2007. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *Journal of Statistical Physics* 128:485–510.
- Etienne, R. S., A. M. Latimer, J. A. Silander, Jr., and R. M. Cowling. 2006. Comment on “Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot.” *Science* 311:610b.
- Etienne, R. S., and H. Olf. 2004. A novel genealogical approach to neutral biodiversity theory. *Ecology Letters* 7: 170–175.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* 3:87–112.
- Holyoak, M., M. Loreau, and D. Strong. 2006. Neutral community ecology. *Ecology* 87:1368–1369.
- Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.
- Hubbell, S. P. 2006. Neutral theory and the evolution of ecological equivalence. *Ecology* 87:1387–1398.
- Karlin, S., and J. McGregor. 1972. Addendum to a paper of W. Ewens. *Theoretical Population Biology* 3:113–116.
- Latimer, A. M., J. A. Silander, Jr., and R. M. Cowling. 2005. Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science* 309:1722–1725.
- McGill, B., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. *Ecology* 87:1411–1423.
- Moran, P. A. P. 1958. Random processes in genetics. *Proceedings of the Cambridge Philosophical Society* 54:60–72.
- Pascal, J.-P. 1988. Wet evergreen forests of the Western Ghats of India. French Institute of Pondicherry, Pondicherry, India.
- Ramesh, B. R., and M. H. Swaminath. 1999. Assessment and conservation of forest biodiversity in the Western Ghats of Karnataka, India. French Institute of Pondicherry, Pondicherry, India.
- Slatkin, M. 1982. Testing neutrality in subdivided populations. *Genetics* 100:533–545.
- Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393–430.
- Vallade, M., and B. Houchmandzadeh. 2003. Analytical solution of a neutral model of biodiversity. *Physical Review E* 68(6): 061902–061905.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

#### APPENDIX

Sequential construction of species/ancestry abundance distributions in neutral community samples, with reference to a common migrant pool (*Ecological Archives* E088-149-A1).

#### SUPPLEMENT

Species abundances in the 50 1-ha forest plots (Karnataka State, South India) (*Ecological Archives* E088-149-S1).