



Comparison of statistical tests for habitat associations in tropical forests: A case study of sympatric dipterocarp trees in a Bornean forest

Akira Itoh^{a,*}, Tatsuhiro Ohkubo^b, Satoshi Nanami^a, Sylvester Tan^c, Takuo Yamakura^a

^a Graduate School of Science, Osaka City University, Sugimoto 3-3-138, Sumiyoshi-ku, Osaka 558-8585, Japan

^b Department of Agriculture, Utsunomiya University, Utsunomiya 321-0943, Japan

^c Sarawak Forestry Corporation, Sarawak, Malaysia

ARTICLE INFO

Article history:

Received 26 February 2009

Received in revised form 19 October 2009

Accepted 19 October 2009

Keywords:

Habitat association

Lambir Hills National Park

Sarawak

Spatial distribution

Torus shift

Tropical rain forest

ABSTRACT

Habitat difference is an important mechanism for maintenance of tree diversity in tropical forests. The first step in studies of habitat difference is to statistically analyze whether the spatial distributions of tree populations are skewed to species-specific habitats; this is called a habitat association test. We propose a novel habitat association test on the basis of the probability of tree occurrence along a continuous habitat variable. The test uses torus shift simulations to obtain a statistical significance level. We applied this test to 55 common dipterocarp species in a 52-ha plot of a Bornean forest to assess habitat associations along an elevation gradient. The results were compared to those of three existing habitat association tests using the same torus shift simulations. The results were considerably different from one another. In particular, the results of two existing tests using discrete habitat variables varied with differences in habitat definitions, specifically, differences in elevation break points, and the number of habitat classes. Thus, definitions of habitats must be taken into account when habitat association tests with discrete habitat variables are used. Analyses of artificial populations independent of habitat showed that all of the tests used were robust with respect to spatial autocorrelation in tree distributions, although one existing test had a higher risk of Type I errors, probably due to the use of multiple tests of significance. Power analysis of artificial populations in which distributions were skewed to certain elevations showed that the novel test had comparable statistical power to the most powerful existing test. Statistical power was affected not by the total number of a given tree but by the number of clumps in a plot, suggesting that >5 clumps were required for a reliable result.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Habitat niche difference is an important mechanism for maintaining local species diversity in tropical tree communities (Davies et al., 2005; Gunatilleke et al., 2006; John et al., 2007) among other mechanisms, such as trade-offs between colonization and competition (Coomes and Grubb, 2003), maximum size and recruitment (Kohyama, 1993), stress-tolerance and growth (Wright et al., 2003), conspecific density-dependent mortality (Janzen, 1970; Connell, 1971), and stochastic processes (Hubbell, 2001). These mechanisms are not mutually exclusive and do function simultaneously in a community. It is therefore significant to quantify the relative importance of each mechanism within and among various tropical forests. One possible way to evaluate the relative importance of habitat niche difference is to calculate the

proportion of species that show statistically significant habitat associations. We may expect that more species show significant habitat associations in communities where habitat difference plays a larger role. We should note that the proportion will depend on the alpha-level set for the significance test.

Habitat associations of individual species do not necessarily indicate habitat difference among species, since these species possibly associated with the same habitat (Noguchi et al., 2007). Thus, analysis of species habitat associations is only the first step in quantitative analysis of habitat difference. Although we are cognizant of this issue, habitat association tests are discussed in this paper because we believe that the proportion of habitat specialist species strongly supports the potential role of habitat difference in a community, and because appropriate habitat association tests are useful as a first step in the analysis of habitat difference.

In using habitat association tests, we often face two major problems: (1) spatial autocorrelation and (2) definition of habitats. The problem of spatial autocorrelation is due to the spatially

* Corresponding author. Tel.: +81 6 66053167; fax: +81 6 66053167.

E-mail address: itoha@sci.osaka-cu.ac.jp (A. Itoh).

aggregated distributions observed for many tropical tree species (Condit et al., 2000). Most conventional statistical tests, such as the Chi-squared test, are invalid for habitat association analysis of spatially aggregated populations (Legendre, 1993; Itoh et al., 2006). For example, the conventional Chi-squared test determines whether the numbers of trees observed in different habitats are consistent with the null hypothesis that the tree distribution is independent of habitat differences if complete spatial randomness (CSR) applies. The null hypothesis is rejected either when the tree distribution does not follow CSR or when the tree distribution is affected by habitat differences. Therefore, the null hypothesis for autocorrelated tree distributions may be rejected because they do not follow CSR, even if there is no significant relationship between the habitat and tree distribution.

For a proper habitat association analysis, it is better to use the following null hypothesis: the observed relationship between the habitat and tree distribution is consistent with a “stationary” distribution model that generates spatial patterns, which are caused by factors independent of habitat. The term “stationary” here means that the expected point density is the same at any location within the study area, and therefore the expected tree density is similar in any habitat (no habitat associations). CSR is a stationary point process with no spatial autocorrelation. However, tree distributions are often spatially autocorrelated at least partly due to factors independent to habitat, such as limited seed dispersal. We therefore need to use stationary point processes that have spatially autocorrelated distributions. We should note that we assume that spatial autocorrelation is caused only by factors independent to habitat under the null hypothesis. However, spatial autocorrelation are often caused by both habitat difference and limited dispersal. It is impossible to separate effects of these factors only from spatial patterns. Therefore, the assumption in the null hypothesis may underestimate the effect of habitat difference in structuring the spatial patterns of tree populations.

A general procedure for habitat association tests is as follows:

1. Select an appropriate stationary point process model that results in autocorrelated patterns similar to the observed tree distribution.
2. Simulate tree distributions many times using the selected point process model.
3. Calculate certain statistics indicating the relationships between habitat and tree distribution, such as the Chi-squared value, based on simulated tree distributions and the observed habitat distribution map.
4. Compare the observed and simulated statistics to estimate the probability that the observed value is obtained by the stationary model.

To generate stationary and autocorrelated distributions, some recent studies on tropical trees have adopted the Poisson cluster model (e.g., Plotkin et al., 2000; John et al., 2007), while others used the torus shift (toroidal shift) simulations (e.g., Webb and Peart, 2000; Harms et al., 2001; Itoh et al., 2003; Gunatilleke et al., 2006). Plotkin et al. (2000) demonstrated that the Poisson cluster model successfully simulated tree distributions that were similarly autocorrelated to those observed for most populations in a 50-ha plot of a tropical rain forest at Pasoh, Malaysia. For the Poisson cluster model, values of model parameters need to be estimated from observed tree distributions. The torus shift is technically easier to use because estimation of parameters is not necessary. In the torus shift, an original tree map for a rectangular plot is shifted randomly on a two-dimensional torus. The torus shift retains the overall autocorrelated patterns, except for trees that cross the plot edges. The expected tree density is the same at any location within the plot due to the torus shift, hence it is stationary (Fortin and

Dale, 2005). Note that there are two theoretically equivalent methods for torus shift simulation: in the first, the tree map shifts and the habitat map is held as the original; in the other, the habitat map shifts and the tree map is retained. In this study, we applied a torus shift of the habitat map to obtain null relationships between the trees and habitats. This was because the spatial resolution was larger for the habitat map (20 m) than tree distributions (0.1 m) in our study. Torus shifts of tree distributions by <20 m are not equivalent to those >20 m because the former torus shifts maintain many trees within the original habitat. By shifting the habitat map by 20-m steps, we avoided torus shifts of <20 m.

