ORIGINAL ARTICLE

Kenichiro Shimatani · Keiko Kitamura
Tatsuo Kanazashi · Hisashi Sugita

# Genetic inhomogeneous Poisson processes describing the roles of an isolated mature tree in forest regeneration

**Abstract** The spatial distribution of genotyped adult plants and offspring can be modeled by genetic inhomogeneous Poisson processes. This paper reports the development of a previously proposed point process model to cover safe site conditions for sapling survival, unknown seed sources and wider ranges of dispersal kernels. Suppose that a species has limited seed dispersal and shade-tolerance, and that young trees are clustered around highly isolated adults. The clustering might be formed solely by dispersed seeds from adults. However, the survival of the offspring might be influenced by shading by the adults' crowns. The new genetic processes are applicable to such cases, as demonstrated for a young population of *Fagus crenata*, a shade-tolerant canopy tree species, in a 55-year-old stand regenerated after shelterwood logging. Isozyme analysis revealed that the regenerating trees were genetically related to the nearest adults, but some were not their respective daughters. The maximum likelihood method has led us to the following regeneration: seeds were dispersed mostly within 20 m of their mothers; two residual adults in the plot had equal female reproductive success; about 10% of offspring even close to the adult were not their daughters; the adults' crowns reduced the survival of offspring up to 20–25 m away. However, the degree of model fitting was unsatisfactory. Hence, our models, in principle, can quantify two roles of highly isolated adults: providing seed sources and safe sites, but the case study suggests that other, unknown factors influence *F. crenata* regeneration after shelterwood harvesting.

**Keywords** Clustering · Inhomogeneous Poisson process · Isozyme · Reproductive success · Safe site · Seed dispersal

K. Shimatani (✉)
The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato, Tokyo 103-8569, Japan
E-mail: shimatan@ism.ac.jp
Tel.: +81-3-54218774
Fax: +81-3-54218796

K. Kitamura
Hokkaido Research Center, Forestry and Forest Products Research Institute, Sapporo, Japan
E-mail: kitamq@ffpri.affrc.go.jp

T. Kanazashi
Forestry and Forest Products Research Institute, Tsukuba, Japan
E-mail: kana@ffpri.affrc.go.jp

H. Sugita
Tohoku Research Center, Forestry and Forest Products Research Institute, Morioka, Japan
E-mail: sugitah@ffpri.affrc.go.jp

## Introduction

In species with limited seed dispersal, such as members of the Fagaceae, seedlings tend to emerge near their mothers. Thus, when mature trees are isolated and seedlings are clustered around the adults, the seedling population seems to consist of maternal half-sib patches. However, recent studies using genetic markers have revealed, interestingly, that the nearest adult to an offspring is not necessarily its mother, and long-distance immigrants born elsewhere are found even beneath mother trees (Dow and Ashley 1996; Kitamura et al. 1997; Konuma et al. 2000; Godoy and Jordano, 2001; González-Martínez et al. 2002). In addition, genetic point process models have indicated that clusters of seedlings with compatible genotypes to nearby adults are not formed simply by limited dispersal of seeds around the adults (Shimatani 2004), and that the secondary forest under highly isolated adults is not regenerated solely from the isolated adults but that progenies of other adults are also likely to be present (Shimatani 2002).

Spatial sapling distribution appears to be related not only to the provision of seeds by an adult but also a

"safe site", i.e. appropriate micro-environmental conditions for germination and growth (Harper 1977; Houle 1998; Calviño-Cancela 2002). For example, suppose that a species is shade-tolerant, mature trees of that species are highly isolated and no adults of other species are present (such highly stratified stands can be seen in shelterwood systems, in which some adults of specific species are preserved and all other stems are harvested). If the abundance of light far from any adult promotes the growth of other, shade-intolerant species (Loftis 1990; Brose and Van Lear 1998), but shading by adults suppresses such species' regeneration, isolated adults may provide relative competitive advantages (safe sites) for the shade-tolerant seedlings, resulting in higher frequencies of survivors near the adults. If the shading also suppresses the survival of the shade-tolerant seedlings, sapling density will be high under the edges of the adult crowns. In such cases, isolated adults have two roles in regeneration; providing both seed sources and safe sites for offspring, and the combination of the two mechanisms results in the clustering of saplings around adults.

The objective of the study presented here was to improve previously proposed inhomogeneous point processes capturing genetic information (Shimatani 2004), and to demonstrate their applicability. We first describe a case study site and the characteristics of its secondary forest (including genetic clustering) to illustrate the kinds of populations our models are intended to cover. We then review previous genetic inhomogeneous Poisson processes and describe the improvements we have made. The improved genetic processes reflect (1) the spatial distribution of dispersed seeds, (2) the variation of female reproductive success, (3) the contribution of unknown seed sources such as distant sources and advanced regeneration, and (4) shading effects of adults on seedling survival. Applying the models to the regeneration mechanisms in a young *Fagus crenata* population under highly isolated adults, we obtained a quantitative description of the regeneration mechanism at the study site. Unfortunately, even the optimized models did not satisfactorily fit the observations, and we briefly discuss the efficiency and limitations of our modeling approach.

## Materials and methods

### Study site

The study site was located in the Kurosawajiri Experimental Forest for Natural Forest Management in northern Honshu, Japan (39°14′N, 140°62′E). The 104.2 ha experimental forest was established in 1944 to develop natural forest management techniques for beech (*Fagus crenata*)-dominated forests. For our study we selected a shelterwood compartment of about 3 ha and established a permanent 60×40 m monitoring plot in a flat area within it. The harvesting was carried out during the winter of 1948–1949, and six mature *F. crenata* trees were preserved per hectare. Brush was cut in 1951 after the mast year of 1950, primarily to remove shrubs. No residual adult has been harvested. Yanagiya et al. (1969) reported that excellent regeneration of *F. crenata* occurred after harvesting, and that there were 16,990 seedlings per hectare on average, most of which were 18 or 19 years old in 1968, i.e., they germinated before the brush cutting in the second to third spring after harvesting [here and in the cited study by Yanagiya et al. (1969), a current-year seedling is referred to as being 1 year old, hence a plant that was a current-year seedling in the spring of 1951 was 18 years old in 1968).

During 2002–2004, we mapped all living stems taller than breast height that were found to have a minimum diameter at breast height, DBH, of 2.5 cm, as well as those of the two mature *F. crenata* trees (named L and R) found in the plot, which had crowns with 5–10 m radii from their trunks. The stand was dominated by *F. crenata* (accounting for 73% of the dominant regenerating trees, and 80% of the basal area; Sugita et al. (2006)), suggesting that *F. crenata* had successfully regenerated. Young *F. crenata* trees were clustered around the two large individuals, but more densely near the edges of the crowns (8–16 m from the trunks) than beneath the crowns (Figs. 1, 2). The nearest large *F. crenata* individual from the plot was 16.5 m below the bottom edge in Fig. 1.

### Fagus crenata

*Fagus crenata* Blume, which belongs to the beech family (Fagaceae), is a deciduous, wind-pollinated, self-incompatible canopy tree species that produces large acorns dispersed mostly by gravity, although some are spread by rodents and jays. It is widely distributed in cool temperate forests in Japan, ranging from southern Kyushu to southern Hokkaido. In some areas, forests are strongly dominated by *F. crenata*, and in mast years the forest floor in such stands is carpeted by current-year seedlings of the species. For further details regarding the ecology of *F. crenata*, see Kitamura et al. (1997) and references therein. Beech-dominated forests have high conservation values, and thus should be sustainably managed. Management regimes involving the use of shelterwoods may be among the most suitable, provided that *F. crenata* regenerates immediately after harvesting.

