Evaluating the performance of neutrality tests of a local community using a niche-structured simulation model

Yayoi Takeuchi and Hideki Innan

Y. Takeuchi (takeuchi.yayoi@nies.go.jp), Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Tsukuba, JP-305-8506 Ibaraki, Japan. – H. Innan, Graduate Univ. for Advanced Studies, Hayama, JP-240-0193 Kanagawa, Japan.

Understanding the processes that underlie species diversity and abundance in a community is a fundamental issue in community ecology. While the species abundance distributions (SADs) of various natural communities may be well explained by Hubbell’s neutral model, it has been repeatedly pointed out that Hubbell’s SAD-fitting approach lacks the ability to detect the effects of non-neutral factors such as niche differentiation; however, our understanding of its quantitative effect is limited. Herein, we conducted extensive simulations to quantitatively evaluate the performance of the SAD-fitting method and other recently developed tests. For simulations, we developed a niche model that incorporates the random stochastic demography of individuals and the nonrandom replacements of those individuals, i.e. niche differentiation. It therefore allows us to explore situations with various degrees of niche differentiation. We found that niche differentiation has strong effects on SADs and the number of species in the community under this model. We then examined the performance of these neutrality tests, including Hubbell’s SAD-fitting method, using extensive simulations. It was demonstrated that all these tests have relatively poor performance except for the cases with very strong niche structure, which is in accordance with previous studies. This is likely because two important parameters in Hubbell’s model are usually unknown and are commonly estimated from the data to be tested. To demonstrate this point, we showed that the precise estimation of the two parameters substantially improved the performance of these neutrality tests, indicating that poor performance can be owed to overfitting Hubbell’s neutral model with unrealistic parameters. Our results therefore emphasize the importance of accurate parameter estimation, which should be obtained from data independent of the local community to be tested.

Ecological communities in nature comprise complex consortia of species with intricate structure, e.g. in a tropical forest, over a thousand tree species coexist in one area (Condit et al. 2006). One of the major goals in community ecology is to understand the underlying processes that govern species diversity and abundance in a community (Tilman 1982, Lande et al. 2003). Community ecologists have developed numerous models to explore community structure; their accuracy to explore empirical community data has been examined. The species abundance distribution (SAD) is a basic metric used to describe the relative abundance of species in a community, and observed SADs are often used for testing these theoretical models (Fisher et al. 1943, Preston 1948, Tokeshi 1990, Hubbell 2001, Ulrich et al. 2010, Locy and White 2013).

Two major categories of theories have been developed to explain the data of the community structure. The first of these theories is the niche theory, which incorporates deterministic factors such as interspecies competition and niche differentiation with some models allowing for a stochastic (random) process. The other is the neutral theory, which considers random drift as the major factor in community composition without including any deterministic factors. Traditionally, deterministic factors have been considered to play a major role in shaping species composition and diversity in a community (Tilman 1982, Tokeshi 1990, 1992, Chesson 2000, Sugihara et al. 2003). Niche theories assume that each species in a community would be specialized to particular combinations of resources through interspecies competition (Westoby et al. 2002). This competition involves several deterministic factors, including tradeoffs, and consequently, drive interaction between species, which thereby result in the coexistence of multiple species at equilibrium. Niche models are widely accepted because there are numerous field observations exhibiting clear evidence for niche differentiation (Wright 2002). In addition, theories under niche models predict that SADs should be approximated by a lognormal distribution, which is consistent with many field observations (Tilman 1982, Tokeshi 1990, 1992, Sugihara et al. 2003, Harpole and Tilman 2006).

On the other hand, neutral theories have also been advocated in the last decade. Caswell (1976) first introduced three neutral models into ecology, but they were not well accepted in the 20th century because they failed to provide a good fit to data from natural communities. Hubbell’s neutral model (Hubbell 2001) changed this situation because his model
was able to explain a wide range of empirical observations. His model assumes that all individuals are ecologically or functionally equivalent, i.e. no differences in reproduction and mortality among individuals. Thus, the composition of a local community is determined only by stochastic extinction, local birth, and dispersal from the nested metacommunity by random speciation. This process is elegantly summarized by only three parameters, the fundamental diversity number ($\theta$), the migration rate ($m$) from the metacommunity to the local community and the number of individuals in the local community ($J$), and the shape of the expected SAD in the local community can be characterized by a function of $\theta$, $m$ and $J$. The distribution derived from the neutral model is the so-called zero-sum multinomial distribution. This model is considered to be the strictest form of neutral model and contains several simplified assumptions. Despite these strict assumptions, Hubbell's neutral model fits the field data quite well, as can be seen in SADs from a wide range of communities (e.g. tropical forests: Etienne 2005, Volkov et al. 2007, fish communities: Etienne and Ollff 2005, and birds: He 2005).

The success of Hubbell's neutral model is particularly surprising because it provides a good fit to the data from tropical forests (Etienne 2005, Volkov et al. 2007), despite previous assumptions that niche differentiation would be the major force in maintaining high species diversity (Wright 2002). Moreover, Hubbell's neutral model, in some cases, shows a better fit to the abundance of rare species than those predicted by deterministic models (Volkov et al. 2005, He 2005). The historical reason behind the rise in popularity of Hubbell's neutral model was partly due to investigating the effects of increasing sample size. Communities represented by a small number of individuals have a SAD that is well-fitted by a lognormal distribution (Preston 1948) or a log-series distribution (Fisher et al. 1943). Preston (1948, 1962) initially predicted that if the sample size of a community was sufficiently large, the SAD would be symmetric or lognormal. However, the situation changed when community data with a large sample size in a closed community became available (e.g. 50-ha forest dynamics plots from the Smithsonian Tropical Research Institute). It was found that such SADs are negatively skewed with a large excess of rare species. Since Hubbell's neutral model was a better fit to the data for rare species, it became popular although assumptions of the underlying theory were difficult to accept for some ecologists. His model has been used as the first null model to be tested, which was suggested in a recent review by Alonso et al. (2006) (see also Gotelli and McGill 2006). Meanwhile, lognormal and log-series distributions became alternatives to SADs that represent some non-neutral process as already demonstrated by theoretical studies (May 1975, Sugihara 1980, Engen and Lande 1996, Magurran 2004).

