

Seed dispersal kernels estimated from genotypes of established seedlings: does density-dependent mortality matter?

Etienne K. Klein^{1,2*}, Aurore Bontemps² and Sylvie Oddou-Muratorio²

¹INRA, UR546 Biostatistique et Processus Spatiaux (BioSP), F-84914 Avignon, France; and ²INRA, UR629 Ecologie des Forêts Méditerranéennes (URFM), F-84914 Avignon, France

Summary

1. The seed dispersal kernel is a major determinant of spatial population dynamics and spatial distribution of genetic diversity. Among the main methods to estimate it, inverse modelling (IM) and gene shadow model (GSM) rely on seed counts in traps, whereas competing source model (CSM) and spatially explicit mating models (SEMMs) rely on compositions of seed pools. Moreover, GSM, CSM and SEMM exploit genetic information from molecular markers, whereas IM only exploits seed counts ignoring seed origins. These methods were also applied to established seedlings.

2. In the presence of post-dispersal density-dependent mortality (DDM), the effective dispersal kernel, describing the spatial distribution of established seedlings relatively to the seed source, is notoriously different from the basic dispersal kernel, describing the spatial distribution of seed deposition sites relatively to the source. Using simulated data sets, we investigated whether IM, GSM, CSM and SEMM applied to established seedlings estimate the basic or the effective dispersal kernel. In our simulations, DDM resulted in a shift of the mean basic dispersal distance (10 m) towards substantially higher effective mean dispersal distances (15 m and 20.8 m).

3. We demonstrated that CSM and SEMM estimate the basic seed dispersal kernel, independently from the presence of post-dispersal mortality. By contrast, GSM estimates the effective dispersal kernel. IM failed to provide satisfactory estimates in the presence of DDM in our sampling design. Besides, for all methods, seed migration was inflated in the presence of DDM, due to lower mortality among randomly distributed immigrants relatively to local seedlings.

4. It could seem intuitive that estimates based on seedlings or seeds provide effective or basic dispersal kernels respectively. Our results showed that it is not true for estimates obtained with CSM or SEMM because they rely on the composition of seed/seedling pools and not seed/seedling counts such as IM or GSM. This has important consequences for life stage studies where the discordance of dispersal kernels estimated from different cohorts is used to investigate post-dispersal density-dependent mortality.

Key-words: microsatellite markers, basic dispersal kernel, effective dispersal kernel, Janzen–Connell hypothesis, Inverse Modelling, Spatially Explicit Mating Model, Competing Source Model, Gene Shadow Model

Introduction

Estimation of seed dispersal kernels has been an active field of research in the last two decades (Cousens, Dytham & Law 2008; Nathan *et al.* 2012). In plants, seed dispersal is a major process determining the spatial distribution of the next generation relatively to the distribution of the parental population (Schupp & Fuentes 1995; Clark *et al.* 1999a). Post-dispersal processes such as seed germination and seedling mortality then reshape the primary seed dispersal pattern to provide the effective pattern of dispersal from one generation to the next (Clark *et al.* 1999a; Nathan & Muller-Landau 2000; Nathan & Casagrandi 2004). In particular, seedlings generally suffer from

pervasive density-dependent mortality (DDM) that can be driven by seedling density or distance to the mother, for example, due to herbivory, pathogens or competition for light or water (Janzen 1970; Connell 1971). When DDM is strong enough to result in increasing density of surviving seedlings up to a certain distance from the mother tree, this is acknowledged as the Janzen–Connell effect (Janzen 1970; Connell 1971; Schupp & Jordano 2011). Less intense DDM does not inverse the trend of decreasing the density of seedlings from the mother tree, but still increases the effective dispersal distance of surviving seedlings (Hubbell 1980). In these cases, seed primary dispersal and post-dispersal mortality are intrinsically linked in determining the effective dispersal distance between a mother plant and its progenies that reach older life stages (Augsburger 1983; HilleRisLambers, Clark & Beckage 2002).

*Correspondence author. E-mail: etienne.klein@avignon.inra.fr

Basic and effective dispersal kernels are two curves that characterize dispersal and account for these processes (Nathan *et al.* 2012). Formally, a dispersal kernel is a 2-dimensional probability density function describing the probability for a seed (or seedling) to disperse (or disperse and establish) to any position relative to the maternal tree. The basic dispersal kernel is defined as encompassing only the movement and survival of the seeds during dispersing movements. The effective dispersal kernel additionally includes the post-arrival survival of seeds that lead to successful establishment. When survival probability depends on the distance to the nearest adults, the two kernels are expected to differ (González-Martínez *et al.* 2006; Martin & Canham 2010; Steinitz *et al.* 2011; Tautenhahn *et al.* 2012). In published studies, it is not always clear which of the two kernels is estimated, but it seems implicitly accepted that estimations resulting from established seedlings provide effective kernels, whereas estimations from seeds provide basic kernels. However, we expect that the statistical method used to derive the dispersal kernel will also have an effect.

Four major statistical methods were developed to estimate dispersal kernels from seed or seedling data. Inverse modelling (IM) is the classical method to analyse the spatial pattern of dispersed seeds (Clark *et al.* 1999b). This method accounts for the overlap of seed shadows by comparing seed counts obtained at trapping locations to the expected number of seeds computed from the basic seed dispersal kernel and individual adult fecundities. IM has also been applied to established seedlings (Ribbens, Silander & Pacala 1994; LePage *et al.* 2000; Kunstler *et al.* 2007), in which case it estimates an effective dispersal kernel. Hence, comparing IM estimates of the kernels obtained at successive recruitment stages in the same site allows disentangling seed dispersal and post-dispersal mortality (e.g. LePage *et al.* 2000; Kunstler *et al.* 2007; Amm *et al.* 2012). However, using IM to estimate the basic dispersal kernel can pose some challenges. For example, most seed-trap devices do not integrate secondary animal-assisted dispersal. More generally, in masting species, in species continually producing seeds or when sampling seed just after dispersal is technically complicated, sampling established seedlings is a simpler option.