### Isozyme analysis

In June 2004, we collected leaves from all 255 *F. crenata* trees (including the two adults) in the plot. Samples were separately collected from multiple stems with the same base in order to check whether they originated from stump sprouts or from different seeds. The samples were

**Fig. 1a–c** Spatial allelic distributions of *Fagus crenata* trees. The grid spacing is 10 m. *Single marking* denotes a homozygote of the respective allele while *double marking* indicates a heterozygote of the two alleles. **a** *Amy2* locus (*open square a, Δ b, open circle c, − d, × e, + f*), **b** *Aap1* locus (*open circle a, × b, − c*) and **c** *Fum* locus (*open circle a, × b, − c*). The designations and diameter at breast height (DBH) of the two adults (*large symbols*) are shown in **a**
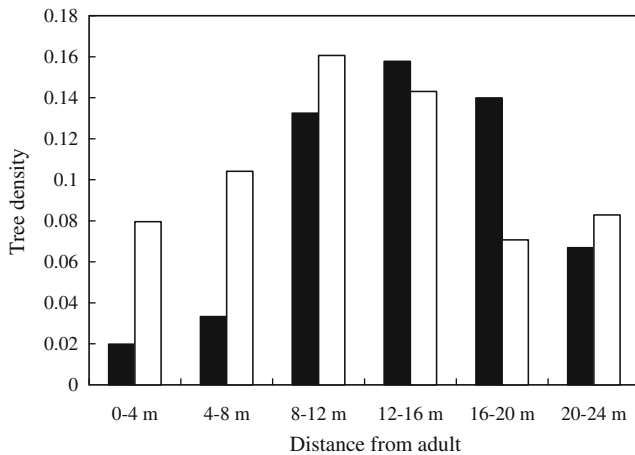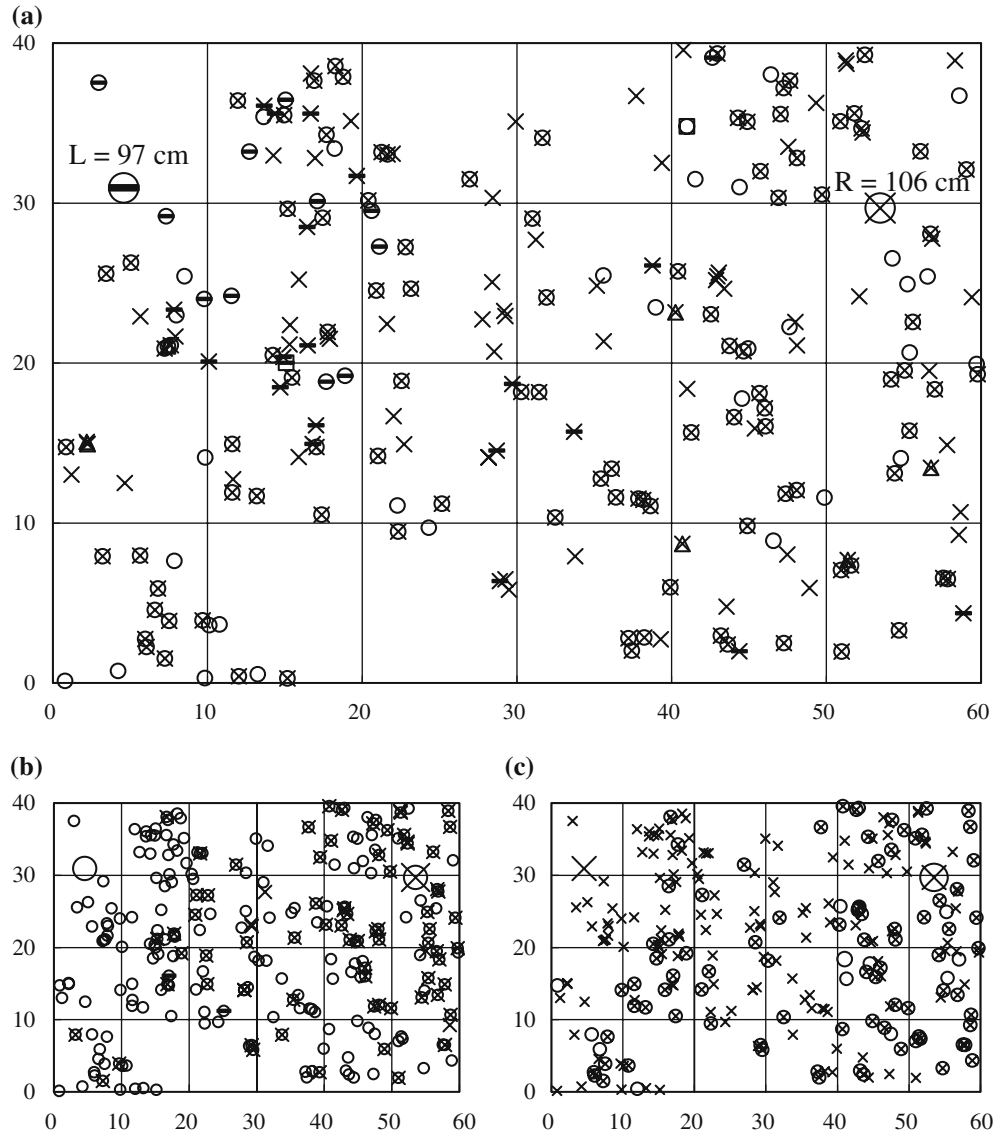




**Fig. 2** The densities (per m$^2$) of regenerating *Fagus crenata* trees along distance classes from adults *L* (*black bars*) and *R* (*white bars*)

kept at approximately 4°C before they were moved to the laboratory, then stored in a deep freezer at −80°C until enzyme extraction. Isozymes of ten enzymes were analyzed by vertical slab polyacrylamide gel electrophoresis, and we detected 11 polymorphic genetic loci. For details concerning the procedures used in the isozyme analysis, see Kitamura et al. (1997) and references therein. Alleles were designated using the same notation as Kitamura et al. (1997). When a new allele was obtained, it was named c' if it appeared between alleles c and d, and z if before a.

Spatial genetic distribution

Table 1 shows the allele frequencies of the young *F. crenata* population and the two mature trees, and

**Table 1** Observed allele frequencies of *Fagus crenata* saplings and two adults for 11 loci, and estimated frequencies in the haploid pollen pool received by the adults for the optimized genetic process (2Dt-Weibull dispersal kernels, $\beta = 1,000$ fixed) assuming equal female reproductive success and equal pollen pools

| locus | allele | Frequency (%) | | |
|---|---|---|---|---|
| | | Observed saplings | Two adults | Model |
| *Aap1* | z | 0.00 | 0.00 | 0.00 |
| | a | 81.33 | 75.00 | 85.81 |
| | b | 18.26 | 25.00 | 13.67 |
| | c | 0.41 | 0.00 | 0.52 |
| *Amy2* | a | 0.41 | 0.00 | 0.52 |
| | b | 1.45 | 0.00 | 1.83 |
| | c | 37.55 | 50.00 | 34.70 |
| | d | 6.85 | 25.00 | 3.46 |
| | e | 53.53 | 25.00 | 59.23 |
| | f | 0.21 | 0.00 | 0.26 |
| *Fum* | a | 23.86 | 25.00 | 22.19 |
| | b | 76.14 | 75.00 | 77.81 |
| *Mdh* | z | 0.21 | 0.00 | 0.26 |
| | a | 87.55 | 100.00 | 84.31 |
| | b | 12.24 | 0.00 | 15.43 |
| *6Pg* | a | 0.83 | 0.00 | 1.05 |
| | b | 98.76 | 100.00 | 98.43 |
| | c | 0.41 | 0.00 | 0.52 |
| *Dia* | b | 3.53 | 0.00 | 4.45 |
| | c | 96.06 | 100.00 | 95.03 |
| | d | 0.41 | 0.00 | 0.52 |
| *Got1* | a | 0.21 | 0.00 | 0.26 |
| | c | 98.96 | 100.00 | 98.69 |
| | d | 0.83 | 0.00 | 1.05 |
| *Got3* | b | 0.83 | 0.00 | 1.05 |
| | c | 97.72 | 100.00 | 97.12 |
| | c′ | 0.21 | 0.00 | 0.26 |
| | d | 1.24 | 0.00 | 1.57 |
| *Pgi* | c | 0.41 | 0.00 | 0.52 |
| | d | 99.17 | 100.00 | 98.95 |
| | e | 0.41 | 0.00 | 0.52 |
| *Pgm1* | z | 1.66 | 0.00 | 2.09 |
| | a | 97.51 | 100.00 | 96.86 |
| | b | 0.62 | 0.00 | 0.78 |
| | c | 0.21 | 0.00 | 0.26 |
| *Pgm2* | a | 6.43 | 0.00 | 8.11 |
| | b | 3.11 | 0.00 | 3.92 |
| | c | 89.21 | 100.00 | 86.40 |
| | d | 0.62 | 0.00 | 0.78 |
| | e | 0.62% | 0.00% | 0.78% |