There has been a great deal of debate on the interpretation of Hubbell's neutral model. It is obvious that Hubbell's neutral model cannot be the exclusive explanation for shaping a SAD and, consequently, has been challenged by a number of authors. Several studies demonstrated that non-neutral models fit the observed SADs better than Hubbell's neutral model (i.e. in grassland communities: Harpole and Tilman 2006, coral reefs: Dornelas et al. 2006, tropical forests: Etienne 2005, aphids: He 2005, and fish communities: He 2005). One of the main criticisms of Hubbell's neutral model is that it is too flexible, such that it can predict SADs that are generated by non-neutral models (Adler et al. 2007, Chave 2004, Bell 2005, Chisholm and Pacala 2010). This is because the model, which predicts the SAD in the local community of interest, is dependent upon $\theta$ and $m$, which are usually unknown. Therefore, in the fitting process, $\theta$ and $m$ are conventionally estimated from the data of the 'local' community to be tested. As these two estimated parameters are optimized to the local community, it is not surprising that Hubbell's neutral model often fits the observed SAD. Consistent with this intuitive understanding, there are a number of theoretical reports demonstrating that non-neutral models can predict very similar patterns of SADs as do other summary statistics, to those expected under Hubbell's neutral model. For example, Chisholm and Pacala (2010) have recently presented an analytical framework to prove that niche-structure could predict a similar pattern of SADs of neutral communities (see also Purves and Pacala 2005). The consensus seems to be that niche and neutral models can generate similar patterns if parameters are adjusted (Adler et al. 2007, Chave 2004, Volkov et al. 2005, Bell 2005). Therefore, it is apparent that the major problem with Hubbell's SAD-fitting approach (2001) is that it likely misses the signature of non-neutral factors, since it generally has a poor ability to reject neutrality. Although this has been pointed out repeatedly (Adler et al. 2007, Chave 2004, Bell 2005, Chisholm and Pacala 2010, Clark 2012, Rosindell et al. 2012), there has not been a systematic likelihood-based quantitative test for this. For example, Chave et al. (2002) visually compared SADs generated from neutral and niche models, but did not provide statistical tests of the neutral model.

Motivated by this problem, various statistical methods have been recently developed by many researchers in the field. One method proposed is the 'exact test' by Etienne (2007), which is based on Fisher's exact test and similar tests found in population genetics by Slatkin (1994, 1996) (see also Innan et al. 2005). Note that one cannot expect the 'exact' performance of this test as it also requires estimated values for $\theta$ and $m$ (Etienne 2007), so the same problem still remains. Furthermore, because the 'exact' computation of the probabilities for all possible patterns of species abundance is not computationally feasible, it employs approximations using likelihood.

Another approach to fit the neutral model is that of summary statistic-based tests that are similar to Watterson's homozygosity test used in population genetics (note: the analogue of Watterson's homozygosity test in ecology is Shannon's index). Jabot and Chave (2011) developed a statistical test to examine if the observed Shannon's index is consistent with a null distribution predicted by Hubbell's neutral model, conditional on the number of observed species. As with other tests, it requires estimated values for $\theta$ and $m$. Because these tests are relatively new and their applications to real field data are still limited, it is unclear how and under what conditions they perform.

The main aim of this study is to evaluate the performance of these neutrality tests quantitatively by performing extensive simulations. To this end, we first develop a simple niche model which incorporates stochastic demography. The advantage of this model is that it has a parameter, $p$, which
represents the degree of niche differentiation. \( P \) ranges on the closed interval \([0,1]\). When \( p = 1 \), the model is identical to Hubbell’s neutral model, and as \( p \) decreases, the degree of niche differentiation becomes stronger. In the extreme case when \( p = 0 \), it is assumed that each niche can be occupied by only one particular species. This idea of niche differentiation is similar to previous studies (Gravel et al. 2006, Tilman 2004) where these models consider a stochastic process of birth and death, and each species is assumed to have a preference to a specific environment or niche. As with our model, these models have a parameter to determine the degree of niche overlap among species. Thus, with this type of neutral-niche model, we can quantitatively assess the relationship between the degree of niche differentiation (i.e. \( p \)) and the performance of various neutrality tests.

In this study, we perform extensive simulations by varying \( p \) from 0 to 1 and explore the performance of various neutrality tests. Included in our analysis is Hubbell’s SAD-fitting approach (Hubbell 2001), Etienne’s exact test (Etienne 2007), and summary statistic-based tests, including those using Shannon’s index (Jabot and Chave 2011). In addition, we also develop similar tests using other summary statistics, and assess their performance under various conditions. We discuss the possibility of using more powerful techniques.

Model

Our model is spatially implicit and we focus on the species abundance in a local community (see the works by Rosindell et al. 2008 and Matthews and Whittaker 2014, which developed the spatially explicit neutral SAD models). It is assumed that there is a metacommunity that provides a source of individuals to the local community. Let the metacommunity consist of \( f_{M} \) individuals for a local community of \( f \) individuals, with \( f \) being constant. It is generally assumed that the size of the metacommunity is several orders of magnitude larger than the size of the local community.