Habitat definitions are also critical in habitat association tests because different habitat criteria can produce different results. Many previous habitat association studies have used a priori fixed discrete habitats (e.g., Plotkin et al., 2000; Webb and Peart, 2000; Harms et al., 2001; Itoh et al., 2003; Davies et al., 2005; Gunatilleke et al., 2006), and two have used continuous habitat variables (Itoh et al., 2006; John et al., 2007). Boundaries between habitats have often been defined by selecting break points for continuous habitat variables such as elevation, water content and soil fertility in a more or less arbitrary manner. However, it is sometimes difficult to define appropriate break points for continuous habitat variables. Inadequate break points and/or an inadequate number of habitat categories may produce misleading results if the manner of habitat classification significantly affects the habitat association tests. However, it is not yet clear how the number and values of the break points selected affect the results of habitat association tests.

In this paper, we propose a novel habitat association test based on a continuous habitat variable. We compared this test to three previously published habitat association tests using a common dataset, specifically, distributions of Dipterocarpaceae species along an elevation gradient within a 52-ha plot of a Bornean rain forest. Of the four statistical tests compared, two use continuous habitat variables and the other two use discrete variables. We analyzed how the results of tests using discrete variables are affected by differences in the break points and number of habitats. Finally, we compared the statistical power and the probability of Type I errors (i.e., the probability that habitat generalists are judged wrongly as specialists) among the four tests, using simulated tree distributions.

2. Materials and methods

2.1. Tree distribution data

Tree distribution data were obtained within a 52-ha permanent plot (1040 × 500 m) established in 1992 in a mixed dipterocarp forest located within the Lambir Hills National Park (4°12'N, 114°00'E; c. 60–450 m a.s.l.) in Sarawak, Malaysia. The average annual rainfall is 2725 mm (1967–1998) at Miri Airport, approximately 20 km north of Lambir. There is no distinct dry season, although the period February–August has relatively less rainfall. The topography of the study site is hilly, with undulating terrain with steep slopes, including scars caused by small landslides (Ohkubo et al., 2007). All trees ≥ 1 cm diameter at breast height (dbh; 1.3 m) in the plot were individually labeled, mapped, identified to species, and measured for dbh. Elevations were measured at 20 × 20-m grid points within the plot (Yamakura et al., 1995). Tree enumerations have been conducted four times, in 1992, 1997, 2002, and 2007–2008. In this study, we used data for 55 common Dipterocarpaceae species ($n \geq 100$ in the plot) obtained in 1997.

2.2. Statistical tests of habitat association

All of the tests use the torus shift simulations to obtain significance levels, but each test applies a different statistic to

evaluate the degree of habitat association. In this paper, we refer to these tests as the “Chi-squared test,” “Tree-density test,” “SD test,” and “Adjusted-SD test” based on the statistics used in the tests. The latter test is the novel test proposed in this study. The Chi-squared and Tree-density tests use discrete habitat variables, while the SD and Adjusted-SD tests use continuous habitat variables.

For the torus shift simulations, we used an elevation map with 1300 20×20 -m squares within the 52-ha study plot. The map was shifted along the x - and/or y -coordinates by 20-m steps on a two-dimensional torus, for the 1300 possible combinations. We also used mirror and rotated maps of each shifted image following Harms et al. (2001). As a result, we generated a total of 5200 habitat maps, including the original map. The test statistics for each of the habitat tests were then calculated for all of the shifted maps as a null distribution of the statistics.

The Chi-squared test uses the statistic χ^2 , which is the same as that used in the conventional Chi-squared test:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}, \quad (1)$$

where E_i and O_i are the expected and observed numbers of trees in the i th habitat (elevation class in this study), respectively, and n is the total number of categories for the discrete habitat variable. The value of E_i is calculated from the total tree number of a focal species in the study plot multiplied by the proportion of the area of the i th habitat in the plot. Habitat association is considered to be statistically significant ($p < 0.05$) if the observed χ^2 -value is within the largest 5% of χ^2 -values obtained by the torus shift simulations. In this paper, we refer to a species that shows a significant habitat association as “a specialist species” or merely “a specialist,” and one that shows non-significant habitat association as “a generalist species” or “a generalist.” Examples of the use of this test can be found in Cannon and Leighton (2004) and Noguchi et al. (2007).

The Tree-density test is a modified version of the Torus-translation test of Harms et al. (2001). Harms et al. used the relative number of trees (the proportion of trees of a focal species to the total number of trees) observed in each category of discrete habitat as the test statistic. The Tree-density test uses the actual numbers of trees rather than relative ones, because the actual number of trees is equivalent to O_i in Eq. (1), which allows easy comparison to the other tests. The observed number of trees is compared to that obtained by the torus shift simulations for each habitat class. A species is considered a specialist if the observed number of trees is either larger than the largest 2.5% of values obtained by torus shift simulation or smaller than the smallest 2.5% of values in at least one of the habitat classes used. Because this test applies the significance test for each habitat class, it may suffer from multiple testing which inflates the probability of Type I errors. Examples of the use of this test can be found in Harms et al. (2001) and Gunatilleke et al. (2006).

The SD test uses the standard deviation (SD) of a continuous habitat variable (elevation in this study). A value for the habitat variable is assigned to each individual tree within a study plot and the SD is then calculated for each species. A species is considered a specialist if the observed SD is smaller than the smallest 5% of values of the simulations. Previous studies using this test (e.g., Enoki and Abe, 2004; Itoh et al., 2006; John et al., 2007) adopted the species mean as well as the SD as test statistics. However, for the purposes of analyzing habitat associations, evaluation of the SD is sufficient and using two statistics in one test invokes the problem of multiple tests. Thus, we used only the SD in our test.

2.3. Adjusted-SD test

Here, we propose a novel habitat association test, the Adjusted-SD test. In the SD test, the SD for a focal species is calculated based

on the observed frequency of habitat variables for individual trees. Therefore, the value of the SD should be affected by the frequency distribution of a habitat variable in the study plot. For example, the SD would become smaller if a large area of the plot was occupied by one habitat, as opposed to the case in which all habitats are distributed evenly within the plot, even if there is no relationship between tree distribution and habitat. To address this problem, we adjusted the SD based on the conditional probability of tree occurrence (E) given a habitat variable x , notated as $p(E|x)$. For example, when we use elevation for the habitat variable and select a point randomly within the study plot, $p(E|x)$ represents the probability that a focal species exists at the selected point when the elevation of the point is x m. We can use $p(E|x)$ as an index of habitat niche of a species. Values of $p(E|x)$ are not affected by differences in habitat distribution within a study plot, although the probability that the randomly selected point is at x m should vary among plots with different habitat distributions. We therefore calculated the SD of $p(E|x)$, or Adjusted-SD, as the test statistic.

We can estimate $p(E|x)$ from the field data on the habitat structure and tree distributions within a study plot. The conditional probability $p(E|x)$ is:

$$p(E|x) = \frac{p(x|E)p(E)}{p(x)}, \quad (2)$$

where $p(x|E)$ is the conditional probability of x when a tree exists at the selected point, $p(E)$ is the probability that a tree exists at the point, and $p(x)$ is the probability that the point is within habitat x . We can think of $p(E)$ as a constant in terms of x , because:

$$p(E) = \int_{x_{\min}}^{x_{\max}} p(E|x)p(x)dx, \quad (3)$$

where x_{\min} and x_{\max} are the minimum and maximum elevations in the plot. Therefore:

$$p(E|x) \propto \frac{p(x|E)}{p(x)} \quad (4)$$

We can determine the shape of $p(E|x)$ if we know the density functions of $p(x|E)$ and $p(x)$.