Fig. 1 illustrates the spatial allelic distribution for three loci that showed polymorphisms in the two adults. Among 17 multiple-stem patches (each had two or three stems), 7 consisted of stems with different genotypes. In a particularly noteworthy case, three stems are combined, giving the appearance of a single tree with three branches, although the central stem has a different genotype from the other two. The two outer stems probably originated from stump sprouts, while the other originated from a seed that germinated on the stump. In the following analysis, multiple stems with a common genotype were identified as single individuals (these multiple stems are pooled in Table 1 and Figs. 1–2). Consequently, there were 241 individuals, 11 (4.6%) of which had genotypes that were incompatible with them being the offspring of the two adults, suggesting that they originated from long-distance seed movement or advanced regeneration prior to harvesting.

Allelic clustering was observed for allele *d* at the *Amy2* locus around the *cd*-heterozygous adult L (Fig. 1a), for *b* at *Aap1* around the *ab*-heterozygous adult R (Fig. 1b), and for *a* at *Fum* around the *ab*-heterozygous adult R (Fig. 1c), suggesting that short-distance seed dispersal had occurred. However, many individuals with a homozygous *ee* genotype at *Amy2* were found near the *cd*-heterozygote L (Fig. 1a). Some *aa*-homozygotes at *Fum* were found very far from the *ab*-heterozygote R (Fig. 1c), and some individuals were *bb*-homozygous at *Mdh*, although both adults are *aa*-homozygotes (Table 1). Therefore, some of the regenerating trees near the adults appear to be daughters scattered from the respective adults, but others likely originated from elsewhere.

Genetic similarity

The genetic similarities between adults and young trees were examined by calculating the average numbers of alleles in common (NAC, Surles et al. 1990) when the two were within *r* meters (cf. Berg and Hamrick 1995):

$$nac_j(r) = \{\text{average NAC}(i, j) \text{ over } i \text{ satisfying}$$
$$\|Z_j - X_i\| \leqslant r\},$$

where $Z_j$ is the location of adult *j*, and $X_i$ that of offspring *i*. NAC $= 2$ if the two individuals have an identical genotype, NAC $= 1$ if, for example, one is AA and the other is AB, and NAC $= 0$ between AA and BC. NAC(*i, j*) indicates the average NACs over the 11 loci. Thus, when NAC(*i, j*) is high, *i* and *j* have similar genotypes, and are likely to have a kinship such as a mother-daughter relationship.

Fig. 3a illustrates $nac_L(r)$ and $nac_R(r)$ ($r \leq 20$ m) together with 95% confidence envelopes if the spatial genotypic distribution is random; genotypes at the given locations were randomly permutated 100 times, $nac(r)$ was calculated and their 5 and 95 percentiles were connected. Thus, if a part of the curve is beyond the envelopes, such genetic similarity occurs less than 5 times per 100, if the distribution is random.

Young trees near adult L had significantly higher NAC values while young trees further away had progressively lower NAC values, suggesting that the majority of nearby young trees were offspring of adult L and the proportion of L's daughters decreased with distance from L. Such a tendency is less clear around R.

Therefore, the regenerating population was formed, at least in part, by short-distance seed dispersal from the two residual adults, but some must have resulted from long-distance travel (or advanced regeneration of harvested stems and/or sprouting from stumps). In order to assess the relative contributions of these mechanisms, as

**Fig. 3 a** The observed numbers of alleles in common [*nac(r)*] (*bold lines*) between adults (*open squares* L, × R) and saplings. The dotted lines show 95 confidence envelopes drawn from 100 simulations, using the same symbols for the adults. **b** The expected *nac(r)* from adult-L for the optimized 2Dt × Weibull model ($\beta = 1,000$, thin line with *open squares*), exponential power × Weibull (thin line with *open diamonds*), and model (5) based solely on seed dispersal of the log-normal distribution (*dotted line* with *triangles*). The observed *nac_L(r)* is indicated by a *bold line* with *open squares* (same curve as in **a**). **c** The expected *nac(r)* (*thin lines*) between adults (*open squares* L, × R) and saplings according to the optimized 2Dt × Weibull model ($\beta = 1,000$). The observed *nac(r)* is indicated with *bold lines* (same curves as in **a**), and the *dotted lines* are 95 confidence envelopes drawn from 1,000 simulations
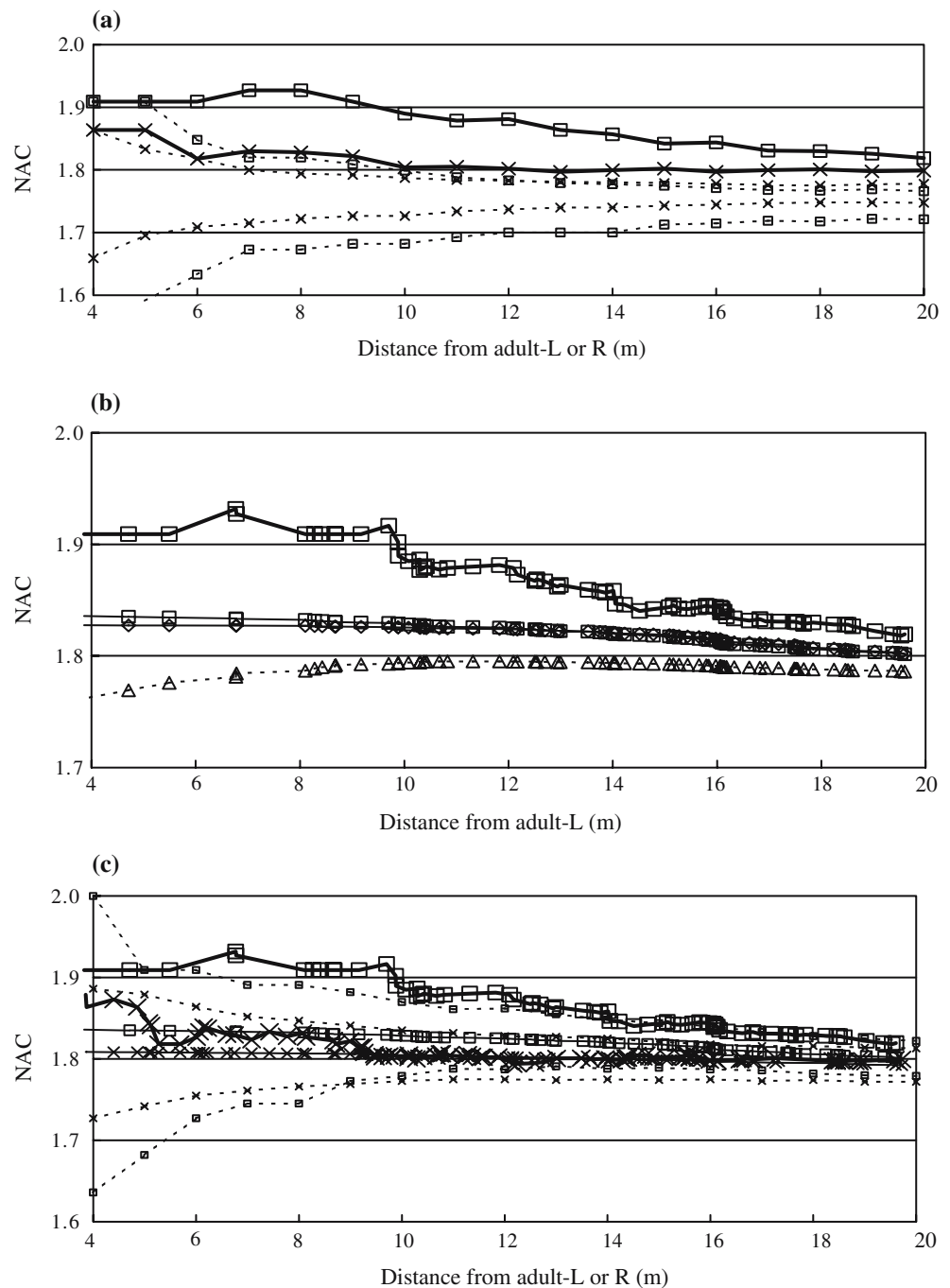