Hubbell’s neutral model

As our niche model is very similar to Hubbell’s neutral model (Hubbell 2001) with the exception of one process, we first explain how a local community can be simulated under Hubbell’s neutral model (Hubbell 2001). Here, it is assumed that we can count the number of individuals in a local community in the field, so we fix \( f \). Thus, we consider that the neutral model has only two parameters, \( \theta \) and \( m \). Each simulation run can be described with the following steps.

1) Create the metacommunity

The diversity and relative abundance of species in the metacommunity are predetermined by the composite parameter \( \theta \), which is referred to as the ‘fundamental biodiversity number’ \((\theta = \frac{1}{\nu} (f_{M} - 1))\), where \( \nu \) is the probability of speciation per birth. The configuration of the metacommunity is governed by Ewens’ sampling formula (Ewens 1972) in which the SAD follows a log-series distribution (Hubbell 2001). For further details, see Etienne and Alonso (2007). Under a given value of \( \theta \), a random configuration of the metacommunity with \( f_{M} \) individuals can be obtained by the Hubbell’s method (2001) (see Supplementary material Appendix 1 for detailed algorithms). Let \( S_{M} \) be the total number of species in the simulated metacommunity. This configuration of the metacommunity will be fixed for the subsequent steps in simulating the local community.

2) Create the initial local community

The initial state of the local community with \( f \) cells is randomly created; i.e. all \( f \) cells are filled by randomly choosing individuals from the metacommunity. Depending on this initial state, the dynamics of local community can be simulated forward in time.

3) Simulate the dynamics of the local community

Next, we simulate the dynamics of the local community by randomly replacing individuals from this community. The simulation can be performed by repeating many small time steps. At each time step, individuals die at a given mortality rate (all individuals have equal susceptibility to mortality). Empty cells due to deaths are randomly filled by immigrants from the metacommunity with probability \( m \) and offsprings of the remaining local community members with probability \( 1 - m \). Thus, there are no empty cells because a death is always replaced by either a birth or an immigrant (i.e. the ‘zero-sum dynamics’). This demographic stochasticity is called ‘ecological drift’ (Hubbell 2001). Another important assumption is that of ecological equivalence among species or individuals. In short, this considers that all individuals have equal mortality rates, equal fecundities, and equal probabilities of their offsprings taking over the cell on which they land, regardless of the previous occupant of the cell.

4) Evaluate the configuration of the local community

The end result of the simulation of the local community is obtained after 20 000 time steps, whereby the diversity and relative abundance of species in the local community can be evaluated.

Niche model

In our niche model, we modify steps 2) and 3) of Hubbell’s neutral model in order to incorporate the effect of niche differentiation in the local community.

2) Create the initial local community

It is assumed that there are \( N \) different niches in the local community. Each cell in the local community belongs to one niche, and the number of cells in each niche is determined by a multinomial distribution with parameters \( \left( \frac{1}{N}, \frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N} \right) \). \( N \) is determined such that it does not exceed the total number of species in the metacommunity, \( S_{M} \) (which was given in the step 1). \( q_{ij} \) is the parameter specifying the property of the \( i \)-th niche \((i = 1, 2, 3, \ldots N)\), which is determined such that \( q_{ij} = 1 \) if the \( i \)-th niche allows the \( j \)-th species to occupy, otherwise \( q_{ij} = 0 \). Therefore, the property of niche adaptation of the entire local community is described by the \( N \times S_{M} \) matrix denoted by \( M \):
th niche:
lows a binomial distribution, can be a random value from
i.e. the
there is a one-to-one relationship between niche and species;
i.e. the
for niche differentiation with
overall niche specificity. Let us first consider the strictest case
matrix is given by
We now define
the expected proportion of species that can occupy a niche.
We also consider an intermediate case, where
as the proportion of species that can be
within the local community. If it is the former case, then
is an immigrant from the metacommunity or a birth
Determine whether the next individual to fill this cell I.
Let
be the species number of the chosen individual.
Fill the cell if II.
Pick a random individual from the metacommunity. Let
be the species number of the chosen individual. Fill
the cell if III.
Pick a random individual from the local community. Let
be the species number of the chosen individual. Fill
the cell if I.

\[
M = \begin{bmatrix}
q_{1,1} & q_{1,2} & q_{1,3} & q_{1,4} & \cdots & q_{1,SM} \\
q_{2,1} & q_{2,2} & q_{2,3} & q_{2,4} & \cdots & q_{2,SM} \\
q_{3,1} & q_{3,2} & q_{3,3} & q_{3,4} & \cdots & q_{3,SM} \\
q_{4,1} & q_{4,2} & q_{4,3} & q_{4,4} & \cdots & q_{4,SM} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
q_{N,1} & q_{N,2} & q_{N,3} & q_{N,4} & \cdots & q_{N,SM}
\end{bmatrix}
\]  

Next, we introduce the parameter \( p \), which characterizes the overall niche specificity. Let us first consider the strictest case for niche differentiation with \( p = 0 \), in which we assume that there is a one-to-one relationship between niche and species; i.e. the \( i \)th niche can be occupied only by the \( i \)th species. The matrix is given by

\[
M_{p=0} = \begin{bmatrix}
1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix}
\]