Alternative and increasingly popular statistical approaches rely on genetic parentage assignment (Jones *et al.* 2005; Burczyk *et al.* 2006; Robledo-Arnuncio & García 2007; Oddou-Muratorio & Klein 2008; Klein & Oddou-Muratorio 2011). Parentage assignment typically entails mapping of seeds or seedlings and their potential parents within a circumscribed area and genotyping all individuals with molecular markers. Ideally, a unique mother can be retrieved for each seedling (favoured by the genotyping of maternal tissues present on seeds or newly germinated seedlings). Then the GSM (gene shadow model, Jones & Muller-Landau 2008) and CSM (competing source model, Robledo-Arnuncio & García 2007) can both estimate a dispersal kernel, accounting for the location of all seed sources. The GSM and CSM rely, respectively, on the number and proportion of seeds from each source in each trap. Both methods were compared on simulated seed genotypes in Jones & Muller-Landau (2008) and provided similarly good

estimates of the dispersal kernel. To our knowledge, no comparable study was published concerning the analysis of seedling genotypes in the presence of DDM.

When the genetic system and/or genotyped tissue do not enable the mother of each progeny to be retrieved categorically, the dispersal kernel can be estimated using full-probability methods belonging to the broad class of spatially explicit mating models (SEMMs) (Burczyk *et al.* 2006; Oddou-Muratorio & Klein 2008; Klein & Oddou-Muratorio 2011; Moran & Clark 2011). Similar to the methods above, SEMMs consider the overlap of seed shadows from different trees when computing the probability of a given seedling genotype at a given location. This probability is a function of the seed and pollen dispersal kernels, the locations and male and female fecundities of adult trees and the genotypes of the candidate parents. In contrast with IM and GSM, but similarly to CSM, SEMMs focus on the composition of seed/seedling pools instead of seed/seedling rain intensity.

Here, we used mathematical arguments and individual-based simulations to characterize the seed/seedling dispersal kernels estimated by IM, GSM, CSM and SEMM applied to established seedlings, after post-dispersal recruitment processes. We questioned whether these methods estimate the basic or the effective dispersal kernel. We considered successively the impact of density-dependent mortality (DDM) and environmentally driven mortality (EDM).

Materials and methods

THEORETICAL ARGUMENTS

We assume that the modelling framework of dispersal kernels correctly represents seed dispersal patterns. The density I_t of seed rain in a trap t , that is, the amount of seeds per unit area, is given by (e.g. Nathan *et al.* 2012):

$$I_t = \sum_{j, \text{sources}} Q_j k_b(x_t - x_j, y_t - y_j), \quad \text{eqn 1}$$

where k_b is the 2D basic dispersal kernel, Q_j is the intensity of seed source j (i.e. the effective amount of seeds released), and (x_j, y_j) and (x_t, y_t) are the locations of the sources j and the traps t .

The composition of seed rain in trap t , that is, the proportion of seeds originating from one source j , is given by (e.g. Nathan *et al.* 2012)

$$\pi_{tj} = \frac{Q_j k_b(x_t - x_j, y_t - y_j)}{\sum_{j', \text{sources}} Q_{j'} k_b(x_t - x_{j'}, y_t - y_{j'})} \quad \text{eqn 2}$$

If post-dispersal mortality occurs with probability $d(x, y)$ at each location (x, y) , the density \tilde{I}_t of established seedling (i.e. the amount of seedlings per unit area) is given by

$$\tilde{I}_t = \left(\sum_{j, \text{sources}} Q_j k_b(x_t - x_j, y_t - y_j) \right) (1 - d(x_t, y_t)). \quad \text{eqn 3}$$

For DDM determined by the distance to neighbour competitor plants (e.g. adult trees) following a field-of-neighbourhood approach (Berger *et al.* 2008), then $d(x_t, y_t)$ is defined from an influence function $\text{Comp}()$ as follows:

$$\sum_{j, \text{competitors}} \text{Comp}(x_t - x_j, y_t - y_j), \quad \text{eqn 4}$$

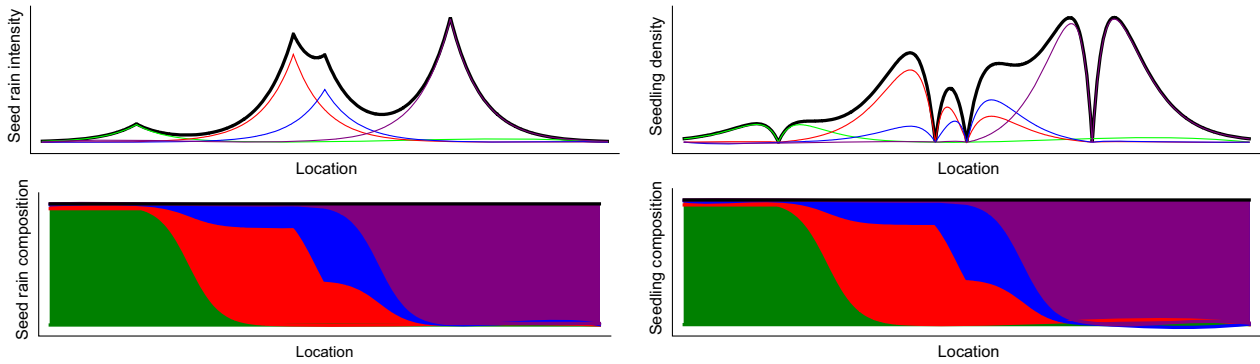


Fig. 1. Schematic representation of a 1D case with four sources (one colour for each source). Top panels represent the seed rain intensity I_t and seedling density \tilde{I}_t (black curves). Bottom panels represent seed rain composition π_{ij} and seedling composition $\tilde{\pi}_{ij}$.

The effective dispersal kernel k_e can then be defined at this stage as the kernel that best satisfies over all traps t

$$\tilde{I}_t \approx \sum_{j, \text{sources}} \tilde{Q}_j k_e(x_t - x_j, y_t - y_j), \quad \text{eqn 5}$$

where \tilde{Q}_j is an effective amount of ‘established seedlings’ released by tree j . Note that generally the exact equality cannot be reached, but the effective kernel is defined as an ‘average’ kernel (Tautenhahn *et al.* 2012).