well as the ranges of short-distance seed dispersal and the effects of shading by the adults, we constructed genetic point process models as described below and applied them to this population.

## Models

### Inhomogeneous Poisson processes

Suppose that there are $M$ adult trees at $\{Z_1, Z_2, ..., Z_M\}$, and that each produces $u_j$ offspring according to the Poisson distribution of intensity $U_j$, and disperses them according to a non-negative function $f(r; \Theta)$ satisfying $\int_{+0}^{+\infty} f(r)2\pi r dr = 1$ [the dispersal kernel; Stoyan and Wagner (2001) called $f(r)$ the probability density] in which $\Theta$ is a parameter set in $f(r)$. Point patterns generated by this stochastic process can be formulated as an inhomogeneous Poisson process with first-order intensity (Rathbun and Cressie 1994; Stoyan and Penttinen 2000; Shimatani 2004):

$$\lambda_0(x; \mathbf{U}, \Theta) = \sum_{j=1}^{M} U_j \cdot f(\|x - Z_j\|; \Theta), \qquad (1)$$

where $\|\cdot\|$ indicates the Euclidean distance and $\mathbf{U} = (U_1, U_2, ..., U_M)$. $\lambda_0(x; \mathbf{U}, \Theta)\Delta x$ indicates the probability that an offspring is present in a small area $\Delta x$ centered at $x$.

As a dispersal kernel function, we tested two decreasing functions; the exponential power function $f(r; \alpha, \beta) = \dfrac{\beta}{2\pi\alpha^2\Gamma(2/\beta)} e^{-(r/\alpha)^\beta}$ and the two-dimensional $t$-distribution (2Dt) $f(r; \alpha, \beta) = \dfrac{\beta - 1}{\pi\alpha^2}(1 + r^2/\alpha^2)^{-\beta}(\alpha > 0$ and $\beta > 1)$ (Clark et al. 1999; Austerlitz et al. 2004). The average distances of these seed dispersals are given by $\displaystyle\int_0^\infty rf(r)2\pi r dr = \alpha\left(\dfrac{\Gamma(1.5)\Gamma(\beta - 1.5)}{\Gamma(\beta - 1)}\right)$ and $= \alpha\dfrac{\Gamma(3/\beta)}{\Gamma(2/\beta)}$, respectively (Austerlitz et al. 2004).

Practically, we can map adults only in a limited area. Suppose that trees at $\{Z_1, Z_2, ..., Z_L\}$ in a rectangular plot $A$ $(=[0, a] \times [0, b])$ are mapped and all the other potential mothers $\{Z_{L+1}, Z_{L+2}, ..., Z_M\}$ are not. There is no way to calculate $f(\|x - Z_j\|)$ for $j = L + 1, L + 2, ..., M$. If we set $m = \displaystyle\sum_{j=L+1}^M U_j f(\|x - Z_j\|)$ and consider this amount as one unknown parameter that covers all unmapped potential mothers such as those outside the plot (long-range dispersal) and those that are already dead (e.g. advanced regeneration, stump sprouts), (1) is written:

$$\lambda(x; \mathbf{U}, \Theta, m) = \sum_{j=1}^M U_j \cdot f(\|x - Z_j\|; \Theta) + m. \quad (2)$$

Then, $m\Delta x$ indicates the probability that a seedling born from unmapped adults appears in a small area $\Delta x$ centered at $x$. When an offspring emerges at $X$, it is a daughter of adult 1, 2, ..., $M$ and an unmapped adult with the probability proportional to $U_1 f(\|X - Z_1\|)$, $U_2 f(\|X - Z_2\|)$, ..., $U_M f(\|X - Z_M\|)$ and $m$.

*Likelihood*

Suppose that $N$ offspring are observed at $\{X_1, X_2, ..., X_N\}$ in a subset $B$ in adult plot $A$ (an offspring plot is often smaller than an adult plot because of its greater density. In the case of our study site, $B = [0, 60] \times [0, 40] \subset A = [-15, 80] \times [-15, 70]$). The likelihood of the above inhomogeneous Poisson process is given as (Cressie 1991: pp 650–656):

$$\prod_{i=1}^N \lambda(X_i; \mathbf{U}, \Theta, m) \cdot e^{-\int_B \lambda(x; \mathbf{U}, \Theta, m)dx},$$

if $N > 0$. (For $N = 0$, it is $e^{-\int_B \lambda(x; \mathbf{U}, \Theta, m)dx}$). The maximum likelihood method can be applied to its logarithm to optimize the parameters:

$$\sum_{i=1}^N \ln(\lambda(X_i; \mathbf{U}, \Theta, m)) - \int_B \lambda(x; \mathbf{U}, \Theta, m)dx. \quad (3)$$

*Genetic processes*

Suppose that all adults and offspring are genotyped, and $p_l(j, c) = \{$the frequency of allele $c$ at genetic locus $l$ in the set of haploid pollen received by adult $j\}$ are given. Let:

$P(j, G_l) = \{$the probability that offspring between mother $j$ and arbitarily chosen haploid pollen has genotype $G_l$ under the Mendelian law of inheritance$\}$.

For a long-distance traveler, suppose that its genotype is governed by random mating over the population:

$P(G_l) = \{$the probability that genotype $G_l$ arises by random mating$\}$.