We now define \( q_{ji} = 1 \) (\( i = 1, 2, 3, \ldots, N \)) for convenience, so that the remaining species (from species \( N + 1 \) to \( S_M \)) cannot survive in any niche in the local community. For the other extreme case with \( p = 1 \), it is assumed that all niches can be occupied by any of the \( S_M \) species, so that \( M_{p=1} \) is given by,

\[
M_{p=1} = \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & 1 & 1 & 1 & \cdots & 1 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 1 & 1 & \cdots & 1
\end{bmatrix}
\]

We also consider an intermediate case, where \( p \) represents the expected proportion of species that can occupy a niche. Let us define \( \bar{q}_i \) as the proportion of species that be accepted in the \( i \)th niche:

\[
\bar{q}_i = \sum_{j=1, i \neq j}^{S_M} q_{ij} / S_M - 1
\]

Then, \( M_p \) is given such that

\[
E(\bar{q}_i) = p
\]

holds for all rows.

For a simulation given a specified value of \( p \), we can construct a random matrix \( M_p \) by defining a certain function for \( \bar{q}_i \). Since any function can work, we chose a beta function \( Beta(1, b) \), where \( b \) is adjusted so that the mean of \( Beta(1, b) \) is \( p \) (For example, \( b = 1 \) if \( p = 0.5 \)). A beta distribution provides a relatively wide range of values between 0 and 1, so that the local community can consist of variety of niches, from strong to weak, for an intermediate \( p \). Let \( q'_i \) be a random value from \( Beta(1, b) \). Then, \( Q_i \), the number of species that are able to survive in the \( i \)th niche and follows a binomial distribution, \( Binom(S_M - 1, q'_i) \). With \( Q_i \), a vector \( \{q_{1,1}, q_{1,2}, q_{1,3}, \ldots, q_{1,SM}\} \) can be constructed as follows: First, \( q_{ij} = 1 \) by definition. Next, \( Q_i \) columns are randomly chosen from the remaining \( S_M \) columns. By using this method, all rows of the matrix \( M \) can be determined.

Then, the initial state of the local community can be created once this matrix \( M \) is specified. Note that, as stated earlier, it is already determined which cells in the local community belong to which niches. With this setting, each of the \( J \) empty cells is filled by the following procedure. For a cell that belongs to the \( i \)th niche,

I. Pick a random individual from the metacommunity. Let \( j \)
be the species number of the chosen individual.
II. Fill the cell if \( q_{ij} = 1 \), otherwise go to [I]. Continue until
this cell is filled.

This initial setting is fixed by the simulation of the local community. The configuration of the metacommunity is also fixed.

3) Simulate the dynamics of the local community
We next simulate the dynamics of the local community by randomly replacing individuals from this community, as before. At each time step, individuals die at a given mortality rate, and these empty cells are randomly recolonized by new individuals. This process is similar to that of constructing the initial local community. That is, if an empty cell belongs to the \( i \)th niche,

I. Determine whether the next individual to fill this cell is an immigrant from the metacommunity or a birth within the local community. If it is the former case, then choose with probability \( m \) and go to step [II], otherwise, go to step III.
II. Pick a random individual from the metacommunity. Let \( j \)
be the species number of the chosen individual. Fill the cell if \( q_{ij} = 1 \), otherwise repeat this step until the cell is filled.
III. Pick a random individual from the local community. Let \( j \)
be species number of the chosen individual. Fill the cell if \( q_{ij} = 1 \), otherwise repeat this step until the cell is filled. Note that, although rare, there can be situations where this procedure does not work because some of the individuals in the local community cannot survive in this niche (i.e. \( q_{ij} = 0 \) for all individuals in the local community). For such cases, the cell may be filled by an immigrant from the metacommunity. That is, go to step II.

Simulations
In our simulation, we assume \( J = 10\,000 \) and \( J_M = 10\,000 \) 000. A single simulation run consists of 20 000 time steps with a mortality rate of 1% per step, which are based on previous studies of neutral models in tropical forest (Condit et al. 2006). We set \( \theta = 50 \) and \( m = 0.1 \), which are very close to estimates in tropical forests under the neutral model (Etienne 2005). We consider five different values of niches, \( N = \{1, 5, 10, 100, N_{max}\} \), where \( N_{max} \) is the potentially maximum number of niches and is identical to \( S_M \). Note that \( S_M \) is a variable determined by \( \theta \) and \( J_M \). For example, if \( \theta = 50 \) and \( J_M = 10\,000 \) 000, \( S_M \) would be around 650. Suppose \( S_M \) is randomly determined to be 652 in step I, then we assume \( N_{max} = 652 \) when we analyze the result of this iteration. This is a commonly used method found in previous neutral model simulation studies (Hubbell 2001).
For \( p \), we used values of \( p = [0.01, 0.5, 0.8] \), in addition to the completely neutral case where \( p = 1 \). In this study, simulations were performed for all pairs of values for \( p \) and \( N \), except for \((p, N) = (0, 1)\) because this is obviously a meaningless parameter set, i.e. the case where the community is composed of only one species with 10 000 individuals.

**Model selection**

A common approach to test Hubbell’s neutral model is to compare the goodness-of-fit between the neutral model and alternative models by using AIC (Akaike’s information criterion, Akaike 1973). Both lognormal (Preston 1948) and log-series distributions (Fisher et al. 1943) have been commonly used to represent non-neutral cases. We also include our niche model as an alternative, so that we can select the best fit amongst Hubbell’s neutral model and the three non-neutral models (lognormal, log-series and our niche models). Below, we explain the process of how these four models are fit to an observed SAD, which is generated by a simulation as described in the previous section. Note that AIC can be suitable for model selection among nested models (simpler cases are special cases of more complex models) although it is commonly used in ecology to compare non-nested models (Johnson and Omland 2004). In this sense, our niche model is one of few alternative models in which Hubbell’s neutral model is nested. Therefore, it is possible to statistically compare these two models with the AIC approach, or with more sophisticated likelihood-based methods. Nevertheless, in order to include all three alternative models, we use the conventional AIC-based approach (i.e. non-nested cases).