If post-dispersal mortality occurs with probability $d(x, y)$ at each location (x, y) , the composition of established seedling pools can be written as follows:

$$\tilde{\pi}_{ij} = \frac{Q_j k_b(x_t - x_j, y_t - y_j)(1 - d(x_t, y_t))}{\sum_{j', \text{sources}} Q_{j'} k_b(x_t - x_{j'}, y_t - y_{j'})(1 - d(x_t, y_t))} = \pi_{ij} \quad \text{eqn 6}$$

$\tilde{\pi}_{ij}$ thus simplifies to π_{ij} as long as the mortality rate $d(x, y)$ at any location does not depend on the source origin j of the seed. This encompasses DDM driven by adults or other seedlings density and also environmentally driven mortality (EDM), as soon as all genotypes respond similarly to density or environment.

Figure 1 provides a 1D illustration of the densities I_t and \tilde{I}_t and compositions π_{ij} and $\tilde{\pi}_{ij}$. Although the intensity of seed rain is largely modified by DDM to determine the seedling density, the composition of the seed rain is not modified. Therefore, the statistical methods taking as input density of seed/seedlings (IM and GSM) should be sensitive to post-dispersal mortality, whereas the statistical methods taking as input the composition of seedling pools (i.e. CSM and SEMM) should estimate the basic dispersal kernel, even if mortality largely modifies the spatial distribution of seedlings. Using a simulation study, we investigate below this expected result.

SIMULATION STUDY

We ran simulations to investigate whether post-dispersal mortality occurring between seed dispersal and seedling stage affects the four methods classically used to estimate kernels (IM, GSM, CSM and SEMM). We used a three-step procedure: (1) simulate positions, diameters and genotypes for adults in a 200×200 m plot and simulate seed rain in a 100×100 m central plot according to known seed and pollen dispersal kernels; (2) simulate post-dispersal mortality and sample among surviving seedlings; and (3) estimate the dispersal and fecundity parameters using IM, GSM, CSM and SEMM. We chose parameter

values and sampling designs following experimental studies in beech (*Fagus sylvatica*) reported recently (Bontemps, Klein & Oddou-Muratorio 2013; Millerón *et al.* 2013).

Simulation procedure (steps 1 and 2)

We simulated a 4-ha plot (200×200 m) in which adult trees emitted seeds with individual fecundity determined by diameter at breast height (DBH). Trees were distributed at random with a density of 50 trees/ha (200 trees on average in the plot). DBH followed a truncated Gaussian distribution with average = 15 cm and SD = 4 cm. Adult genotypes were drawn following Hardy–Weinberg frequencies, at 10 independent loci with 10 alleles each and even allelic frequencies (single parent exclusion power = 0.9999998; see Appendix S3 for results with different genetic systems and uneven allelic frequencies).

Then we simulated 500 seedlings collected in 49 pre-defined traps regularly located on a grid in the 1-ha (100×100 m) central plot of the site. Each trap was a 1-metre-radius disc (Fig. 2). We considered three types of scenarios to simulate seed dispersal, post-dispersal mortality and sampling.

Scenario with no mortality (NoM)

Each tree j with diameter at breast height DBH_j and position (x_j, y_j) released an amount of seed Q_j

$$Q_j = q_s \text{DBH}_j \rho_s,$$

with the fecundity parameters $\rho_s = 2.0$ and q_s to be estimated (see eqn 10 and eqn 11). The distribution of DBH and $\rho_s = 2.0$ lead to a coefficient of variation of Q_j equal to 50%. Each seed dispersed at a location (x, y) following an exponential-power dispersal kernel centred on the source location:

$$k_b(x - x_j, y - y_j) = \frac{a_s}{2\pi\Gamma(2/b_s)} \exp\left(-\frac{r^{b_s}}{a_s}\right), \quad \text{eqn 7}$$

where k_b is the basic dispersal kernel, a_s is the scale parameter (homogeneous to a distance), b_s is the shape parameter determining the tail of the kernel, and r is the distance from (x_j, y_j) to (x, y) . The mean dispersal distance is $\delta_s = a_s \Gamma(3/b_s) / \Gamma(2/b_s)$. We used $\delta_s = 10$ m and $b_s = 0.5$ in all simulations.

Additionally to seeds generated by trees within the plot, a proportion m_s of seeds originating from unknown parents immigrated into the

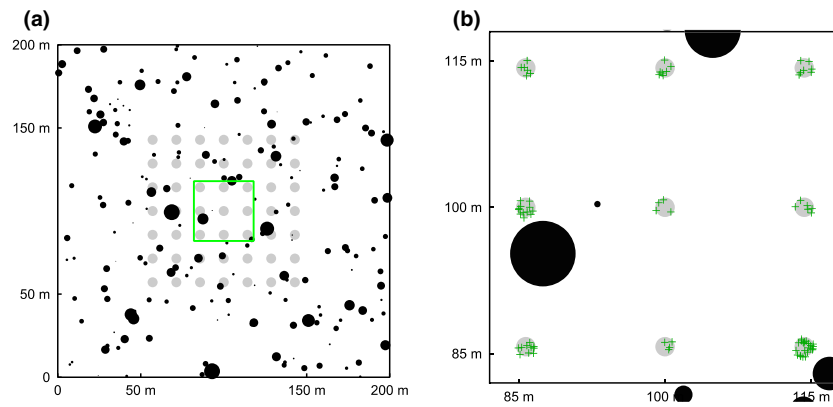


Fig. 2. Example of one simulated population. Black discs represent adult trees, radii being proportional to DBH. Grey discs represent 1-m-radius seed traps. Panel b: zoom on the green box from (a). Green '+' represent seedlings located in the seed traps.

plot. Their locations were drawn uniformly at random, and their genotypes were drawn from the allelic frequencies. We used $m_s = 0.3$.

We drew seeds until we gathered 500 of them over all traps (see Appendix S1 for technical aspects). This corresponds to a density of immigrant seeds, Im_s , of 0.97 seeds/m².

For each trapped seed s , we recorded its mother tree M_s and drew a father using an exponential-power pollen dispersal kernel with parameters $\delta_p = 75$ m and $b_p = 0.5$, no selfing, a pollen migration rate of 0.6 and male fecundities proportional to $DBH_j^{0.9}$ with

$\rho_p = 1.5$ (coefficient of variation for male fecundity = 40%) (see algorithm 3 in Appendix S1). A genotype was drawn for the seed from the genotypes of the mother and father trees following Mendelian rules.

This scenario 'no mortality' is equivalent to a scenario where the mortality rate between seed and seedling stages is equal for all seeds. We simulated 100 data sets following this scenario (NoM) (Fig. 3a).