We extended (2) to its genetic version for every genotype $G_l$ as:

$$\lambda_{G_l}(x; \mathbf{U}, \Theta, m, \{p_l(j, c)\})$$
$$= \sum_{j=1}^M P(j, G_l)U_j \cdot f(\|x - Z_j\|; \Theta) + mP(G_l). \quad (4)$$

$\lambda_{G_l}(x)\Delta x$ expresses the probability that an offspring with genotype $G_l$ appears in a small area centered at $x$ (Shimatani 2004). This can be extended to multilocus cases as

$$\lambda_G(x) = \sum_{j=1}^M P(j, G)U_j \cdot f(\|x - Z_j\|; \Theta) + mP(G), \quad (5)$$

where $G = \Pi_l G_l$, $P(j, G) = \Pi_l P(j, G_l)$, $P(G) = \Pi_l P(G_l)$. Note that since $\sum_G P(j, G) = \sum_G P(G) = 1$, the sum of $\lambda_G(x)$ over all possible genotypes is equal to the first-order intensity of the spatial individual distribution:

$$\lambda(x) = \sum_G \lambda_G(x). \quad (6)$$

If the reproduction independently occurs for each genotype, the likelihood of given genotypic data $\bigcup_G \{X_1^G, X_2^G, ..., X_{N_G}^G\}$ ($N_G$ refers to the number of offspring with genotype $G$ and $X_i^G$ to the location of the $i$-th offspring of genotype $G$) is the product over the multilocus genotypes:

$$\prod_G \left\{\prod_{i=1}^{N_G} \lambda_G(X_i^G) \cdot e^{-\int_B \lambda_G(x)dx}\right\},$$

and because of (6), the log-likelihood is written as:

$$\sum_G \sum_{i=1}^{N_G} \ln(\lambda_G(X_i^G)) - \int_B \lambda(x)dx,$$

or

$$\sum_{i=1}^{N} \ln(\lambda_{G_i}(X_i)) - \int_B \lambda(x)dx, \tag{7}$$

where $G_i$ is the genotype of offspring $i$. In this study, $P(G_l)$ was calculated from the average over estimated allele frequencies in haploid pollen pools [$p_l(L, c)$ and $p_l(R, c)$] under Hardy-Weinberg equilibrium.

*Survival model*

As mentioned above, there were fewer *F. crenata* trees very near the trunks of the isolated adults than further away, and their density peaked at distances in the range 8–16 m. Suppose that, in general, original seeds are dispersed around their mothers according to a decreasing function $f(r)$, but habitat conditions vary over the stand, and survival rates are given by a function $s(x)$ on $B$. Then, the distribution of survived saplings with genotype $G$ is expressed by another inhomogeneous Poisson process with first-order intensity

$$\lambda_G(x) = s(x) \cdot \left\{ \sum_{j=1}^{M} P(j,G)U_j \cdot f(\|x - Z_j\|) + mP(G) \right\}. \tag{8}$$

Under highly isolated adults in a flat stand like the case study site, shading from adults' crowns might largely control survival chances of seedlings/saplings. If isolated adults negatively affect the survival of all species beneath their crowns, and if fast-growing species have relative advantages in open spaces, the survival rates of shade-tolerant saplings might be high at a certain distance from an adult. Hence, using another dispersal kernel $F(r, \Psi)$, we assumed that $s(x)$ is proportional to $F(\|x - Z_{j(x)}\|, \Psi)$ where $j(x)$ refers to the nearest adult from $x$;

$$\lambda_G(x) = C_0 F(\|x - Z_{j(x)}\|; \Psi)$$
$$\left\{ \sum_{j=1}^{M} P(j,G)U_j \cdot f(\|x - Z_j\|; \Theta) + mP(G) \right\}. \tag{9}$$

In attempts to describe this dispersal kernel, we tested the lognormal distribution $F(r;a,b) = e^{-(\ln r - a)^2/2b^2}/r^2(2\pi)^{3/2}b$ (Stoyan and Wagner 2001) and the Weibull distribution $F(r;a,b) = b/2\pi\,a^2 \cdot (r/a)^{b-2} e^{-(r/a)^b}$ (Clark et al. 1999; Austerlitz et al. 2004). Both of these functions can express skewed bell-shaped curves with a maximum near the origin. Note that the proportional constants $C_0$ cannot be distinguished from the productivity $U_j$ and $m$, hence $C_0$ was fixed to reduce the redundant constant [$C_0 = (2\pi)^{3/2}$ for the log-normal and $C_0 = 2\pi$ for the Weibull distributions]. Ecologically, this implies that high seed production with low survival rates give the same results as low production and high survival rates.

*Necessity condition for maxima*

If $(U_j, m, \Theta, \Psi)$ is at the (local) maximum, all the partial derivatives of the log-likelihood [Eq. 7 using $\lambda_G(x)$ given by Eq. 9] with respect to these parameters are zero, hence,

$$\begin{cases} \dfrac{\partial L}{\partial U_j} = \sum_i \dfrac{s(x)P(j,G_i)f(\|X_i - Z_j\|)}{\lambda_{G_i}(X_i)} \\ \qquad - \int_B s(x)P(j,G_i)\,f(\|X_i - Z_j\|)dx = 0 \\ \dfrac{\partial L}{\partial m} = \sum_i \dfrac{s(x)P(G_i)}{\lambda_{G_i}(X_i)} - \int_B s(x)P(G_i)dx = 0 \end{cases}$$

Since $U_j \cdot \dfrac{\partial L}{\partial U_j} + m \cdot \dfrac{\partial L}{\partial m} = 0$ yields $\sum_i \dfrac{\lambda_{G_i}(X_i)}{\lambda_{G_i}(X_i)} -$
$\int_B \lambda(x)dx = 0$, we have

$$\int_B \lambda(x)dx - N = 0. \tag{10}$$

This necessity condition can be used to check if an optimization process has worked appropriately.

*Maximum likelihood method*

We first assumed that the two adults had equal reproductive success and received equal haploid pollen pools [$U_L = U_R$ and $p_l(L, c) = p_l(R, c)$] and applied the maximum likelihood (quasi-Newtonian) method by changing six parameters $\{U, \alpha, \beta, a, b, m\}$ while keeping $p_l(\bullet, c)$ fixed at the initial values (the observed frequencies over the 243 trees), then $p_l(\bullet, c)$ (28 parameters in total) were optimized. This process was repeated until further changes did not increase the likelihood, in a couple of trials. In practice, $p_l(\bullet, c)$ influenced the likelihood much more weakly than the above six parameters, especially for the nine non-polymorphic loci in the two adults.

Next, we tested unequal reproductive success by separately optimizing $U_L$ and $U_R$, followed by the unequal pollen pool ($7 + 56$ parameters in total). The results were evaluated by calculating the Akaike Information Criterion (AIC) defined as:

AIC = $-2 \times$ {the maximum log $-$ likelihood}
$+ 2 \times$ {the number of parameter}.

If the AIC of a model with more parameters is higher, the added parameters are considered redundant and the simpler model is judged to be better. In general, AIC improvement greater than 2.0 are desirable for selecting a model with additional parameters (Sakamoto 1991); hence, greater than 2.0 and 29.0 are required to include the unequal reproductive success and unequal pollen pools, respectively.

*Comparisons with simple models*

We may omit the survival rate function and directly use the skewed bell-shaped dispersal kernel functions as seed

dispersal functions; i.e., model (5) using the log-normal or Weibull distribution for the dispersal kernel. Then, we can test if the clusters could be formed solely by the mechanism of limited seed dispersal. This type of models is equal to the previous model (Shimatani 2004), but more dispersal (or seed production) near the crown edge (Stoyan and Wagner 2001).

Finally, excluding genetic information, we may simply apply

$$\lambda(x) = C_0 F(||x - Z_{j(x)}||) \left\{ \sum_{j=1}^{M} U_j \cdot f(||x - Z_j||) + m \right\} \tag{11}$$

to the spatial distribution data. We also tested this model to examine the effects and importance of incorporating genetic information.

### Model fitting tests

The degree of model fitting was tested for both the spatial distribution of the young trees and their genotypic distribution. For an inhomogeneous Poisson process, the number of points in domain $D$ follows the Poisson distribution of intensity $\int_D \lambda(x)dx$. Hence, we compared the observed numbers of trees in $[x_1, x_1 + 10] \times [y_1, y_1 + 10]$ for $x_1 = 0, 10, ..., 50$ and $y_1 = 0, 10, ..., 30$, with the confidence envelopes drawn by the Poisson distribution of intensity ($=$ expected value $=$ variance) $\int_{x_1}^{x_1+10} \int_{y_1}^{y_1+10} \lambda(x)dxdy$.