To evaluate the performance of this model selection approach, we simulated a large number of SADs under our niche model, and investigated which model was selected for each SAD. For a given simulated SAD, the four models were fitted and the maximum log-likelihoods were computed. Although there are a number of methods and software available to estimate the best-fit parameters, in this study, it was necessary to modify them in order to make a statistically fair comparison of the likelihoods for each of the four models. Because it is not possible to obtain a reliable SAD under our niche model (particularly for very rare abundance due to the lack of an analytical expression), we had to use a binned SAD. Therefore, to be consistent, the likelihoods for all four models were computed for a binned SAD. In practice, we employed Preston’s method (Preston’s octave, Preston 1948) for creating a log-based binned SAD, and the log likelihood was computed by:

\[
LL' = \sum r \left[ n_r \log \left( \frac{E_r}{S} \right) \right]
\]

where \( n_r \) and \( E_r \) are the observed number of species and the expected number of species in the \( r \)th abundance bin, respectively, and \( S \) is the total number of species. The expected SAD can be computed either analytically or by a simulation using the estimated parameters in each model (see below for each model). We searched the parameter set that maximizes the likelihood for all four models. In the following, we detail this process for each of the four models.

**Hubbell’s neutral model**

Under Hubbell’s neutral model, the binned SAD is given as a function of \( \theta \) and \( m \) (\( f \) is assumed to be known). Because there is no adequate analytical expression for SADs (Volkov et al. 2003) for given values of \( \theta \) and \( m \), the binned SAD was obtained by averaging over 100 independent artificial species-abundance configurations by using Etienne’s program (2007).

The best-fit parameters were searched using a wide range, where \( \theta = \{1.2, 3, ..., 300\} \) and \( m = \{0.05, 0.1, 0.15, ..., 1\} \) using the likelihood function:

\[
LL = \sum r \left[ n_r \log \left( \frac{E_r}{n_r} \right) - [E_r - n_r] \right]
\]

(Kempton and Taylor 1974, Hubbell 2001). We confirmed that the best-fit to \( \theta \) and \( m \) are almost identical to their maximum likelihood estimates obtained by using Ewen’s sampling formula implemented by Etienne (2005). Using the best-fit parameter set, the likelihood of the binned SAD was re-calculated by Eq. 6 for subsequent model selection.

**Niche model**

This model has four parameters which are \( \theta, m, N \), and \( p \). For a parameter set, the binned SAD was obtained by averaging over 100 independently simulated SADs (20 000 time steps for each replication), and the best-fit parameters set was searched using Eq. 7 with wide ranges for the four parameters \( \{\theta = \{1, 2, 3, ..., 300\}, m = \{0.05, 0.1, 0.15, ..., 1\}, N = \{1, 5, 10, 100, N_{max}\} \), and \( p = \{0 (N \neq 1), 0.1, 0.5, 0.8\} \). Using the best-fit parameter set, the likelihood of the binned SAD was calculated by Eq. 6 for model selection.

**Lognormal function**

It is traditionally accepted that there are occasions in which SADs can be approximated well by a lognormal distribution. For example, a community under many ecological factors or a community with multidimensional resource utilization (May 1975, Magurran 2004) can be fitted with a lognormal distribution. This type of distribution can be specified by two parameters – the mean and variance. To fit a lognormal distribution to an observed SAD, it is necessary to search for the best-fit parameters. To accomplish this, it is common to use a SAD binned in Preston’s octave (O’Hara and Oksanen 2003), to which a generalized linear regression model with a standard lognormal distribution or a Poisson lognormal distribution is fitted. Here, we employ the former one because it shows a better fit to our model than a Poisson lognormal. It is known that this method provides identical estimates to the maximum likelihood method. After estimating the mean and variance of the lognormal distribution, we computed the log-likelihood of the binned SAD by Eq. (6).

**Log-series function**

A log-series distribution can approximate a typical SAD in the following: a community where the dynamics is simply dominated by one or a few ecological factors, a community where the dominant species preempt the major part of the limited resource or, a community that is not in equilibrium (May 1975, Magurran 2004). Note that a log-series distribution is usually applied to an open community, although Hubbell’s neutral SAD and lognormal distribution consider
a fully census or closed community. Nevertheless, we apply a log-series distribution to the ‘closed’ local community because this type of community can be considered to be a special case of a subsampled or an open community. Fitting a log-series distribution involves estimating Fisher’s two parameters, \( \alpha \) and \( x \) (Fisher et al. 1943). For any pair of \( \alpha \) and \( x \), the expected binned SAD was numerically obtained, and the log-likelihood was computed by Eq. 6. The best-fit parameter pair was searched using wide ranges of \( \alpha \) and \( x \) (\( \alpha = \{1.00, 1.01, 1.02, \ldots, 200\} \) and \( x = \{0.9500, 0.9502, \ldots, 1\} \)).

### Performance of neutrality tests

The performance of several neutrality tests are compared by applying them to the simulated data. We consider Etienne’s exact test (Etienne 2007, Etienne and Rosindell 2011) and other summary statistic-based tests, as summarized in Table 1. Etienne’s exact test can be considered as a summary statistic-based test because the likelihood of the configuration is treated as a summary statistic.