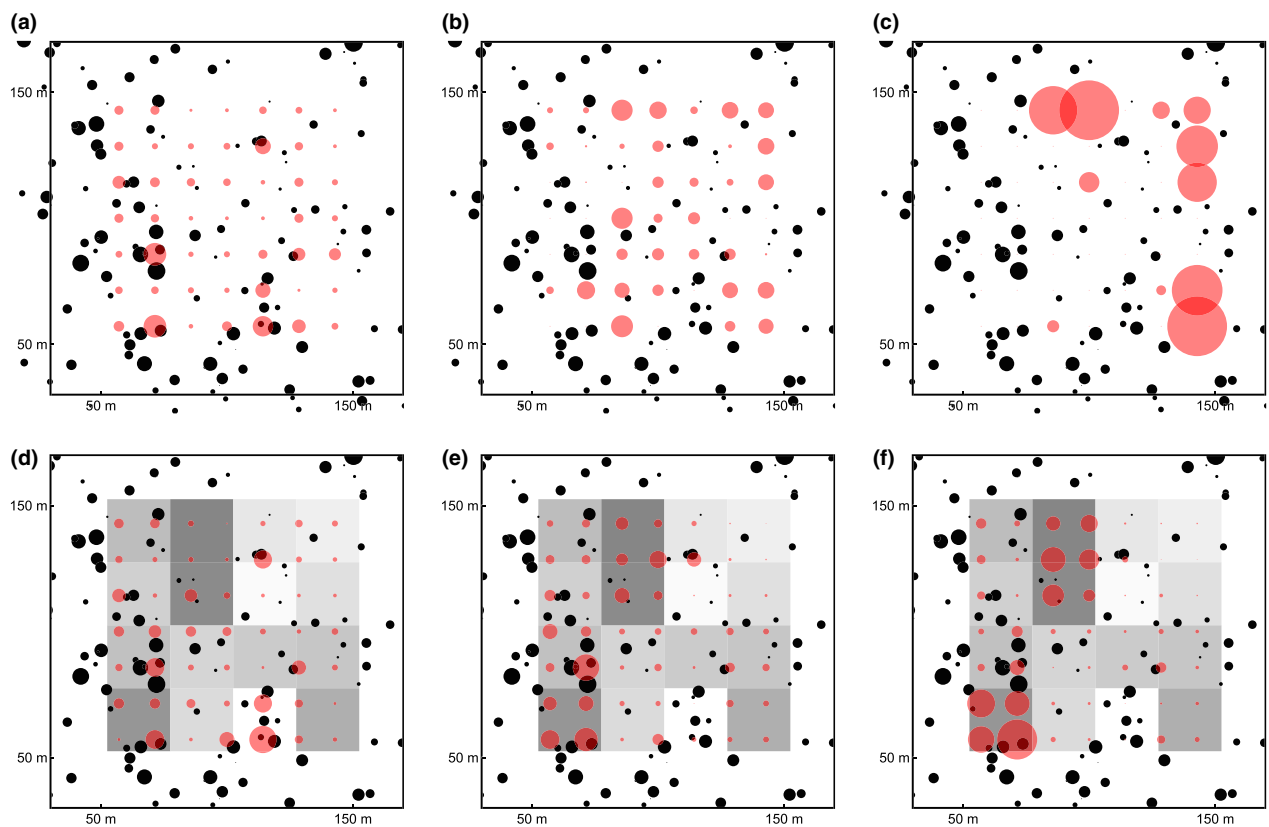


Fig. 3. Seedling counts in the 49 seedling traps under different mortality regimes. Red discs represent the 49 sampling traps, diameters being proportional to the number of seedlings sampled (total always equal to 500). Top panels represent different levels of density-dependent mortality (DDM): (a) NoM; (b) low DDM; (c) high DDM. Bottom panels represent different levels of environmentally driven mortality: (d) NoM; (e) low EDM; (f) high EDM. Grey squares represent mortality rates driven by environmental conditions (darker grey stands for better survival).

Scenarios with density-dependent mortality (low and high DDM)

Using the same scheme to simulate seed dispersal as in the NoM scenario, we additionally simulated seedling mortality events driven by proximity to adult trees. The level of competition CI_s experienced by each seedling s at location (x_s, y_s) was related to the distances to all adults following a Gaussian competition kernel (i.e. *Comp* in eqn 4 is Gaussian):

$$CI_s = \sum_{j, \text{competitors}} \exp\left(-\frac{(x_s - x_j)^2 + (y_s - y_j)^2}{2\tau^2}\right), \quad \text{eqn 8}$$

where τ determines the spatial scale at which competition occurs. We used $\tau = 7$ m.

Each seed/seedling died before sampling with probability

$$d_s = \min(1, \alpha \times CI_s) \quad \text{eqn 9}$$

where α determines the intensity of mortality. We used $\alpha = 0.5$ and 1 for low- and high-mortality scenarios (Fig. 3b,c).

We drew seeds until we had 500 recruited seedlings in the traps (Appendix S1, algorithm 5). Fathers and genotypes were drawn for the seedlings just as in the NoM case. We simulated 100 data sets following this procedure for $\alpha = 0.5$ (low DDM) and $\alpha = 1$ (high DDM).

Using algorithm 6 in Appendix S1, we computed the mortality rates, the true effective dispersal kernels and the effective migration rates after mortality under each DDM scenario.

Scenarios with environmentally dependent mortality (low DDM and high EDM)

Using the same scheme to simulate seed dispersal as in the NoM scenario, we additionally simulated seedling mortality events driven by a spatially structured environmental variable. For each simulation, the 1-ha plot was divided into 16 quadrats, each quadrat q with a random mortality rate d_q uniformly distributed between 0 and 1. Each seed at location (x, y) in a trap t in a quadrat q died at random with mortality rate d_q (Fig. 3d–f).

We drew seeds until we had 500 recruited seedlings in any trap. Fathers and genotypes were drawn as before. We simulated 100 data sets following this procedure using mortality rates d_q (low EDM) and 100 data sets using mortality rates $1 - (1 - d_q)^2$ (high EDM) (see Fig. 3e,f).