For the genotypic distribution, we simulated genotypes at the observed locations ($X_i$) by first assigning their mothers to adult L, adult R or unknown seed sources according to the probability proportional to $[U_L f(r_{L,i}), U_R f(r_{R,i}), m; r_{L,i}$ and $r_{R,i}$ are the distance from $X_i$ to L and R, respectively] then determining their genotypes according to Mendelian law. Next, we calculated $nac(r)$ between the adults and the offspring with simulated genotypes, repeated the process 1,000 times and drew 95% confidence envelopes by connecting their 5 and 95 percentiles. Moreover, since the probability that the genotype at $X_i$ is $G_i$ can be similarly calculated, we derived the expected NAC between the adults and offspring at $X_i$, and calculated the expected $nac_L(r)$ and $nac_R(r)$ for $r \leq 20$ m.

## Results

The first four rows in Table 2 show the values of the optimized parameters and the maximum log-likelihood for the four combinations of the two dispersal kernels under equal female reproductive success and equal pollen pool. Note that when the 2Dt distribution was used in the seed dispersal kernel, greater $\beta$ value provided greater likelihood and the maximum likelihood method did not reach a maximum, although the resulting dispersal curves were almost the same regardless of $\beta$ values (see Fig. 4 below). Hence, two examples ($\beta = 3$ and 1,000) are shown in Table 2. For the other kernels, the optimization continued until the necessity condition (10)

**Table 2** Parameter values for four types of genetic inhomogeneous Poisson processes and their maximum log-likelihood. The average distances of seed dispersals derived from each parameter set are also shown. The 1,000 and 3 are italicized to show that these values were fixed in the maximum likelihood process

| Dispersal kernel | | Estimated parameters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Seed dispersal | | Average distance | Crown effects | | Productivity | | Random factor | Maximum log-likelihood |
| Seed dispersal | Crown effects | $\alpha$ | $\beta$ | | a | b | UR | UL | | |
| Equal female reproductive success and equal pollen pool | | | | | | | | | | |
| 2Dt | Weibull | 331.0 | *1000* | 9.29 | 32.15 | 3.66 | 671.5 | = UR | 0.203 | −2,125.4 |
| Exp[a] | Weibull | 15.02 | 4.03 | 10.37 | 35.58 | 3.29 | 438.6 | = UR | 0.187 | −2,124.3 |
| 2Dt | lognormal | 362.7 | *1000* | 10.17 | 7.20 | 1.26 | 30,085,048 | = UR | 8123 | −2,127.6 |
| 2Dt | Weibull | 14.65 | *3* | 11.51 | 32.43 | 3.50 | 613.8 | = UR | 0.176 | −2,127.0 |
| Unqual female reproductive success and equal pollen pool | | | | | | | | | | |
| 2Dt | Weibull | 330.0 | *1000* | 9.26 | 32.17 | 3.69 | 558.4 | 830.4 | 0.204 | −2,123.9 |
| Exp | Weibull | 14.42 | 3.60 | 10.17 | 34.83 | 3.38 | 395.2 | 568.0 | 0.190 | −2,123.1 |
| Unequal female reproductive success and unequal pollen pool | | | | | | | | | | |
| 2Dt | Weibull | 319.62 | *1000* | 8.97 | 31.70 | 3.78 | 581.3 | 955.6 | 0.210 | −2,108.0 |
| Seed dispersal only | | | | | | | | | | |
| lognormal | – | 2.86 | 0.58 | 20.72 | – | – | 112.1 | 148.0 | 0.053 | −2,135.4 |
| Weibull | – | 14.41 | 3.70 | 13.00 | – | – | 66.8 | 81.1 | 0.0684 | −2,135.4 |
| Without genetic information | | | | | | | | | | |
| 2Dt | Weibull | 320.0 | *1000* | 8.98 | 33.24 | 4.14 | 1,522.6 | = UR | 0.205 | – |

[a]Exp refers to the dispersal kernel given by the exponential power function
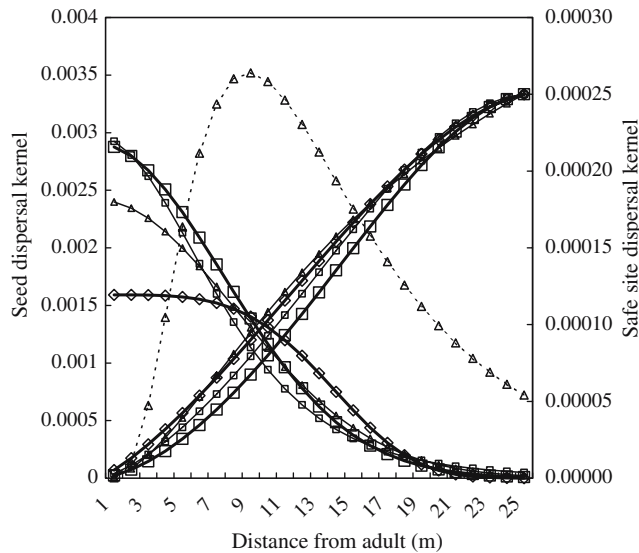
**Fig. 4** Dispersal kernels of the seed dispersal functions (*curves decreasing with distance*) and the survival rates (*curves increasing with distance*) according to the four optimized models [the latter are normalized at $F(25) = 0.00025$]. *Bold lines/open squares* 2Dt × Weibull ($\beta = 1,000$], *bold lines/open diamonds* exponential power × Weibull, *solid lines/triangles* 2Dt × log-normal ($\beta = 1,000$), *solid lines/open squares* 2Dt × Weibull ($\beta = 3$), *dotted line/triangles* model (5) based on the log-normal seed dispersal

was satisfied at the $10^{-4}$ level, i.e., left-hand-side of (10) $< 10^{-4}$, in which the solution was very likely at a (local) maximum.

For the safe site effects, the Weibull distribution exhibited better likelihood, thus we tested $U_L \neq U_R$ cases for this kernel function. The unequal female reproductive success assumption improved the log-likelihood, but only slightly (Table 2). Assuming that pollen pools were unequal did not sufficiently increase the likelihood either (Table 2).

For the equal reproduction and pollen pool cases, 2Dt and exponential power functions exhibited similar maximum log-likelihoods. However, for the expected $nac(r)$, the 2Dt showed better fitting than the exponential power function (as illustrated in Fig. 3b, for the adult-L case).

Excluding the safe site effects (model 5) resulted in solutions with much poorer maximum log-likelihoods (Table 2) for both Weibull and log-normal kernels. In particular, their expected $nac(r)$ showed the opposite tendency from the observations; increasing with distance from the adults (as illustrated in Fig. 3b, for the log-normal case). This is because under these skewed bell-shaped dispersal distributions, offspring near an adult mostly originate from external seed sources. On the other hand, when we applied the simple inhomogeneous process without genetic information (Eq. 11), we obtained similar parameter estimates to those obtained when we included genetic information (Table 2, the bottom row).

Therefore, we incorporated the assumption that the female reproductive success of the two adults and pollen

pools they received were equal, and implications from this type of models are discussed below.

According to the 2Dt and exponential power dispersal kernels, initial dispersal of offspring followed the curves illustrated by the bold lines that decrease with distance in Fig. 4, suggesting that most of the seeds fell within 20 m of their mother tree and that their average dispersed distances are similar (about 10 m, Table 2). However, the exponential power kernel exhibits a rapid decline around 15 m, whereas the 2Dt kernel merely gently decreases. Seed dispersal curves for 2Dt × lognormal (thin lines in Fig. 4) and 2Dt × Weibull ($\beta = 3$, the dotted line) functions are also illustrated, suggesting similar dispersal ranges. The survival rates increased with distance from the nearest adult, indicating that the crowns have a negative influence, up to a distance of 20–25 m, and this tendency is common to all the models (since the survival rates cannot be distinguished from the productivities, these curves were normalized so that they all take the probability 0.00025 at 25 m).