Given a set of simulated data, we computed the summary statistics and the corresponding p-values, as shown in Table 1. We obtained the distributions of the summary statistics by randomly generating data (1000 replications) dependent on the maximum likelihood estimates for \( \theta \) and \( m \). We define the p-value as the proportion of iterations that rejected Hubbell’s neutral model for the estimates for \( \theta \) and \( m \). This procedure is shared by all summary statistic-based tests as described below (also see Supplementary material Appendix 1 Fig. A1).

I. Simulate data \( D \) to be tested, where \( D \) represents the configuration of species abundance. Then, compute the summary statistic (SS) for \( D \), which is denoted by \( SS_D \).

II. Determine the P-value for \( SS \). First, estimate \( \theta \) and \( m \) from \( D \) using the maximum likelihood method based on Ewens’ sampling formula implemented in the PARI/gp program by Etienne (2005). Note that \( J \) is known since we consider a closed local community, that is, we have data for all individuals in the community. Then, we independently generate 1000 iterations of the species-abundance configuration under Hubbell’s neutral model according Etienne’s algorithms implemented in the PARI/gp program (Etienne 2005, 2007). For each random configuration, we calculate various summary statistics (Table 1).

III. Determine the P-value. Etienne’s exact test is treated as a one-tailed test, while all the others are two-tailed tests. Let \( r \) be the proportion of the simulation runs with \( SS \) (or likelihood) less than \( SS_D \). Then, the P-value for Etienne’s exact test is identical to \( r \), and for the other two-tailed tests, the p-value is \( 2r \) if \( r < 0.5 \) otherwise \( 1 - 2r \).

Table 1. Summary statistics investigated in this study.

<table>
<thead>
<tr>
<th>Summary statistics</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S )</td>
<td>Species richness (Number of species)</td>
<td>Magurran (2004)</td>
</tr>
<tr>
<td>( H )</td>
<td>Shannon’s index, ( H = \sum u_j \times \ln(u_j) ), where ( u_j ) is relative abundance of species ( j ) in the community</td>
<td>Magurran (2004)</td>
</tr>
<tr>
<td>( D )</td>
<td>Simpson’s index, ( D = 1 - \sum u^2_j )</td>
<td>Magurran (2004)</td>
</tr>
<tr>
<td>( \text{inv}N )</td>
<td>( \text{inv}N = \sum_j \frac{1}{n_j} )</td>
<td>Jabot and Chave (2009)</td>
</tr>
<tr>
<td>( \text{inv}N^2 )</td>
<td>( \text{inv}N^2 = \sum_j \left( \frac{1}{n_j} \right)^2 )</td>
<td>Jabot and Chave (2009)</td>
</tr>
<tr>
<td>( \text{var}N )</td>
<td>variance of ( n_j )</td>
<td>Jabot and Chave (2009)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Fisher’s ( \alpha )</td>
<td>Fisher et al. (1943)</td>
</tr>
<tr>
<td>( \text{rarefaction} \times 0.1 )</td>
<td>Rarefaction species richness, which is the number of species by randomly resampling with sample size ( x = J \times 0.1 )</td>
<td>Magurran (2004)</td>
</tr>
<tr>
<td>( \text{rarefaction} \times 0.5 )</td>
<td>Rarefraction species richness, which is the number of species by randomly resampling with sample size ( x = J \times 0.5 )</td>
<td>Magurran (2004)</td>
</tr>
</tbody>
</table>

### Results

Our simulations clearly demonstrate that there is a strong effect of niche-preference on the pattern of species diversity (i.e. SAD and \( S \)) in the local community (Fig. 1, A2). When \( p = 1 \) (complete neutral case), the average \( S \) is 197.8 ± 5.5 (± SD), which is consistent with Hubbell’s neutral model. The other extreme case would be when \( p = 0 \) and \( N = 1 \), where all cells in the local community belong to one kind of niche and only one species is allowed to occupy the niche, so that there is only one species with abundance \( J = 10000 \). As the number of niches (\( N \)) increases (with fixed \( p = 0 \)), the number of species (\( S \)) increases to the theoretical maximum (\( S = 650 \), which is the approximate number of species in the...
Figure 1. The expected SADs of the niche model for various pairs of \( p \) and \( N \). The expected SAD was obtained by averaging over 100 iterations of the simulation. Error bars represent SD. Four models (Hubbell’s neutral, lognormal, log-series, and our niche model) were fitted to each average SAD. The estimated \( q \) and \( m \) for fitting Hubbell’s neutral model are shown in blue, lognormal in yellow, log-series in orange, and our niche model in red.

metacommunity when \( J M = 1000000 \) and \( \theta = 50 \). As \( p \) increases (with \( N \) fixed), the SAD becomes close to the neutral model. Thus, our model enables us to explore situations with various degrees of niche differentiation.