Kernel estimation procedure

To apply inverse modelling (IM), we computed the number of seeds n_t collected in each trap t . Parameters δ_s , b_s , ρ_s , Im_s , q_s were estimated by maximizing the likelihood

$$l_{IM} = \prod_t e^{-A \times Im_t} \frac{(A \times Im_t)^{n_t}}{n_t!}, \quad \text{eqn 10}$$

$$\text{with } I_t = Im_s + \sum_k q_s DBH_k^{\rho_s} k(x_t - x_k, y_t - y_k),$$

where A is the area of each trap, Im_s is the density of immigrant seeds (see Jones & Muller-Landau 2008), and δ_s and b_s are the parameters of the exponential-power kernel k (see eqn 7).

We applied CSM and GSM (Robledo-Arnuncio & García 2007; Jones & Muller-Landau 2008), by assuming that the genetic system is powerful enough to retrieve surely the mother tree M_s that generated

each sampled seedlings s . For the GSM (Jones & Muller-Landau 2008), we computed the number of seeds $n_t(j)$ originating from each source j collected in each trap t and $n_t(\infty)$ originating from outside the study site. Parameters δ_s , b_s , ρ_s , Im_s , q_s were then estimated by maximizing the likelihood

$$l_{GSM} = \prod_t e^{-A \times Im_t} \frac{(A \times Im_t)^{n_t(\infty)}}{n_t(\infty)!} \prod_j e^{-A \times I_t(j)} \frac{(A \times I_t(j))^{n_t(j)}}{n_t(j)!}, \quad \text{eqn 11}$$

$$\text{with } I_t(j) = q_s DBH_j^{\rho_s} k(x_t - x_j, y_t - y_j)$$

For the CSM (Robledo-Arnuncio & García 2007), parameters δ_s , b_s , ρ_s , m_s were estimated by maximizing the likelihood

$$l_{CSM} = \prod_s \pi_{sM_s}, \quad \text{eqn 12}$$

where the index s runs over all genotyped seedlings and with

$$\pi_{sj} = (1 - m_s) \frac{DBH_j^{\rho_s} k(x_s - x_j, y_s - y_j)}{\sum_{j'} DBH_{j'}^{\rho_s} k(x_s - x_{j'}, y_s - y_{j'})} \quad \text{eqn 13}$$

for each j corresponding to a tree inside the plot and $\pi_{s\infty} = m_s$ for $j = \infty$ (i.e. mother tree outside the plot).

For the SEMM (Burczyk *et al.* 2006; Oddou-Muratorio & Klein 2008), parameters δ_s , b_s , ρ_s , m_s and δ_p , b_p , ρ_p , m_p were estimated by maximizing the likelihood

$$l_{SEMM} = \prod_s \left[m_s T(g_s | \text{BAF}) + \sum_j \pi_{sj} \left[m_p T(g_s | G_j, \text{BAF}) + (1 - m_p) \sum_k \Phi_{jk} T(g_s | G_j, G_k) \right] \right] \quad \text{eqn 14}$$

where π_{sj} is defined in eqn (13) and Φ_{jk} follows eqn (1) in Appendix S1. The genotypes of seeds s and adults j or k are, respectively, g_s , G_j and G_k . $T(g|\dots)$ stands for Mendelian probabilities to draw an offspring genotype g given the genotypes of the parents or background allelic frequencies (for detailed definition, Oddou-Muratorio & Klein 2008).

Because CSM and SEMM rely on the proportions π_{sj} , they cannot estimate the absolute fecundity parameter q_s because it simplifies in the ratio given by eqn (13).

For all four methods, we used either the true exponential-power family to fit a kernel to the observed data or alternatively the log-normal family (Stoyan & Wagner 2001). The first family provides kernels that strongly peaked in 0, and the second family provides hump-shaped curves more suited to effective dispersal kernels in the presence of DDM (Nathan & Casagrandi 2004). All likelihoods were maximized in Mathematica 8 following a quasi-Newton algorithm (Wolfram Research).

EFFECT OF SAMPLING SCHEME AND GENETIC MARKERS FOR CSM AND SEMM

In addition to seed/seedlings sampled in traps, we also simulated an experiment where seedlings were collected at random among all the seedlings that established within the 100x100 m central square in the site (Algorithm 4 in Appendix S1, Figure S1). For this type of sampling, IM and GSM are not easy to implement. We only applied CSM and SEMMs.

Results

BASIC AND EFFECTIVE DISPERSAL KERNELS

Under low DDM and high DDM, mortality due to competition reached, respectively, 79.4% (84% for non-migrants; 69% for migrants) and 94.4% (97% for non-migrants; 89% for migrants). After mortality, the effective migration rate was 0.46 and 0.59 under low DDM and high DDM.

As expected, the two levels of DDM resulted in a difference between basic and effective dispersal kernels: the effective

kernel was more hump-shaped than the basic kernel (Fig. 4). The mean effective dispersal distance increased from 10 m for NoM (equal to δ_s) to 15.0 m and 20.8 m for low DDM and high DDM. The mode of the dispersal distance distribution clearly shifted towards larger values (see Fig. 4).

Under low EDM and high EDM, post-dispersal mortality rates were 51.5% and 68.0% for all types of seeds. The effective dispersal kernel after mortality was not different from the basic dispersal kernel (Figure S2). This is because the quality of environment was distributed independently of the position of adults.