When we use discrete habitat variables, $p(x|E)/p(x)$ in Eq. (4) is proportional to the tree density at habitat x , hence $p(E|x)$ is also proportional to the tree density. Therefore, this method is conceptually similar to those using the weighted mean of the tree occurrence over different discrete habitat types, widely used in vegetation analysis. However, for continuous habitat variables, more complex methods to estimate $p(E|x)$ are required.

We used kernel-based nonparametric probability density estimation (Silverman, 1986) to estimate $p(x|E)$ and $p(x)$. The observed frequency distribution of elevation at the 20×20 -m grid points was used to estimate $p(x)$. The frequency distribution of elevations where trees of a focal species existed within the 52-ha study plot was used to estimate $p(x|E)$. The estimation was carried out using the “density” function in the R-package “stats” ver. 2.9.1 in R ver. 2.9.1 (R Development Core Team, 2009). We used a Gaussian kernel window with a bandwidth parameter ($bw = “nr0”$), which follows Silverman’s “rule of thumb” (Silverman, 1986). To handle the boundary problem, we adopted Schuster’s (1985) mirror image correction. We created mirror images of the data on the outer side of the upper and lower bound of x for the habitat and tree data. After applying the R “density” function on the created data, we re-normalized the tails of $p(x|E)$ and $p(x)$ numerically using values at 512 equally spaced points from the maximum and minimum elevations. Although the mirror boundary correction is not perfect (Cowling and Hall, 1996), it may be a reasonable assumption that tree density is similar on both

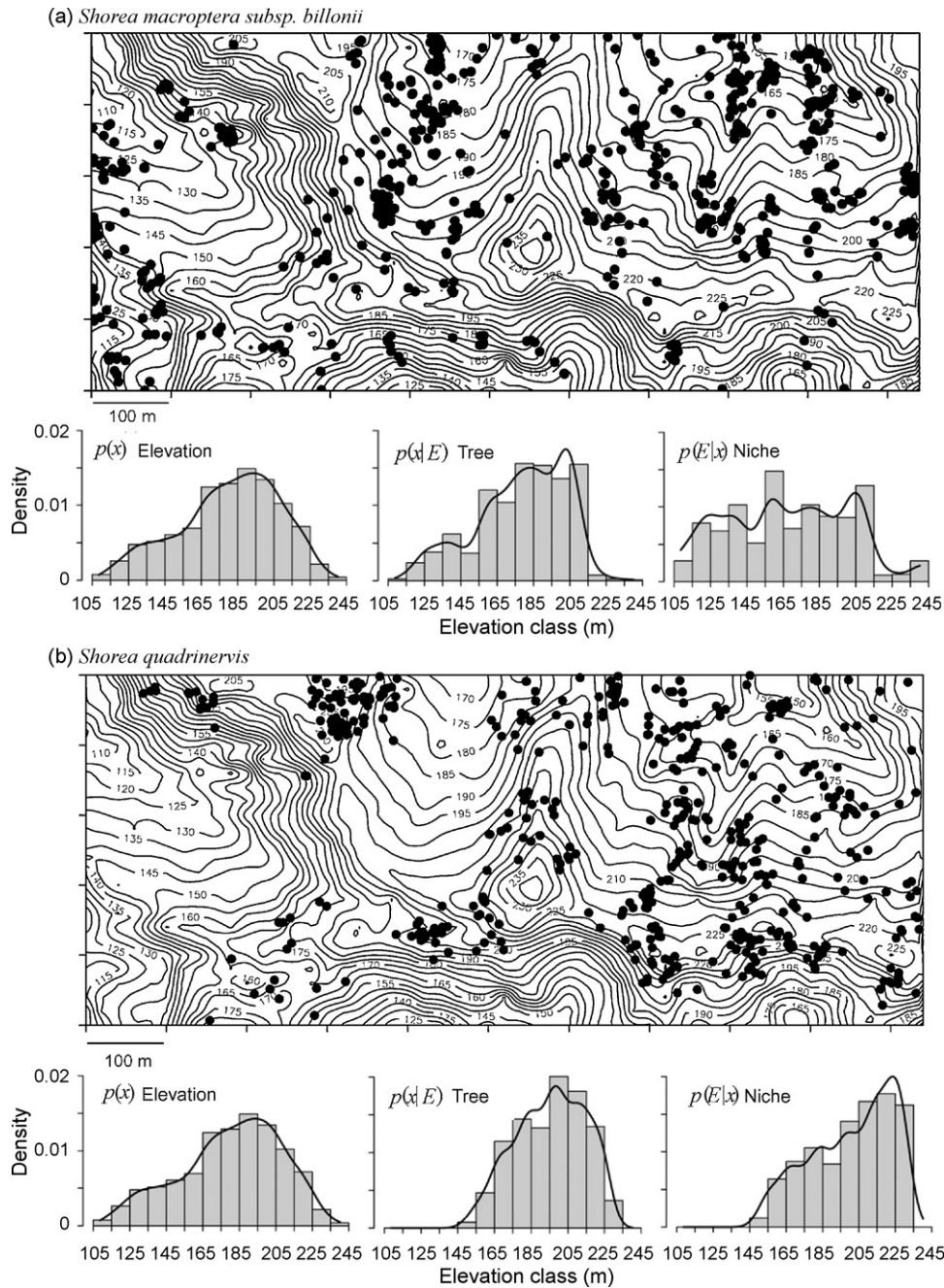


Fig. 1. Two examples of tree distribution maps in a 52-ha Lambir plot and functions of the Adjusted-SD test along an elevation gradient. See text for details of the functions, $p(x)$, $p(x|E)$ and $p(E|x)$.

sides of a boundary, at least near the boundary. The boundary correction resulted in only small changes for our observed and simulated dataset; values of $p(x|E)$ became slightly higher after the boundary correction when most trees were aggregated near the lowest or highest elevations. Comparison to preliminary analyses without the boundary correction indicated that the results were generally very similar with and without the boundary correction, suggesting a limited boundary problem at least in our case.

For convenience in among-species comparison using $p(E|x)$ graphics, we also re-normalized $p(E|x)$ numerically using estimated values of $p(E|x)$ at 512 equally spaced points to satisfy the equation:

$$\int_{x_{\min}}^{x_{\max}} p(E|x) dx = 1. \quad (5)$$

Examples of tree distributions and estimated density functions are shown in Fig. 1. For *Shorea macroptera* subsp. *bilonii* (Fig. 1a), trees were abundant at intermediate elevations; hence, $p(x|E)$, labeled as “Tree” in Fig. 1, had higher values at intermediate elevations. However, this was mostly because these elevations were more abundant in the study plot [see $p(x)$ in Fig. 1a]. Thus, $p(E|x)$, labeled as “Niche”, was nearly flat at a large elevation range, indicating very weak or no habitat association. Fig. 1b shows the results for *Shorea quadrinervis*, which had a clearly unimodal shape for $p(E|x)$, indicating a strong habitat association with higher elevations.

Because a habitat specialist may have a higher probability of $p(E|x)$ within a certain range of x than in other habitats, such as in the case of *S. quadrinervis* (Fig. 1b), we would expect a small SD of $p(E|x)$ for a specialist species. Thus, we used the SD of $p(E|x)$ as the

statistic for the significance test. A species was considered a specialist if the observed SD was smaller than the smallest 5% of values obtained by torus simulation.