Our results demonstrate the strong effects of niche differentiation on the SAD and the number of species. However, as mentioned previously, if we look at the SAD alone, the observed SAD is approximated well by Hubbell’s neutral model by visual inspection (blue line, Fig. 1). This is in good agreement with previous studies which also showed a good fit to Hubbell’s neutral model (Volkov et al. 2007, Chave et al. 2002, Hubbell 2001). The success of Hubbell’s neutral model is due to the estimates of \( q \) and \( m \) which are far from the given values (\( q^* = 50 \) and \( m^* = 0.1 \), especially for small \( p \), (\( q^*, m^* \) = (5.8, 0.2) for \( (N, p) = (1, 0.1) \) in Fig. 1; see also Fig. A3). It was found that a lognormal distribution also provides the best fit to the observations for a wide range of \( p \) and \( N \), while the fit of a log-series distribution is not very good (Fig. 1, 2A). This result is consistent with previous studies (Adler et al. 2007, Chave 2004, Volkov et al. 2005, Bell 2005). The major purpose of this study is to quantitatively evaluate this problem in model selection. We performed several simulations, and the results are summarized in Fig. 2A. Under the complete neutral model (\( p = 1 \)), the observed SADs in 98 of the 100 iterations are best explained by Hubbell’s neutral model. The inferred parameters were \( \hat{\theta} = 50 \) and \( \hat{m} = 0.1 \) (Fig. A3), which were close to the given parameters. The pattern is not very different when \( N = 1 \) or \( p = 0.8 \) as the neutral model is best supported in \( \sim 70\% \) of iterations, and the fit of the lognormal distribution is second best. For \( N \geq 5 \) and \( p \leq 0.5 \), either the lognormal or our niche model ranks first. The fit of our niche model is particularly good with very strong niche-structure (i.e. \( N \geq 5 \) and \( p \leq 0.1 \)). Thus, when the parameters \( \theta \) and \( m \) are estimated from the local community itself, unless the degree of niche preference is very strong (for small \( p \) and large \( N \)), it can be concluded that the fit of Hubbell’s neutral model is quite good, and the power to reject Hubbell’s neutral model is limited. The major reason for this overfitting of Hubbell’s neutral model is that we do not know the precise values of \( \theta \) and \( m \). To demonstrate this, we performed the same analysis by assuming we know the values of \( \theta \) and \( m \) (\( \theta = 50 \) and \( m = 0.1 \)). We then found that the ability to reject Hubbell’s neutral model is substantially improved especially for intermediate values of \( p \) (Fig. 2B).

Throughout this study, we used Preston’s octave classes (Preston 1948), which are log₂-based bins (i.e. \{1, 2, 3–4, 5–8, 9–16, \ldots\}) with some adjustment at the borders between adjacent bins. Although Preston’s octave classes are commonly used, our results may differ if we use other bins. To check the effect of bin sizes, we repeated the same analysis with two additional bins, normal log₂ and log₁₀. As shown in the Supplementary material Appendix Fig. A4 and A5, we obtained essentially identical results to those with Preston’s octave classes, except that the fit of our niche model is generally better.

We also explored the performance of other neutrality tests, namely, Etienne’s exact test and the summary statistic-based tests summarized in Table 1. Note that all these tests are usually performed with \( \theta \) and \( m \) estimated from the local community to be tested, so that they share the same problem of the SAD-fitting approach, but the extent of the sensitivity
Figure 2. Summary of the model selection by AIC. The proportion of simulation runs that support the neutral, lognormal, log-series, and our niche model are shown in different colors. (A) Results when estimated $q$ and $m$ are used. (B) Results when parameters are known ($q^\wedge = 50$ and $m^\wedge = 0.1$). See text for details.

to these unknown parameters may differ depending on the test. Fig. 3A shows the $p$-values for all 10 tests when $\theta$ and $m$ were used. It was found that Hubbell's neutral model was rejected in almost all cases (when $p \geq 0.8$ or $N = 1$), which is consistent with Hubbell's SAD fitting approach. Etienne's exact test, species richness, Fisher's $\alpha$, and rarefaction $J \times 0.5$ failed to reject Hubbell's neutral model in most cases (except for Etienne's exact test when $p = 0$ and $N = 100$,
Figure 3. Summary of the Etienne’s exact test (1) and neutrality tests based on nine summary statistics (2–10). The proportion of simulation runs that rejected Hubbell’s neutral model are shown in orange, others are in blue. (A) Results when estimated $\theta$ and $m$ are used. (B) Results when parameters are known ($\hat{\theta} = 50$ and $\hat{m} = 0.1$).

650). Rarefaction $J \times 0.1$ rejected Hubbell’s neutral model only when the niche structure was strong (when $p$ is small, and $N$ is large). On the other hand, Shannon’s $H$, Simpson’s $D$, $invN$, $invN_p$ and $varN_p$ performed better, suggesting that they are relatively sensitive to niche structure.

We also investigated the manner in which the performance of these tests can be improved if the values of $\theta$ and $m$ are known. As expected, Fig. 3B shows that the performance is substantially improved in comparison to Fig. 3A. Thus, it is reconfirmed that not knowing the true
values of \( \theta \) and \( m \) causes a reduction in the performance of the neutrality tests.

Discussion

It seems that there is a two-fold problem in testing the neutrality in community ecology. First, there are a number of possible neutral models, but the best known one (Hubbell's neutral model), has been so well accepted and widely used that it has become the representative neutral model. However, rejecting Hubbell's model does not necessarily mean that neutrality is rejected. Second, in most cases, it is quite difficult to reject even the simplest neutral model by using the current methods and available data. The focus of this study is the second problem, namely, the problem of current methods because we cannot proceed without solving this technical problem. The first one will be challenging itself and deferred to as a next step, which is beyond the scope of this study.

In this study, we first developed a niche model that incorporates the stochastic demography of individuals together with the mechanism of niche differentiation as a deterministic factor. The model involves two parameters, \( p \) and \( N \), where the former represents the degree of niche preference and the latter is the number of different niches in the local community. Our niche model has a nested-structure within Hubbell's neutral model, which allows us to make a fair statistical comparison between the two models and we can select the best model according to AIC. Furthermore, it is possible to apply more statistically rigorous approaches, such as the likelihood ratio test. Note that the AIC-based comparison of Hubbell's neutral model with the lognormal and log-series distributions is not statistically correct. Since this method is used frequently, we applied it in order to investigate the relative performance.