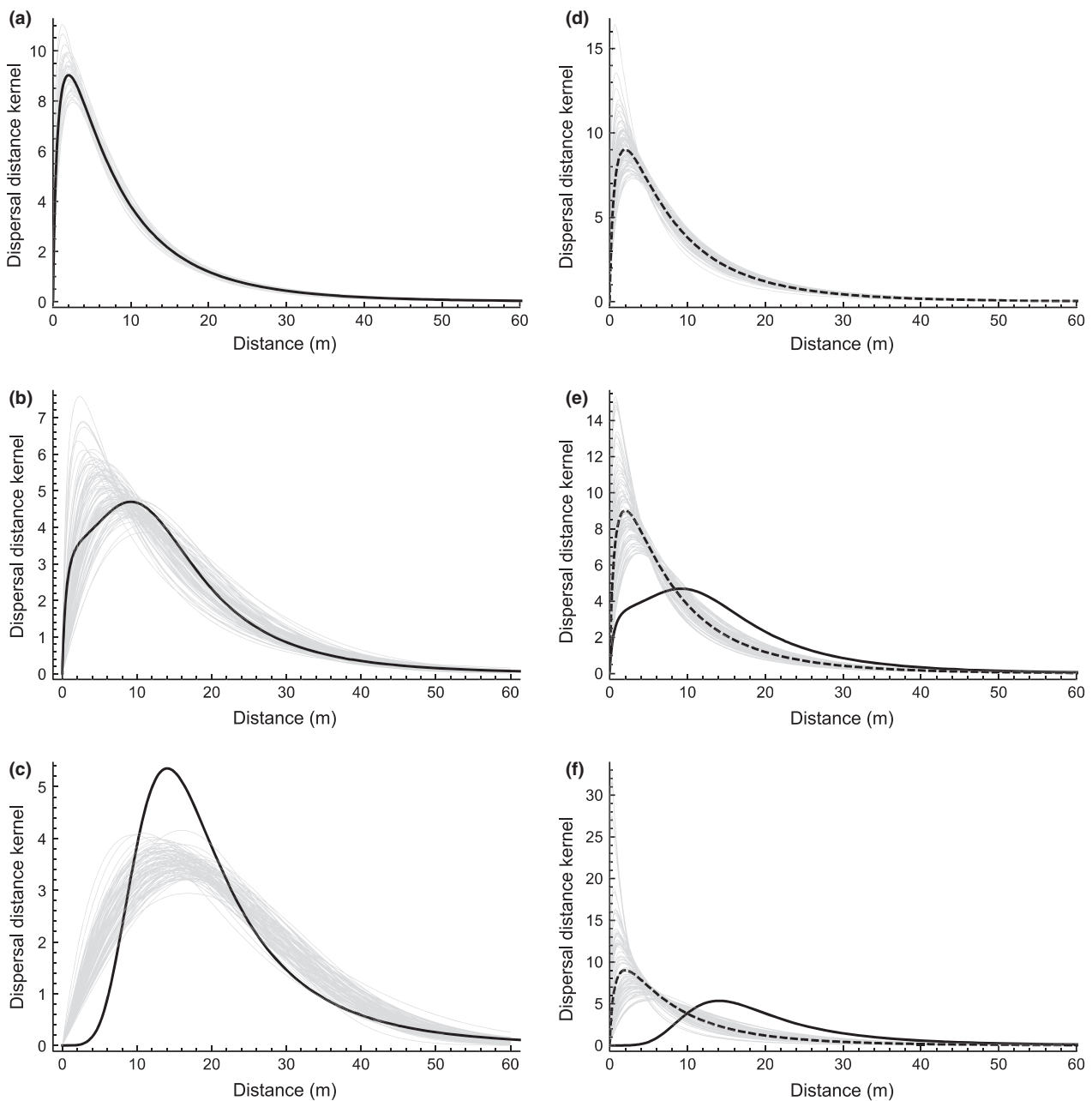


Fig. 4. Basic, effective and estimated dispersal kernels under different regimes of density-dependent mortality (DDM). a–c: 60 kernels estimated with GSM (grey) and effective dispersal kernels (black). d–f: 60 kernels estimated with CSM (grey), basic dispersal kernel (black, dashed) and effective dispersal kernels (black, solid). a & d: No mortality, b & e: low DDM, c & f: high DDM.

SENSITIVITY OF IM, GSM, CSM AND SEMMS TO DENSITY-DEPENDENT MORTALITY

In the absence of mortality, all methods estimated the dispersal parameters with a low bias (Table 1; Figure S3). IM was the less precise method with a larger variability for dispersal parameters δ_s and b_s and an important proportion of cases where the estimation failed (20% of simulated data sets, Table 1). The three other methods, GSM, CSM and SEMM, reached comparable performances (Table 1, Figure S3). Seed immigration was well estimated by the CSM and SEMM, where it is modelled through a migration rate ($m_s = 0.30$, Table 1), and by the IM and GSM, where it is modelled through an immigrant density ($Im_s = 0.97/\text{m}^2$). These two parameters were estimated with low bias and low estimation variability (Table 1).

In the presence of DDM, (i) IM estimates failed to converge for all simulated data sets, (ii) GSM provided mean dispersal distances in close agreement with the mean effective dispersal distance (estimated: 15.3 m and 20.8 m, Table 1, Figure 4; effective mean dispersal distance: 15.0 and 20.8 m, Figure S3), and (iii) CSM and SEMM provided unbiased estimates of the mean basic seed dispersal distance ($\delta_s = 10$ m) (Table 1, Figure S3). For the shape parameter $b_s = 0.5$, CSM and SEMM provided mostly unbiased estimates, whereas GSM provided overestimated values (0.86 for low DDM and 1.64 for high DDM, Table 1). Although reaching correctly the mean

effective dispersal distance, the exponential-power kernels fitted by GSM did not adjust correctly to the effective dispersal kernel, overestimating the frequencies of short and moderately large distances and underestimating the frequencies of intermediate distances (Fig. 4b,c). The shape of the kernels fitted by the CSM was close to the basic dispersal kernel for all levels of mortality, but more variable among simulations when mortality increases (Fig. 4d-f). Finally, for all parameters and all methods, the estimation variability (depicted by the confidence intervals in Table 1 and Figure S3) increased with the intensity of DDM.

The estimates of seed migration varied with DDM for all methods. Estimated seed migration rates (in CSM and SEMM) and seed immigrant densities (in GSM) largely increased with the intensity of DDM (from 0.96 to 1.88 for GSM; from 0.3 to 0.58 for CSM and SEMM, Tables 1 and 2). The estimated values matched closely with the effective migration rate after mortality reported above.

These qualitative results for GSM and CSM held for a variety of values for the adult density, the shape parameter b_s and the fecundity parameter ρ_s (Appendix S2). However, for large values of the dispersal distance δ_s , the mean dispersal distance estimated by the GSM does not match anymore with the mean effective dispersal distance, but still largely varies from NoM to low DDM to high DDM (Appendix S2).