2.4. Effect of habitat classifications

To evaluate the effects of different habitat classifications, we compared the results of the Chi-squared and Tree-density tests for various habitat classifications. First, we divided the plot into two elevation habitats. We used eight different elevations as the break point: 140, 150, 160, 170, 180, 190, 200, and 210 m. Second, we divided the plot into various numbers of even-interval elevation classes: 2, 3, 4, 5, 6, 7, 8, and 9 classes. Thereafter, we applied the Chi-squared and Tree-density tests for each classification and compared the number of specialist species identified by the results. We also calculated Jaccard similarity indices for the specialists for all classification pairs to evaluate how the species compositions of the specialists varied among the different habitat classifications.

2.5. Statistical power

To compare the risk of Type I errors, or the risk of incorrectly identifying a generalist as a specialist, we applied the four tests to artificial generalists. Generalist populations were created using the Poisson cluster model (Neyman and Scott, 1958; Plotkin et al., 2000). First, we randomly selected x - y coordinates for parent points using a Poisson process from the study plot with intensity ρ (ha^{-1}). Second, each parent produced “offspring,” the number of which was determined by a Poisson process with intensity m (per parent). The positions of the offspring relative to the parent were derived from random points following a radially symmetric two-dimensional normal distribution with variance s^2 (m^2). The offspring simulation was applied on a two-dimensional torus to avoid offspring points being located outside of the plot. Finally, we deleted the parent points. This procedure generates clumped distributions. A larger ρ results in a larger number of clumps, and a larger m results in a larger number of points in each clump. A larger s^2 results in a larger clump area, and therefore a less clumped distribution. The Poisson cluster model produces aggregated but non-habitat-dependent distributions (Plotkin et al., 2000). Therefore, this simulation can create artificial populations of generalist species with various individual numbers and clumping intensities.

To evaluate the effects of clump size and clump number on the habitat test results, we selected 36 combinations of six values of ρ (0.05, 0.1, 0.2, 0.4, 0.8, and 1.6 ha^{-1}) and s^2 (10, 50, 100, 500, 1000, and 5000 m^2). The value of m was chosen as $4/\rho$, which held the mean total number of points constant at 4 ha^{-1} (see Appendix A for examples). We simulated 100 populations for each parameter combination (3600 populations in total), and then counted the number of populations that were incorrectly judged as specialists (Type I error) using each statistical method.

To evaluate the statistical power of the four tests, we analyzed 100 artificial specialist populations and counted the number of cases that were correctly judged as specialists as an index of statistical power. The specialist populations were produced using a modified Poisson cluster model; in this model, parent points were not chosen randomly, but rather were selected so that their distributions were skewed to a certain elevation. To generate parent points, we first sampled 20×20 -m squares equal to the target parent number from the 1300 squares in the plot, with replacement that depended on the following probability:

$$N(x_i|M, SD^2) \times p(x_i), \quad (7)$$

where x_i is the mean elevation of the i th square, $N(x_i|M, SD^2)$ is the density of x_i in a normal distribution with a mean equal to M (m)

and a standard deviation equal to SD (m), and $p(x_i)$ is the probability of the i th elevation in the plot calculated by $p(x)$ in Eq. (2). Then, the position of the parent point (x - and y -coordinates) within the sampled square was randomly chosen. This procedure is equivalent to requiring the $p(E|x)$ of parent points to follow a normal distribution with mean = M and standard deviation = SD . The degree of specialization was determined by changing the SD value; a smaller SD represents a stronger habitat specialization. For each population, a value of M was randomly chosen within the range of elevation in the study plot (180–245 m). The sampling of squares was conducted using the “sample” function in the R-package “base” (ver. 2.9.1). In the offspring simulations, we did not apply a two-dimensional torus; offspring points falling outside of the plot were discarded, because the torus method produced some points at elevations greatly different from those of the parents.

To determine the statistical power with the degree of specialization, we used six SD values: 10, 20, 30, 40, 50, and 60 m. For each SD value, we used various combinations of parameters (m , s^2 , and ρ) of the Poisson cluster model to evaluate the effects of sample size and spatial aggregation. We changed one of the three parameters ($m = 2.5, 5, 10, 20, 40,$ and 80 per parent; $s^2 = 10, 50, 100, 500, 1000,$ and 5000 m^2 ; $\rho = 0.05, 0.1, 0.2, 0.4, 0.8,$ and 1.6 ha^{-1}) while holding the values of the other parameters constant ($m = 20$; $s^2 = 500$; $\rho = 0.2$). In total, we simulated 10,800 specialist populations (see Appendices B–D for examples).

3. Results and discussion

3.1. Habitat associations for Dipterocharaceae

Of the 55 species analyzed, 28 (50.9%) and 34 (61.8%) were judged as habitat specialists by the SD and Adjusted- SD tests, respectively (Table 1). We should note that the number of specialists and generalists will change with the alpha-level used in the significance test. For example, when we set the alpha-level as 0.01 instead of 0.05, only 12 and 13 species were specialists by the SD and Adjusted- SD tests, respectively. All of the specialist species identified by the SD test were judged as specialists by the Adjusted- SD test. Six species were judged as specialists only by the Adjusted- SD test; most of them had low mean elevations (Fig. 2). This was probably because only a small proportion of the study plot was occupied by lower elevations (Fig. 1). The difference in the SD between the SD and Adjusted- SD tests is larger for species whose distribution is aggregated to less abundant habitats by definition (Eq. (2)).

The number of specialist species determined by the Chi-squared and Tree-density tests varied depending both on the break point for the two elevation classes and on the number of elevation classes (Fig. 3). The number of specialist species was largest for the break point elevations of 150 and 160 m in the Tree-density test (Fig. 3a). Break points at high elevation, i.e., 200 and 220 m, had a small number of specialists in both tests. The numbers of specialist species were fairly constant (32–36 species) in the Tree-density test for various numbers of habitats, except for the two habitat categories for which only 17 species were judged as specialists (Fig. 3b). The number of specialists in the Chi-squared test decreased with an increasing number of habitat classes. For all break point elevations and habitat numbers other than the two habitats with a break point elevation of 170 m, fewer species were judged as specialists by the Chi-squared test than by the Tree-density test.

The species composition of the specialists also varied among different habitat categorizations. There were only five and three species that were judged as specialists for all habitat definitions (16 cases) in the Tree-density and Chi-squared tests, respectively. The means of the Jaccard similarity indices between pairs of

Table 1
Results of habitat association analysis of 55 common dipterocarp species at Lambir, Sarawak, by the SD and Adjusted-SD tests along an elevation gradient.