Another advantage of our model is that it allows one to explore various degrees of niche preference by varying \( p \). When \( p = 1 \), the model is identical to Hubbell's neutral model, and in the other extreme case, with \( p = 0 \), all species have their specific niches. We demonstrated that strong niche preference influences the pattern of species abundance (i.e. small \( p \)), showing quite different SADs from that expected under Hubbell's neutral model (i.e., \( p = 1 \)). As shown in Fig. 1, \( S \) is a function of both \( p \) and \( N \) and \( S \) decreases as \( p \) decreases. In the niche site, a dead individual is likely to be replaced by the species that are most abundant in the same niche type or by a generalist species. The preference of species in each niche would limit recruitment of rare species or specialists and induce a reduction in \( S \). In other words, the level of niche overlap among species directly affects the neutrality of the community thereby reducing \( S \). Figure A2 shows that \( S \) increases with increasing \( N \). When there are a large number of niches in the local community, even if each niche has strong species preference, \( S \) is not reduced.

Our niche model was used to evaluate the performance of various neutrality tests (Hubbell's neutral model). We found that all neutrality tests we used here did not always perform very well (Fig. 2A, 3A). This is because the most important parameters, \( \theta \) and \( m \), to characterize the metacommunity, which in turn provides the basis of the local community, are unknown. Therefore, we must estimate them from the local community to be tested. Such conventional treatment likely causes overfitting. Such overfitting has been reported repeatedly by many authors (Chave 2004, Chisholm and Pacala 2010, Volkov et al. 2005) as the main problem in Hubbell's SAD fitting approach, but should also be discussed in other models (or distributions) as well. Next, we investigated the quantitative effects of this parameter on the performance of neutrality tests. As seen in Fig. 2B and 3B, the performance was substantially improved if the true values of \( \theta \) and \( m \) are known, thus indicating the importance of having better estimates for \( \theta \) and \( m \).

Our results suggest that to improve performance, we need to either 1) develop new methods that are more robust to unknowns \( \theta \) and \( m \) or 2) estimate \( \theta \) and \( m \) from data that are independent of the local community to be tested. With regards to 1), along the line of the model-fitting approach, more options for alternative distributions are needed in addition to the commonly used lognormal and log-series distributions. We emphasize this because, occasionally, these distributions alone are insufficient to reject the null neutral model, but other mechanistic models could. Indeed, in our simulation, there are several iterations where our niche model exhibited the best fit (Fig. 2A). It is suggested that if more alternative distributions were available, the ability to reject the null neutral model would be significantly better. With regards to summary statistic-based tests, it is highly desired to develop new summary statistics, for example, those that can simultaneously include information of the species abundance such as species richness, evenness and abundance of rare species. Moreover, statistics that are desired to be robust to \( \theta \) and \( m \) are preferred.

It may be more advantageous, if we can solve problem 2). Unfortunately, it would be extremely difficult to estimate \( \theta \) and \( m \) from data that are independent of the local community to be tested. Ideally, \( \theta \) and \( m \) should be estimated from the metacommunity, but accurately sampling the metacommunity is extremely difficult especially when the scale is unknown. It may also be very difficult to use other data from which \( \theta \) and \( m \) can be estimated (but see the work by Chisholm and Lichstein (2009), which estimated one of the parameters, \( m \), from dispersal data in a local community). However, the question that still remains is what can one do when independent estimates of \( \theta \) and \( m \) are not available? A possibility is to use multiple data sets that may share the same, or at least similar, values of \( \theta \) and \( m \). For example, suppose there are multiple local communities that share a single metacommunity. This is not entirely unrealistic and was suggested by Etienne et al. (2007). Comparing these multiple local communities could provide much more information not only about \( \theta \) and \( m \) in the shared metacommunity but also about the mechanisms that shaped the observed species richness and abundance in each local community. It would also be very useful to have data at multiple time points, even within a single local community (Etienne et al. 2007, Magurran 2007, McGill et al. 2007, Tsai et al. 2015). In addition, any community dynamics data through field observations over multiple years would also be informative. For example, information on which individual was replaced by another individual and at what time point. This multidimensional data requires development of new statistical methods for proper analysis.
In summary, our niche model and simulations provided insight into understanding the observed SADs and fitting to models. We quantitatively demonstrated that it is very difficult to reject Hubbell's neutral model from SAD alone and suggested several ideas to solve this problem (at least partially). While the assumptions of Hubbell's neutral model are too simplistic for some ecologists to accept intuitively, the neutral model can be used as a null model, as it is a good approximation to a neutral community structure (Rosindell et al. 2012). Using Hubbell's model can still be useful as a null model since its rejection indicates that some kind of non-neutral processes is occurring and that new models incorporating such processes could lead to a better understanding of the mechanisms shaping the configuration of the community. In this sense, we would like to again emphasize the importance of developing statistical methods with much higher power than those currently available.

Acknowledgements – We are grateful to J. Rosindell, T. Miki and E. Economo, and members of the Innan Lab for valuable comments and discussions on a previous version of the manuscript. The authors declare no conflict of interest.

References

Pullur, D. W. and Pacala, S. W. 2005. Ecological drift in niche-structured communities: neutral pattern does not imply...


Supplementary material (available online as Appendix oik.01703 at <www.oikosjournal.org/readers/appendix>). Appendix 1.