Table 1. Performance of the IM, CSM, GSM and SEMM estimates of the dispersal and fecundity parameters based on 500 seedlings collected in 49 seedling traps. For each mortality scenarios (NoM, low DDM or high DDM, low EDM or high EDM), we estimated the dispersal/fecundity parameters for 100 simulations. We report the mean and quantiles at 2.5% and 97.5% over the successful estimations. Seed dispersal distance (δ_s in metres), shape of seed dispersal kernel (b_s unitless), migration rates (m_s , proportion) or immigrant density (Im_s , # per m^2) and selection gradient for DBH (ρ_s , unitless). The SEMM was applied with 10 loci and 10 alleles per locus

Method	Scenario	δ_s	b_s	m_s or Im_s^1	ρ_s	No. of successful estimations ²
	True value	10.0	0.50	0.30 or 0.97	2.0	
IM	NoM	10.3 (4.2, 40.4)	0.56 (0.16, 1.32)	0.92 (0.00, 1.87)	2.01 (1.01, 3.26)	80
	Low DDM	—	—	—	—	0
	High DDM	—	—	—	—	0
	Low EDM	106.8 (2.3, 178.8)	0.87 (0.03, 3.7)	1.00 (0.00, 2.36)	2.39 (0, 10.)	65
	High EDM	66.3 (0.2, 127.5)	0.73 (0.02, 2.4)	1.06 (0.00, 2.89)	4.91 (0., 22.)	41
GSM	NoM	10.0 (8.8, 11.1)	0.51 (0.42, 0.62)	0.96 (0.80, 1.13)	1.99 (1.52, 2.43)	98
	Low DDM	15.3 (12.8, 18.1)	0.86 (0.4, 1.29)	1.47 (1.22, 1.69)	2.04 (1.01, 3.12)	100
	High DDM	20.8 (18.0, 23.3)	1.64 (1.24, 2.26)	1.88 (1.64, 2.14)	1.95 (1.01, 3.43)	100
	Low EDM	10.1 (8.5, 12.0)	0.50 (0.36, 0.68)	0.98 (0.82, 1.18)	1.94 (1.26, 2.69)	100
	High EDM	10.1 (8.1, 12.0)	0.53 (0.34, 0.76)	0.98 (0.77, 1.24)	2.03 (1.14, 3.16)	100
CSM	NoM	10.0 (8.6, 11.4)	0.50 (0.39, 0.63)	0.30 (0.25, 0.35)	1.99 (1.44, 2.50)	99
	Low DDM	10.1 (7.9, 11.9)	0.52 (0.36, 0.69)	0.45 (0.37, 0.52)	2.00 (1.36, 2.64)	100
	High DDM	10.1 (5.52, 13.8)	0.53 (0.29, 0.80)	0.58 (0.50, 0.67)	2.03 (1.42, 2.78)	100
	Low EDM	10.1 (8.8, 11.5)	0.51 (0.39, 0.63)	0.30 (0.25, 0.36)	1.99 (1.40, 2.53)	100
	High EDM	9.9 (8.6, 11.1)	0.51 (0.39, 0.65)	0.30 (0.24, 0.38)	2.00 (1.47, 2.60)	100
SEMM	NoM	10.0 (8.7, 11.6)	0.50 (0.37, 0.62)	0.30 (0.25, 0.35)	2.01 (1.42, 2.61)	97
	Low DDM	10.2 (8.7, 11.8)	0.52 (0.38, 0.66)	0.45 (0.38, 0.52)	1.99 (1.40, 2.59)	100
	High DDM	10.4 (8.5, 13.8)	0.53 (0.40, 0.74)	0.58 (0.51, 0.67)	2.04 (1.42, 2.67)	100
	Low EDM	10.1 (8.7, 11.4)	0.51 (0.39, 0.66)	0.30 (0.25, 0.36)	1.99 (1.29, 2.54)	100
	High EDM	10.0 (8.7, 11.2)	0.51 (0.39, 0.67)	0.30 (0.25, 0.38)	1.99 (1.51, 2.64)	100

¹CSM and SEMM estimate a migration rate m_s , whereas IM and GSM estimate an immigrant density Im_s .

²The number of simulations where the maximum likelihood procedure converged with a sufficient precision.

Table 2. Performance of the SEMM estimates of the dispersal and fecundity parameters based on 500 seedlings randomly sampled in the 1-ha central plot after survival. For each mortality scenarios (NoM, low DDM, high DDM), we estimated the mating and dispersal parameters for 100 simulations. We report the mean, median and quantiles at 2.5% ($q_{2.5}$) and 97.5% ($q_{97.5}$) for average dispersal distance (δ_p , δ_s), shape of dispersal kernel (b_p , b_s), migration rates (m_p , m_s), selfing rate (s) and selection gradient for diameter (ρ_p , ρ_s). The SEMM was applied with 10 loci and 10 alleles in even frequencies per locus

Scenario	Param.	δ_s	b_s	m_s	ρ_s	δ_p	b_p	m_p	ρ_p
	True value	10.0	0.50	0.30	2.0	75.0	0.50	0.60	1.5
NoM	Mean	10.1	0.49	0.30	2.0	81.8	0.54	0.60	1.5
	Median	10.1	0.50	0.30	2.0	68.0	0.55	0.60	1.5
	($q_{2.5}, q_{97.5}$)	(8.9, 11.4)	(0.41, 0.59)	(0.27, 0.33)	(1.5, 2.5)	(51.5, 135.6)	(0.33, 0.81)	(0.56, 0.64)	(0.9, 2.1)
Low DDM	Mean	10.2	0.50	0.46	2.1	$>10^5$	0.53	0.60	1.6
	Median	10.2	0.50	0.45	2.1	74.5	0.52	0.60	1.6
	($q_{2.5}, q_{97.5}$)	(8.4, 11.8)	(0.38, 0.64)	(0.40, 0.54)	(1.4, 2.7)	(47.4, 332.7)	(0.20, 0.91)	(0.55, 0.65)	(0.8, 2.4)
High DDM	Mean	10.1	0.56	0.58	2.1	$>10^5$	0.58	0.60	1.5
	Median	9.9	0.52	0.58	2.1	68.1	0.52	0.60	1.5
	($q_{2.5}, q_{97.5}$)	(7.1, 14.5)	(0.36, 0.92)	(0.50, 0.65)	(1.4, 2.9)	(42.9, 227.9)	(0.23, 1.15)	(0.53, 0.67)	(0.3, 2.7)

SENSITIVITY OF IM, GSM, CSM AND SEMM TO ENVIRONMENT-DEPENDENT MORTALITY

In the presence of EDM, the effective and basic dispersal kernels remained similar and the effective migration rate remained constant. As expected, the dispersal parameters were then estimated with no bias for the GSM, CSM and SEMM. The estimation variability remained sensibly constant in NoM, low EDM and high EDM scenarios. In contrast, the IM method was affected by the presence of EDM, providing estimates with much larger bias and variability, and estimations failed in a majority of cases (Table 1).