Fig. 5 illustrates the probability that a sapling at ($x$, 30) and ($x$, 15) is a daughter of adult-L, adult-R or another mother for the 2Dt × Weibull ($\beta = 1,000$) cases. Most offspring (about 90%) closer than 5 m to either adult were daughters of the respective adult, but in the area between the two adults (further than 20 m from both of them), more than 90 had unknown mothers.

The spatial distribution of individuals almost satisfactorily fitted the observations, except for the bottom parts in the plot for both solutions (as illustrated in Fig. 6 for the 2Dt, $\beta = 1,000$ case).

However, the observed $nac(r)$ were not covered within the 95% confidence envelopes drawn from 1,000 simulations (the better, 2Dt, case is illustrated in Fig. 3c). Hence, we have not found a model that can satisfactorily express the regeneration process in this *F. crenata* population.

## Discussion

The genetic inhomogeneous Poisson process was improved to include safe site effects on offspring survival. In addition, external seed sources and advanced regeneration are expressed as random constants and various dispersal kernels can be tested under the improved model. When regeneration occurs under simple environmental conditions such as under highly isolated adults in a flat stand, safe site conditions might be largely determined by relative positions from the isolated adults (due to shading effects of their crowns). In such cases, we can explicitly formulate the safe site effects, and our model can reflect these fundamental factors in forest regeneration and quantitatively describe its mechanism. In the case study in the shelterwood logged stand, important aspects of the successful regeneration of *F. crenata* included the following features: (1) seeds were dispersed mostly within 20 m of

**Fig. 5** Relative contributions of adult L (*black*), adult R (*white*) and the external seed sources (*shaded*) to the reproduction along the horizontal transects $y = 30$ (**a**) and $y = 15$ (**b**) derived from the optimized model; 2Dt × Weibull ($\beta = 1,000$). The ratios are proportional to $[U_L f(r_L), U_R f(r_R), m]$. The *x*-positions of L and R are indicated below the *x*-axis
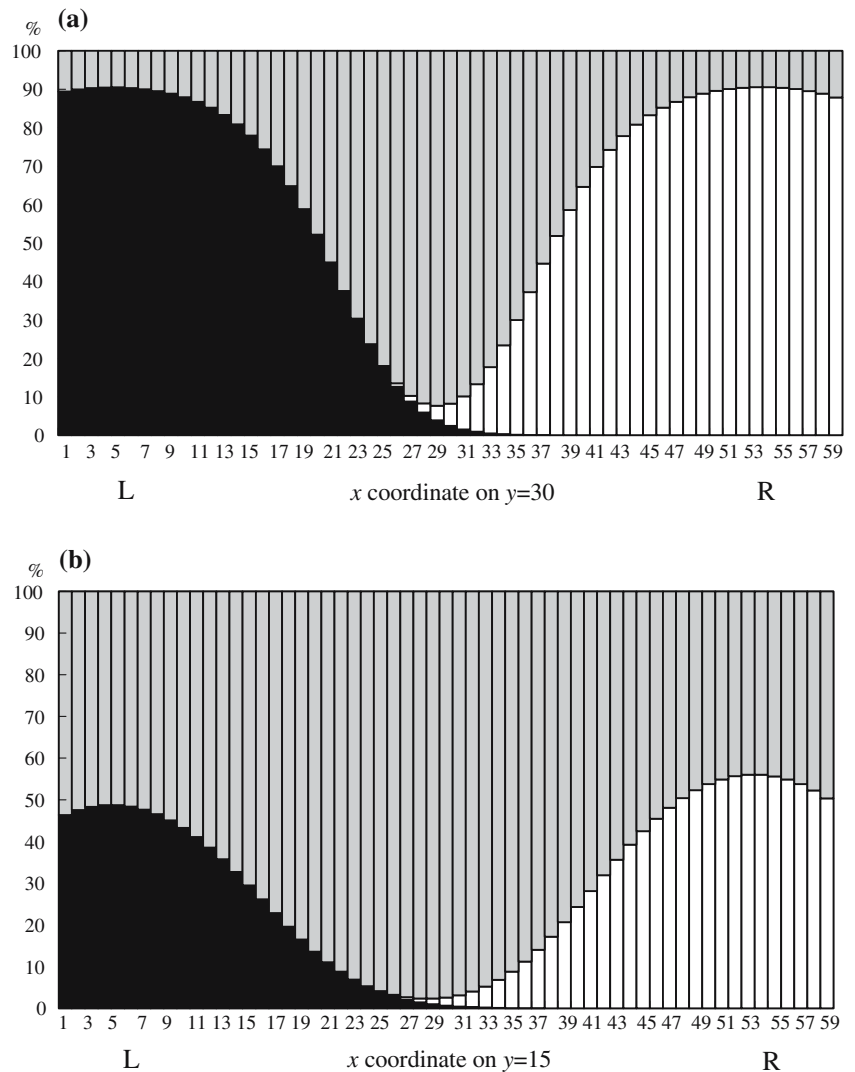


their mothers; (2) the two isolated adults had equal female reproductive success; (3) about 10% of offspring even close to an adult were not its daughters; (4) the adults' crowns reduced the survival of saplings up to 20–25 m away. Such quantitative evaluations have implications not only for population dynamics but also for silvicultural practices regarding aspects such as the ideal spacing between preserved adults, for which both seed dispersal patterns and the negative effects of preserved adults need to be considered (Kashimura et al. 1951; Loftis 1990; Schlesinger et al. 1993; Agestam et al. 2003; Bellocq et al. 2005), and the potential importance of the contribution of advanced regeneration prior to harvesting (Hannah 1987; Schlesinger et al. 1993; Shimatani 2002).

Comparisons with the patterns obtained when only seed dispersal was included in the model revealed that the clustering was not solely due to limited seed dispersal, or that analysis based solely on such a hypothesis would lead to wrong estimates of seed dispersal from established offspring data.
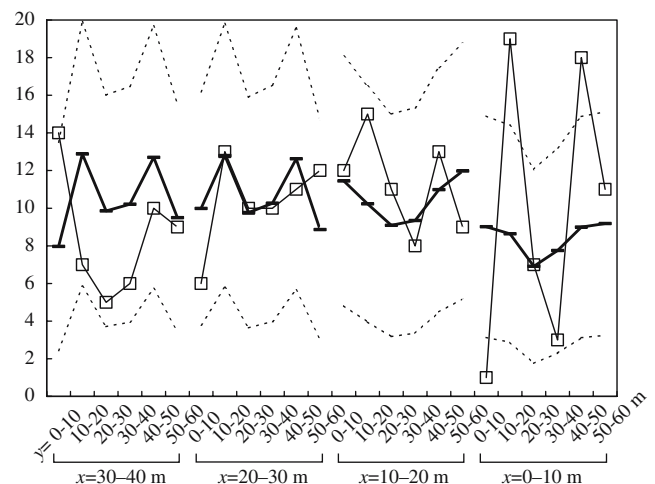


**Fig. 6** Expected (*bold lines/filled bars*) and observed (*thin lines/open square*) numbers of saplings in (*a*) [30, 40] × [0, 10], [30, 40] × [10, 20], ..., [0, 10] × [50, 60]. The dotted lines indicate the 95 %confidence intervals drawn from the Poisson distributions

None of the optimized models satisfactorily fitted the observations when the resulting spatial genetic structures were compared with the observed structures. Moreover, models using different dispersal kernels resulted in similar maximum log-likelihoods although the shapes of seed dispersal curves they generated differed substantially (Fig. 4). Nevertheless, the genetic models again showed a risk using spatial individual distribution data only because the unsatisfactory fitting would be missed if we simply optimized the inhomogeneous Poisson process without including genetic information (Eq. 11), as Fig. 6 exhibited satisfactory fitting.