Species	No. trees (/52-ha)	SD test	Adjusted-SD test	Statistics of $p(E x)$		
				Mean (m)	SD (m)	Mode (m)
<i>Anisoptera grossivenei</i>	153	*	*	184.3	23.9	169.7
<i>Cotylelobium melanoxydon</i>	929	*	*	207.5	22.0	240.3
<i>Dipterocarpus confertus</i>	291	ns	ns	168.1	31.4	203.7
<i>Dipterocarpus crinitus</i>	409	*	*	206.8	20.1	225.3
<i>Dipterocarpus geniculatus</i>	361	ns	*	192.2	25.0	215.3
<i>Dipterocarpus globosus</i>	3,094	*	*	204.2	25.1	224.0
<i>Dipterocarpus kunstleri</i>	118	*	*	129.5	4.7	131.0
<i>Dipterocarpus palembanicus</i> subsp. <i>bornensis</i>	546	*	*	194.5	18.5	205.7
<i>Dipterocarpus palembanicus</i> subsp. <i>palembanicus</i>	251	ns	ns	141.5	24.5	108.6
<i>Dryobalanops aromatica</i>	10,720	*	*	196.2	23.6	202.9
<i>Dryobalanops lanceolata</i>	920	ns	*	128.7	14.8	108.6
<i>Hopea beccariana</i>	193	ns	ns	179.9	43.4	236.9
<i>Hopea bracteata</i>	405	*	*	174.5	20.3	191.3
<i>Hopea dryobalanoides</i>	839	ns	ns	139.4	14.3	147.5
<i>Hopea mesuoides</i>	665	*	*	126.0	9.8	108.6
<i>Hopea pterygota</i>	1,065	ns	*	132.7	15.7	130.0
<i>Parashorea parvifolia</i>	449	ns	*	130.1	19.7	110.7
<i>Shorea acuta</i>	1,052	*	*	202.8	25.9	227.9
<i>Shorea amplexicaulis</i>	2,631	*	*	192.1	26.3	190.5
<i>Shorea asahi</i>	157	ns	ns	217.6	19.1	226.3
<i>Shorea atrinervosa</i>	235	ns	ns	171.3	19.1	168.4
<i>Shorea biawak</i>	3,798	*	*	199.3	22.4	206.5
<i>Shorea curtisii</i>	174	*	*	192.5	24.5	176.1
<i>Shorea exelliptica</i>	100	ns	ns	180.3	26.8	195.2
<i>Shorea faguettiana</i>	117	ns	ns	186.1	21.9	194.4
<i>Shorea falciferoides</i>	466	ns	ns	154.2	32.1	135.9
<i>Shorea ferruginea</i>	172	ns	ns	166.1	25.0	167.4
<i>Shorea flemmichii</i>	192	*	*	183.2	22.8	164.5
<i>Shorea geniculata</i>	519	ns	ns	172.6	7.1	171.7
<i>Shorea havilandii</i>	837	*	*	183.2	24.6	188.5
<i>Shorea kunstleri</i>	186	*	*	200.2	28.4	230.7
<i>Shorea laxa</i>	3,066	*	*	213.6	19.3	216.6
<i>Shorea macrophylla</i>	104	ns	*	129.5	7.4	126.7
<i>Shorea macroptera</i> subsp. <i>billonii</i>	788	ns	ns	168.1	31.9	204.7
<i>Shorea macroptera</i> subsp. <i>macropterifolia</i>	1,440	*	*	158.1	31.6	123.8
<i>Shorea multiflora</i>	322	ns	ns	207.9	30.3	240.3
<i>Shorea ovalis</i>	144	ns	ns	193.9	21.7	206.3
<i>Shorea ovata</i>	440	*	*	215.0	19.1	222.2
<i>Shorea parvifolia</i>	511	ns	ns	166.1	33.8	108.6
<i>Shorea patoiensis</i>	772	ns	ns	151.7	29.8	124.6
<i>Shorea polyandra</i>	283	ns	ns	166.8	31.3	172.3
<i>Shorea pubistyle</i>	343	*	*	187.1	22.7	194.4
<i>Shorea scrobiculata</i>	453	ns	ns	189.3	28.6	203.9
<i>Shorea almon</i>	124	ns	ns	179.6	32.0	203.7
<i>Shorea ochraceae</i>	125	ns	ns	157.3	31.0	178.7
<i>Shorea quadrinervis</i>	513	ns	*	200.6	23.0	223.5
<i>Shorea rubella</i>	283	*	*	166.8	31.3	172.3
<i>Shorea xantophylla</i>	290	*	*	134.0	14.0	130.0
<i>Vatica badifolia</i>	1,017	*	*	188.7	24.3	189.3
<i>Vatica micrantha</i>	6,153	*	*	211.6	19.8	223.8
<i>Vatica nitens</i>	174	*	*	192.5	24.5	176.1
<i>Vatica</i> cf. <i>oblongifolia</i>	113	*	*	181.1	20.3	170.5
<i>Vatica oblongifolia</i> sbsp. <i>crassibata</i>	159	*	*	175.5	21.0	168.4
<i>Vatica oblongifolia</i> sbsp. <i>oblongifolia</i>	775	ns	ns	172.1	34.4	200.1
<i>Vatica sarawakensis</i>	293	*	*	174.9	30.7	171.0

* Stands for a specialist at $p < 0.05$; ns $p > 0.05$.

different break point elevations ($n = 28$ pairs) were 0.527 (range: 0.161–0.947) and 0.499 (range: 0.143–0.900) for the Tree-density and Chi-squared tests, respectively (Table 2). This indicates that, on average, only 50% of the specialist species were common among habitat classifications with different break points. The mean Jaccard indices between different habitat numbers were 0.647 (0.650–0.900) and 0.771 (0.472–0.795) for the Tree-density and Chi-squared tests, respectively (Table 3). These similarities were higher than for different break points; however, 23–35% of the specialist species still were not shared between pairs of different habitat classifications.

The numbers of specialists identified by the Tree-density tests were similar to those identified by the Adjusted-SD test (34 species) for >3 habitat categories (Fig. 3b). However, of the 34

specialist species identified by the Adjusted-SD test, 9–11 species (26–31%) were judged as generalists by the Tree-density test, indicating a large difference in the species composition of specialists between the Tree-density and Adjusted-SD tests.

These results clearly show that the way we define discrete habitats within a study area greatly affects the results of habitat association analysis. Therefore, care must be used in justifying habitat definitions when we divide continuous habitats into discrete habitats for habitat association analysis.

3.2. Probability of Type I errors

Fig. 4 shows the proportions of Type I errors for the four habitat association tests based on simulations with artificial generalist

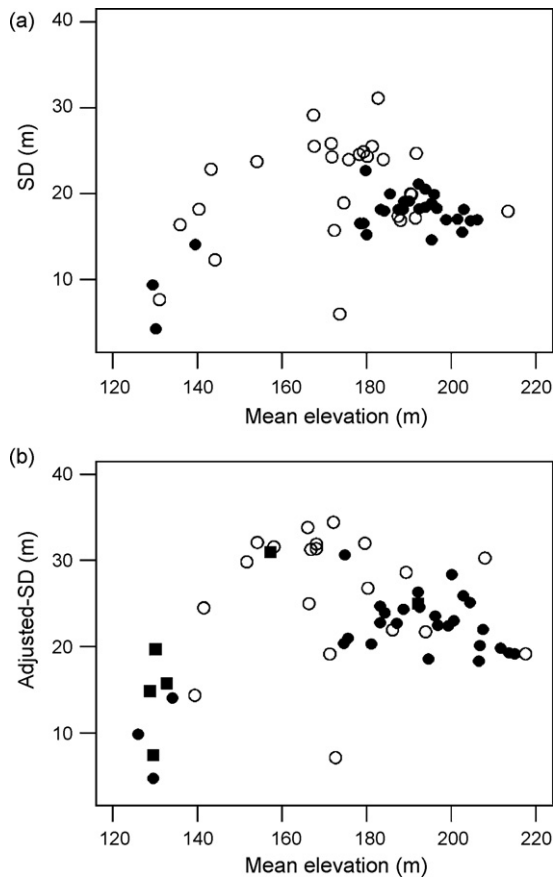


Fig. 2. Relationships between standard deviation (SD) and mean in (a) the SD test and (b) the Adjusted-SD test for 55 dipterocarp species along an elevation gradient in a 52-ha Lambir plot. Open and filled symbols are generalist and specialist species, respectively. Filled squares in (b) stand for the species that were specialists by the Adjusted-SD test but generalist by the SD test.

distributions with various degrees of clumping. The results of conventional Chi-squared tests (Fig. 4e) were also shown to assess the effectiveness of the torus shift. For the conventional Chi-squared test, the probability of Type I error increased with decreasing s^2 values, indicating that generalists were wrongly judged as specialists more often when the degree of clumping was higher. Nearly all populations with s^2 of 10–50 m² were judged as specialists. No inflation in the number of Type I errors was observed in the other four tests, even for highly clumped

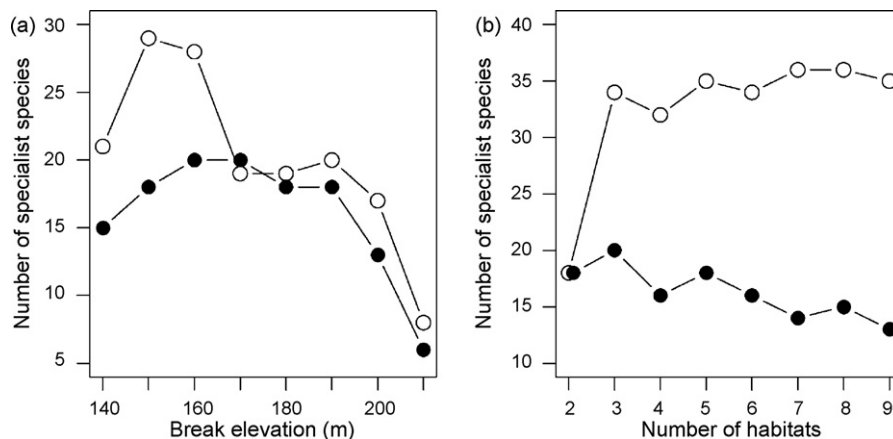


Fig. 3. Numbers of specialist species judged by the Tree-density (closed circles) and Chi-squared (open circles) tests for 55 dipterocarp species with various habitat definitions: (a) two elevation classes with variations in the break elevations and (b) various numbers of regular elevation classes.

distributions, indicating that the torus shift procedure successfully overcame the spatial autocorrelation problem.

The mean proportions of Type I errors were slightly less than 5% in all tests other than the Tree-density test (8.6%). This was reasonable because we adopted $p < 0.05$ for significance tests. The larger Type I errors in the Tree-density test may have occurred because this test repeated multiple significance tests for each population (see Section 2). Thus, we need to bear in mind that the Tree-density test may slightly overestimate the statistical significance of habitat associations, especially when we use many habitat classes.

3.3. Statistical power

Fig. 5 shows the results of the power analysis. The order of statistical power among the four tests was Tree-density > Adjusted-SD > SD > Chi-squared, based on the mean values of statistical power for the same artificial dataset. The higher statistical power of the Tree-density test may be due, at least in part, to its multiple significance tests.

The values of m and s^2 had very little effect on the statistical power in any test, while smaller ρ values reduced the statistical power of each test (Fig. 5). Thus, all of the tests were robust in relation to differences in clump size and tree density within a clump, but were affected by the number of clumps within the plot. Only 40–44% of the cases were successfully judged as specialists for the most specialized populations ($SD = 10$ m) when ρ was 0.05 ha⁻¹ or 2.6/52-ha plot. Thus, we cannot expect sufficient power to detect habitat associations if the number of clumps within a study plot is too small, and should determine the number of clumps rather than the number of individual trees before analyzing habitat associations. Results may not be reliable if there are very few clumps in the study plot, even if there are many individuals.

Accordingly, we evaluated the number of clumps for our study populations. We assumed that the distributions of the study species followed the Poisson cluster model (Neyman and Scott, 1958; Plotkin et al., 2000). Model parameters were estimated for each species based on the Reply’s K -statistic, $K(d)$, expected from the Poisson cluster model according to the following equation (Diggle, 1983; Cressie, 1991; Plotkin et al., 2000):

$$K(d) = \pi d^2 + \rho^{-1} \left(1 - \exp\left(-\frac{d^2}{4s^2}\right) \right), \tag{8}$$

where d is the distance (m) and ρ and s^2 are parameters of the Poisson cluster model. For estimation of the parameters, we used the “pcp” function in the R-package “splancs” ver. 2.01.

Table 2
Jaccard similarity indices between specialist species of two discrete habitat classes with various break points of elevation. The values above and below the diagonal are indices of the Tree-density and Chi-square tests, respectively.

Break elevation (m)	Break elevation (m)							
	140	150	160	170	180	190	200	210
140		0.786	0.690	0.640	0.667	0.519	0.407	0.200
150	0.833		0.833	0.567	0.586	0.469	0.375	0.161
160	0.667	0.727		0.643	0.608	0.484	0.387	0.167
170	0.591	0.583	0.818		0.947	0.727	0.591	0.227
180	0.571	0.565	0.810	0.900		0.762	0.619	0.238
190	0.391	0.400	0.542	0.609	0.667		0.842	0.286
200	0.227	0.250	0.391	0.391	0.429	0.706		0.333
210	0.167	0.143	0.238	0.238	0.263	0.353	0.500	

Table 3
Jaccard similarity indices between specialist species of discrete habitats with various numbers of elevation classes. The values above and below the diagonal are indices of the Tree-density and Chi-square tests, respectively.

Number of elevation classes	Number of elevation classes							
	2	3	4	5	6	7	8	9
2		0.529	0.500	0.486	0.472	0.486	0.486	0.500
3	0.652		0.634	0.690	0.725	0.614	0.690	0.795
4	0.889	0.714		0.707	0.700	0.707	0.795	0.605
5	0.714	0.900	0.789		0.674	0.682	0.682	0.698
6	0.700	0.800	0.778	0.889		0.674	0.714	0.732
7	0.684	0.700	0.765	0.778	0.875		0.762	0.622
8	0.737	0.750	0.824	0.833	0.824	0.706		0.738
9	0.722	0.650	0.813	0.722	0.706	0.800	0.867	