The insufficient results are partly due to the difficulties associated with the maximum likelihood process, especially when the 2Dt distribution was used to model seed dispersal, and partly because we checked the completion of the process only by applying one necessity condition. However, we thought that, more importantly, our results implied the influences of some factors other than (1) – (4) (see Introduction) and difficulties in identifying such an unknown factor as far as using just the spatial individual distributions and isozyme data pertaining to them. Instead, additional approaches such as monitoring forests from seedling to sapling stages, genetic studies in other secondary stands or more polymorphic, microsatellite markers, may provide other descriptions about forest regeneration, and will give hints for improving the genetic inhomogeneous Poisson process.

Although genetic information has often been used to infer distances of seed dispersal for Fagaceae species, most previous inferences were based on descriptive statistics and statistical tests for randomness (e.g., Merzeau et al. 1994; Streiff et al. 1998; Takahashi et al. 2000; Asuka et al. 2004). When spatial data for both adults and juveniles are available, their mutual spatial patterns have also been intensively studied to investigate seed dispersal as well as their biotic interactions (e.g., Nanami et al. 1999; Debski et al. 2000; Arévalo and Fernández-Palacios 2003). However, use of such data for the former purpose is risky if genetic information is not also considered, as shown both here and in previous studies (Shimatani 2002, 2004). Statistics concerning genetic relatedness between adults and juveniles provide better inferences (e.g., Kawano and Kitamura 1997; Parker et al. 2001), and the genetic inhomogeneous Poisson processes can directly estimate a seed dispersal curve and other regeneration features, and also identify the possibility that other factors may be involved. Our approach provides an appropriate first step for modeling forest regeneration and provides useful quantitative information as well as indications of the limitations of the data and hypotheses.

## References

Agestam E, Ekö PM, Nilsson U, Welander NT (2003) The effects of shelterwood density and safe site preparation on natural regeneration of *Fagus sylvatica* in southern Sweden. For Ecol Manage 176:61–73

Arévalo JR, Fernández-Palacios JM (2003) Spatial patterns of trees and juveniles in a laurel forest of Tenerife, Canary Islands. Plant Ecol 165:1–10

Asuka Y, Tomaru N, Nishimura N, Tsumura Y, Yamamoto S (2004) Heterogeneous genetic structure in a *Fagus crenata* population in an old-growth beech forest revealed by microsatellite markers. Mol Ecol 13:1241–1250

Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, Sork EK (2004) Using genetic markers to estimate the pollen dispersal curve. Mol Ecol 13:937–954

Bellocq MI, Jones C, Dey DC, Turgeon JJ (2005) Does the shelterwood method to regenerate oak forests affect acorn production and predation? For Ecol Manage 205:311–323

Berg EE, Hamrick JL (1995) Fine-scale genetic structure of a turkey oak forest. Evolution 49:110–120

Brose PH, Van Lear DH (1998) Responses of hardwood advance regeneration to seasonal prescribed fires in oak-dominated shelterwood stands. Can J For Res 28:331–339

Calviño-Cancela M (2002) Spatial patterns of seed dispersal and seedling recruitment in *Corema album* (Empetraceae): the importance of unspecialized dispersers for regeneration. J Ecol 90:775–784

Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: patterns across temperate and tropical forests. Ecology 80:1475–1494

Cressie N (1991) Statistics for spatial data. Wiley, New York

Debski I, Burslem DFRP, Lamb D (2000) Ecological processes maintaining differential tree species distributions in an Australian subtropical rain forest: implications for models of species coexistence. J Trop Ecol 16:387–415

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5:615–627

Godoy JA, Jordano P (2001) Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. Mol Ecol 10:2275–2283

González-Martínez SC, Gerber S, Cervera MT, Martínez-Zapater JM, Gil L, Alía R (2002) Seed gene flow and fine-scale structure in a Mediterranean pine (*Oinus pinaster* Ait.) using nuclear microsatellite markers. Theor Appl Genet 104:1290–1297

Hannah PR (1987) Regeneration methods of oaks. NJAF 4:97–101

Harper JL (1977) Population biology of plants. Academic, London

Houle G (1998) Seed dispersal and seedling recruitment of *Betula alleghaniensis*: spatial inconsistency in time. Ecology 79:807–818

Kashimura D, Saito H, Kida S (1951) Experiments of the shelterwood system in the buna (*Fagus crenata* Blume) forest - the 1st report: Statistical analysis of the stand structure (In Japanese with English abstract). J Jpn For Soc 33:265–268

Kawano S, Kitamura K (1997) Demographic genetics of the Japanese beech, *Fagus crenata*, in the Ogawa Forest Preserve, Ibaraki, Central Honshu, Japan. III. Population dynamics and genetic substructuring within a metapopulation. Plant Species Biol 12:157–177

Kitamura K, Shimada K, Nakashima K, Kawano S (1997) Demographic genetics of the Japanese beech, *Fagus crenata*, at the Ogawa Forest Preserve, Ibaraki, Central Honshu, Japan. I. Spatial genetic substructuring in local population. Plant Species Biol 12:107–135

Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T (2000) Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol Ecol 9:1843–1852

Loftis DL (1990) A shelterwood method for regenerating red oak in the southern Appalachians. For Sci 36:917–929

Merzeau D, Comps B, Thiébaut B, Cuguen J, Letouzey J (1994) Genetic structure of natural stands of *Fagus sylvatica* L. (beech). Heredity 72:269–277

Nanami S, Kawaguchi H, Yamakura T (1999) Dioecy-induced spatial patterns of two codominant tree species, *Podocarpus nagi* and *Neolitsea aciculate*. J Ecol 87:678–687

Parker KC, Hamrick JL, Parker AJ, Nason JD (2001) Fine-scale genetic structure in *Pinus clausa* (Pinaceae) populations: effects of disturbance history. Heredity 87:99–113

Rathbun SL, Cressie N (1994) A space-time survival point process for a longleaf pine forest in Southern Georgia. J Am Stat Assoc 89:1164–1174

Sakamoto Y (1991) Categorical data analysis by AIC. KTK, Tokyo

Schlesinger RC, Sander IL, Davidson KR (1993) Oak regeneration potential increased by shelterwood treatments. NJAF 10:149–153

Shimatani K (2002) Point processes for fine-scale spatial genetics and molecular ecology. Biom J 44:325–352

Shimatani K (2004) Spatial molecular ecological models for genotyped adults and offspring. Ecol Model 174:401–410

Stoyan D, Wagner S (2001) Estimating the fruit dispersion of anemochorous forest trees. Ecol Model 145:35–47

Stoyan D, Penttinen A (2000) Recent applications of point process methods in forestry statistics. Stat Sci 15:61–78

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Mol Ecol 7:317–328

Sugita H, Kanazashi T, Masaki T (2006) Regeneration of *Fagus crenata* experimentally managed by a shelterwood system 33 or 54 years after cutting in forests with deciduous shrub-type undergrowth in the Tohoku District, northern Honshu, Japan. J Jpn For Soc (in press)

Surles SE, Arnold J, Schnabel A, Hamrick JL, Bongarten BC (1990) Genetic relatedness in open-pollinated families of two leguminous tree species, *Robinia pseudoacacia* L. and *Gleditsia triacanthos* L. Theor Appl Genet 80:49–56

Takahashi M, Mukouda M, Koono K (2000) Differences in genetic structure between two Japanese beech (*Fagus crenata* Blume) stands. Heredity 84:103–115

Yanagiya S, Kon T, Konishi A (1969) The results of natural regeneration with bush clearing by remaining tree management system in Buna (*Fagus crenata*) forest (in Japanese). Annual Report of the Tohoku Branch Forest Experimental Station 10:124–135

www.manaraa.com