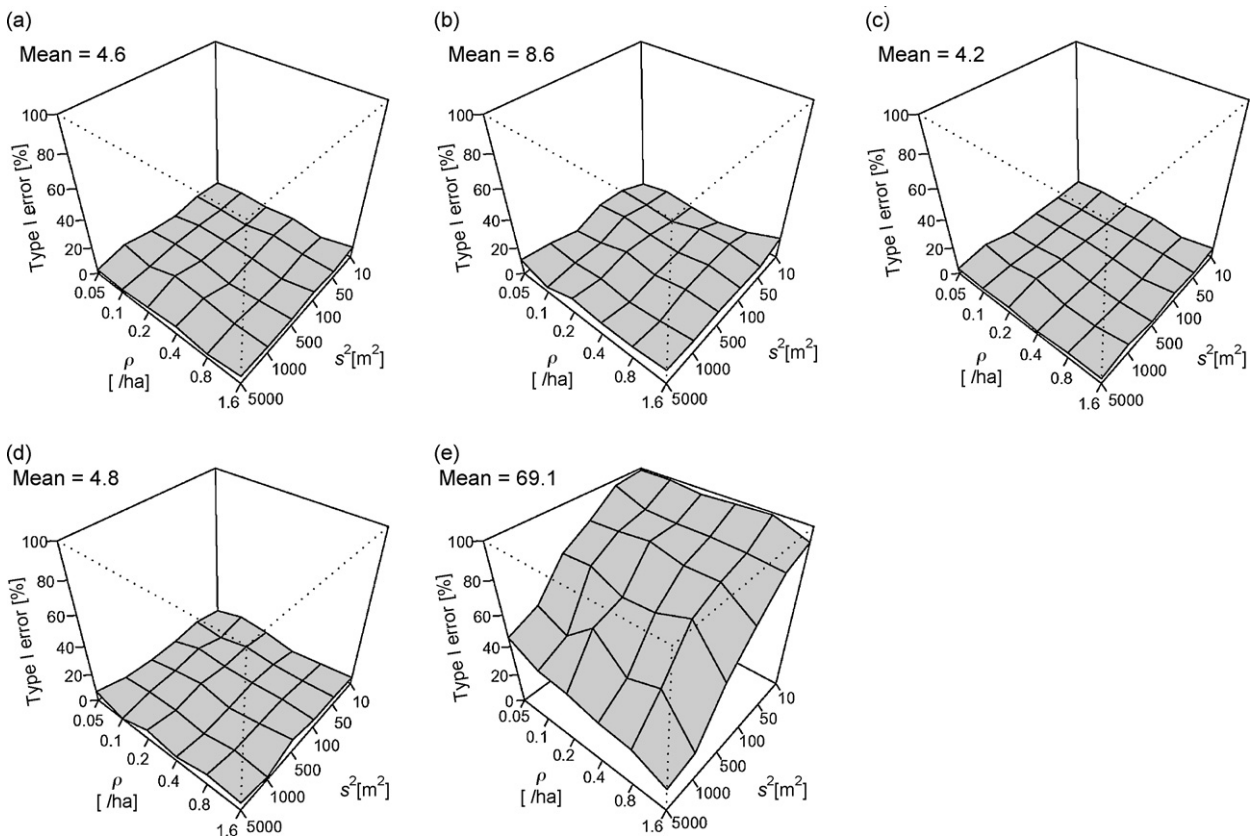


Fig. 4. Proportion of Type I errors in the results of five habitat association tests with 100 artificial generalist populations of various degrees of clumping determined by values of ρ and s^2 in a Poisson cluster model (see text for details of the model). (a) Adjusted-SD test; (b) Tree-density test; (c) SD test; (d) Chi-squared test; and (e) conventional Chi-squared test.

We set the parameters for the function as the upper limit of distance (h_0 in the “pcp” function) = 250 m, the number of distance intervals ($n.int$) = 100, and the tuning constant $expo = 0.25$ (Plotkin et al., 2000).

Fig. 6 shows the relationship between the ρ values and SDs in the Adjusted-SD test for the 55 study species (Table 1). The range for ρ was 0.02–0.95 ha^{-1} , indicating that the numbers of clumps of each species were approximately 1–50 within the 52-ha plot. As

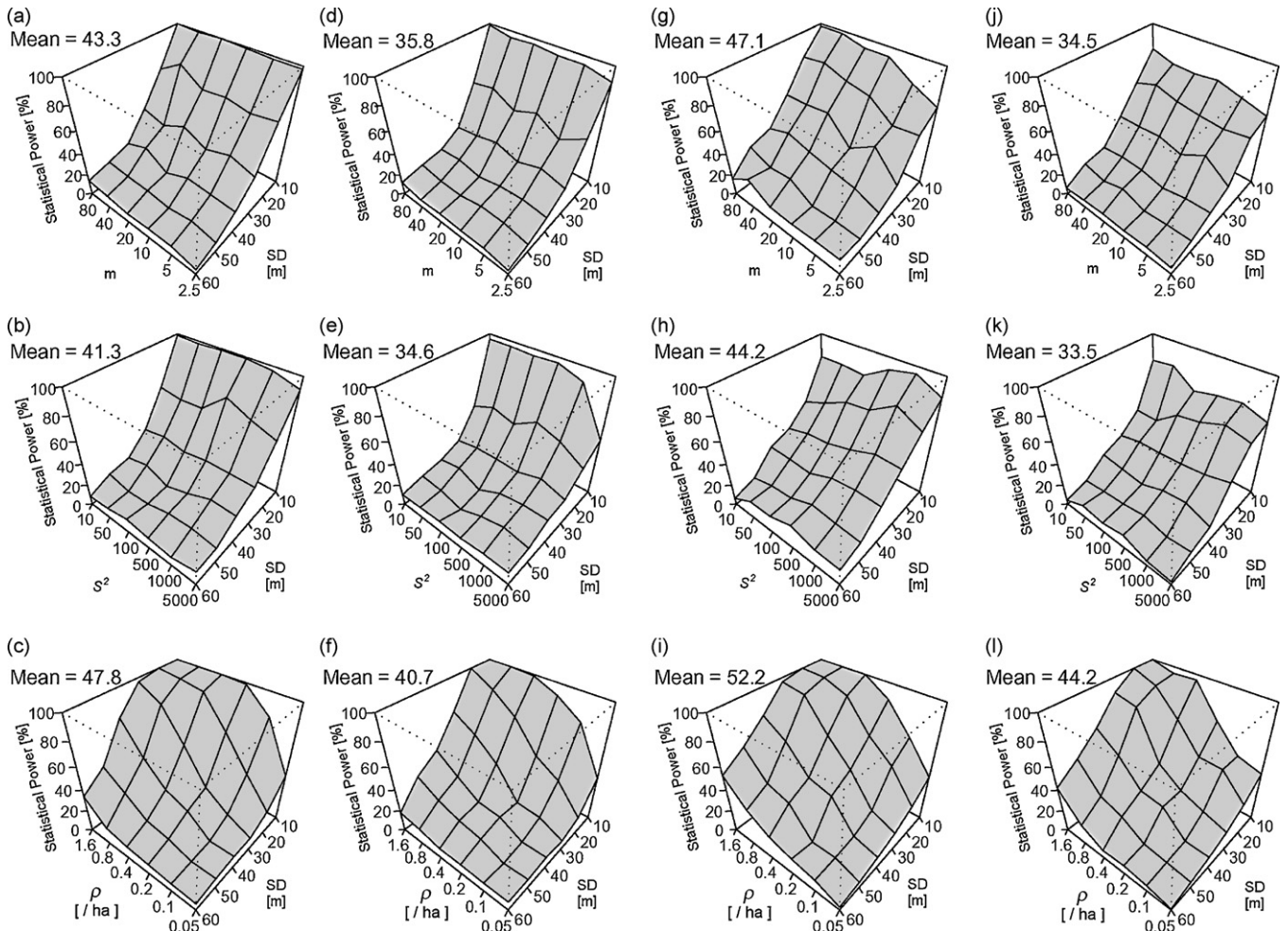


Fig. 5. Statistical power (proportion of specialist species) of four habitat association tests with 100 artificial specialist populations with various degrees of specialization (SD) and clumping (m , ρ and s^2), determined by a Poisson cluster model (see text for details of the model). The first (a–c), second (d–f), third (g–i) and fourth (j–l) columns from the left are, respectively, for the results of the Adjusted-SD, SD, Tree-density, and Chi-squared tests.

expected, species with smaller SDs were generally judged as specialists by the Adjusted-SD tests. However, the marginal SD values between the generalists and specialists decreased with decreasing ρ values, probably due to low statistical power for

species with a small ρ (Fig. 6e). Based on the power analysis and the results for the 55 Dipterocarpaceae species, the Adjusted-SD test can judge species with $SD < 20$ m as specialists when their ρ values are larger than 0.1 ha^{-1} (>5 clumps in a study plot). More specialized distributions (smaller SDs) are needed for species with fewer clumps to qualify as specialists using this test.

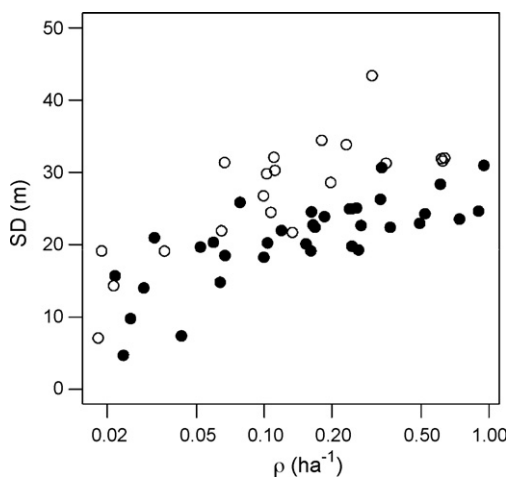


Fig. 6. Relationship between mean clump density (ρ) and degree of habitat specialization (SD) in terms of elevation for 55 dipterocarp species in a 52-ha Lambir plot. Closed and open circles are specialist and generalist species, respectively, based on the Adjusted-SD test. See text for details of ρ and SD.

4. Conclusions

This study demonstrated that all of the habitat association tests considered were robust with respect to spatial autocorrelation of the tree distributions. However, their results were not completely consistent among one another. We should use care regarding the definition of habitat when using discrete habitats, because different classifications provide different results. If possible, it is preferable to adopt methods using continuous habitat variables, especially if there is no clear rationale for ideal habitat classification. In contrast, we may obtain a straightforward interpretation with methods using discrete habitats when a plot includes qualitatively different habitats. It is also noted that the Tree-density test may overestimate, though probably not by much, the number of specialist species, due to its multiple significance testing. On the other hand, the SD and Chi-squared tests may underestimate habitat associations because of their lower statistical power. The Adjusted-SD test attained reasonable statistical power without using multiple tests.

Acknowledgments

We thank Ms. Lucy Chong, Drs. H.S. Lee, S. Tamura, K. Ogino, I. Yamada, P.S. Ashton, and S. Davies for their support during the long-term ecological study of the Lambir forest. The 52-ha study plot was established and maintained through a collaborative effort by the Forest Department of Sarawak, Sarawak Forestry Corporation, Harvard University, the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute, and several Japanese universities. We thank anonymous reviewers for critical reviews of the earlier versions of the manuscript. This work was funded in part by Grants-in-Aid for Scientific Research from MEXT and JSPS, Japan (NPO201, NP0901, 02041071, 15255005, 20405011) and the Global Environment Research Fund of the Ministry of the Environment, Japan (E-091).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.foreco.2009.10.022.

References

- Cannon, C.H., Leighton, M., 2004. Tree species distributions across five habitats in a Bornean rain forest. *Journal of Vegetation Science* 15, 257–266.
- Coomes, D.A., Grubb, P.J., 2003. Colonization, tolerance, competition and seed-size variation within functional groups. *Trends in Ecology and Evolution* 18, 283–291.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P., Foster, R.B., Itoh, A., LaFrankie, J.V., Lee, H.S., Losos, E., Manokaran, N., Sukumar, R., Yamakura, T., 2000. Spatial patterns in the distribution of tropical tree species. *Science* 288, 1414–1418.
- Connell, J.H., 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: den Boer, P.J., Gradwell, G.R. (Eds.), *Dynamics of Populations*. Center for Agricultural Publication and Documentation, Wageningen, pp. 298–312.
- Cowling, A., Hall, P., 1996. On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society: Series B* 58, 551–563.
- Cressie, N.A.C., 1991. *Statistics for Spatial Data*. John Wiley, London.
- Davies, S.J., Tan, S., LaFrankie, J.V., Potts, M.D., 2005. Soil-related floristic variation in the hyperdiverse dipterocarp forest in Lambir Hills, Sarawak. In: Roubik, D.W., Sakai, S., Hamid, A.A. (Eds.), *Pollination Ecology and the Rain Forest—Sarawak Studies*. Springer-Verlag, Tokyo, pp. 22–34.
- Diggle, P., 1983. *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Enoki, T., Abe, A., 2004. Saplings distribution in relation to topography and canopy openness in an evergreen broad-leaved forest. *Plant Ecology* 173, 283–291.
- Fortin, M.-J., Dale, M., 2005. *Spatial Analysis: a Guide for Ecologists*. Cambridge University Press, Cambridge.
- Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Esufali, S., Harms, K.E., Ashton, P.M.S., Burslem, D.F.R.P., Ashton, P.S., 2006. Species-habitat associations in a Sri Lankan dipterocarp forest. *Journal of Tropical Ecology* 22, 371–384.
- Harms, K.E., Condit, R., Hubbell, S.P., Foster, R.B., 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* 89, 947–959.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, New Jersey.
- Itoh, A., Ohkubo, T., Yamakura, T., 2006. Topography and species diversity of tropical rain forests. In: Masaki, T., Tanaka, H., Shibata, M. (Eds.), *Forest Ecology, with Long-Term Perspectives*. Bun-ichi Sogo Shuppan Co., Tokyo, (in Japanese), pp. 219–241.
- Itoh, A., Yamakura, T., Kanazaki, M., Ohkubo, T., Palmiotto, P.A., LaFrankie, J.V., Ashton, P.S., Lee, H.S., 2003. Importance of topography and soil in local dominance and distribution of two sympatric *Dryobalanops* species, emergent trees in a Bornean rain forest. *Ecological Research* 18, 307–320.
- Janzen, D.H., 1970. Herbivores and the number of tree species in tropical forests. *American Naturalists* 104, 501–528.
- John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarette, H., Vallejo, M., Foster, R.B., 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America* 104, 864–869.
- Khoyama, T., 1993. Size-structured tree populations in gap-dynamic forest—the forest architecture hypothesis for the stable coexistence of species. *Journal of Ecology* 81, 131–143.
- Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659–1673.
- Neyman, J.E., Scott, E.L., 1958. A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B* 20, 1–43.
- Noguchi, H., Itoh, A., Mizuno, T., Sri-ngernyuang, K., Kanzaki, M., Teejuntuk, S., Sungpalee, W., Hara, M., Ohkubo, T., Sahunalu, P., Dhanmmanonda, P., Yamakura, T., 2007. Habitat difference in sympatric Fagaceae tree species of a tropical montane forest in northern Thailand. *Journal of Tropical Ecology* 23, 549–558.
- Ohkubo, T., Tani, M., Akojima, I., Takahashi, S., Yamakura, T., Itoh, A., Kanzaki, M., Lee, H.S., Ashton, P.S., Ogino, K., 2007. Spatial patterns of landslide due to heavy rains in a mixed Dipterocarp forest, North-Western Borneo. *Tropics* 16, 59–70.
- Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., LaFrankie, J.V., Ashton, P.S., 2000. Species–area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology* 207, 81–99.
- R Development Core Team, 2009. *R: A language and environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL: <http://www.r-project.org>, ISBN: 3-600051r-r07-0.
- Silverman, B.W., 1986. *Density Estimation*. Chapman and Hall, London.
- Schuster, E.F., 1985. Incorporating support constraints into nonparametric estimation of densities. *Communications in Statistics-Theory and Methods* 14, 1123–1136.
- Webb, C.O., Peart, D.R., 2000. Habitat associations of trees and seedlings in a Bornean rain forest. *Journal of Ecology* 88, 464–478.
- Wright, S.J., Muller-Landau, H.C., Condit, R., Hubbell, S.P., 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84, 3174–3185.
- Yamakura, T., Kanzaki, M., Itoh, A., Ohkubo, T., Ogino, K., Chai, E.O.K., Lee, H.S., Ashton, P.S., 1995. Topography of a large-scale research plot established within a tropical rain forest at Lambir Sarawak. *Tropics* 5, 41–56.