Seed migration rate was correctly estimated at 0.30 on average with CSM and SEMM. Seed immigrant density was correctly estimated with GSM, but slightly overestimated with IM, especially for high EDM (Table 1).

ESTIMATION OF OTHER PARAMETERS

Except IM in the presence of post-dispersal mortality, all methods estimated well the effect of DBH on female fecundity ($\rho_s = 2.$) with no bias and a variability that only moderately increased with post-dispersal mortality intensity (Tables 1 and 2).

Since the fecundity parameter q_s varied much among scenarios and among simulations of a given scenario, we did not interpret the values estimated for this parameter by the IM and GSM methods.

The SEMM can also estimate the parameters related to pollen (pollen migration rate, m_p , selection gradients ρ_p and pollen dispersal parameters δ_p and b_p). m_p and ρ_p were estimated with low bias and high precision even in the presence of DDM (Table 2) and EDM (results not shown). In contrast, the pollen dispersal distance (δ_p) was estimated with a bias that increased for scenarios with low DDM and high DDM (mean $\delta_p > 10^5$, Table 2), but mostly due to rare extremely large estimated values (median δ_p and 95% confidence intervals remained close to the true value of 75.0, Table 2). The quality of the estimated pollen dispersal parameters decreased when the number of seedlings having both local male and female parents decreased.

This is the case when seed migration rate increased (m_s was up to 46% and 56% under low DDM and high DDM): the pollen dispersal kernel became unstable, some simulations producing unrealistically high δ_p values.

Discussion

It could seem intuitive that estimates based on seedling data provide effective dispersal kernels, whereas estimates based on seed data provide basic dispersal kernels. We demonstrated here that dispersal kernels estimated from the CSM or SEMM methods applied to established seedlings are not sensitive to post-dispersal DDM occurring during recruitment. This means that CSM and SEMM estimate the basic seed dispersal kernel, encompassing only the movement of the dispersed seed, and not the effective dispersal kernel, encompassing additionally the 'post-arrival' survival of seeds (Nathan *et al.* 2012). This property stands even when post-dispersal mortality strongly modifies the spatial distribution of surviving seedling (e.g. Figure 4), that is, in cases where the effective dispersal kernel is very different from the basic dispersal kernel (in our simulations, the effective mean dispersal distance was more than twice the basic mean dispersal distance).

Theoretical arguments developed here highlight that this result stands because CSM and SEMM rely on the information carried by the proportions from different sources at each sampling location. These proportions are insensitive to most forms of location-dependent mortality, as long as mortality is independent of genotype/source of origin (eqn 6 and Fig. 1). This means that the proportions of different sources preserve the information about the basic dispersal kernel, even after several episodes of mortality. At the opposite, the density of established seedlings at different locations can be profoundly modified by DDM (as already illustrated in Swamy *et al.* 2011; Tautenhahn *et al.* 2012) and informs on the effective dispersal kernel (see eqn 5). Accordingly, the statistical methods relying on seedling counts (GSM) provided dispersal kernels with a mean distance comparable with the effective mean dispersal distance, when the mean dispersal distances (δ_s) are not too large compared to the dimension of the experimental plot. For

large δ_s , the estimations from GSM no more matched with the mean effective dispersal distance (Appendix S2). This is likely due to the following points: (i) a consequential amount of long-distance events were missing due to the sampling design, and (ii) adjusting an exponential-power function to the effective dispersal kernel, which does not belong to the exponential-power kernel family, results in a compromise (e.g. Fig. 4). The compromise can be largely affected by the spatial scale of the observations, as demonstrated by Kuparinen *et al.* (2007). Fitting a wrong model to the basic dispersal kernel using CSM would probably also result in biased estimates for large δ_s .

We have chosen to model post-dispersal mortality as driven by the distance to all neighbouring adults, which resulted in the well-known Janzen–Connell (JC) pattern, with an increase in mean effective dispersal distance (Fig. 4). We could have obtained comparable JC patterns with DDM driven by the density of seedlings in a neighbourhood, or we could have simulated other patterns of recruitment driven by DDM (McCanby 1985). Although we simulated a single type of DDM, our theoretical argument ensures that CSM and SEMM estimates remain insensitive to DDM in any recruitment pattern where mortality does not depend on seedling genotype or source of origin. At population scale, this does not assume that mortality affects similarly all genotypes/seed sources: for instance, progenies from mother trees in dense areas will indeed suffer more from DDM. But our result stands if at any given location, mortality affects equally the seeds received here from all mother trees (eqn 6). Said differently, mortality must be independent of the source of origin conditionally on seedling location. This assumption would be invalid if, for example, (i) the susceptibility of seedlings to DDM is partially determined genetically (Coleman, Gill & Rebetzke 2001), or (ii) competition generated by an adult is different against its own seedlings as against seedlings from other trees (File, Murphy & Dudley 2012).

The insensitivity to DDM of CSM and SEMM is a useful property complementing the insensitivity to the sampling design (Robledo-Arnuncio & García 2007) and the lower sensitivity to poor fecundity models (Jones & Muller-Landau 2008). Once acknowledged that methods based on proportions (CSM and SEMM) estimate the basic seed dispersal kernel, a way to get consistent estimates between proportion-based and density-based estimates (GSM and IM) is to specify the mortality process in GSM and IM (e.g. Kunstler *et al.* 2007; Amm *et al.* 2012; Tautenhahn *et al.* 2012). Oppositely, we can see no way to estimate the effective dispersal kernel when considering only information about proportions.

While CSM and SEMM estimates of the seed dispersal kernel are unaffected by DDM, the estimates of seed migration rates (m_s) vary much with increasing DDM. This was not a statistical bias, but reflected the increase in effective seed migration rates in populations undergoing DDM (from 30% to 58% in our high DDM simulations). It happened because local seedlings occurred often close to an adult tree (their mother), whereas migrants were assumed uniformly distributed. Therefore, local seedlings have a higher average mortality rate than migrants in the DDM scenarios. Such effect could be lower in

natural populations for species where animal-assisted dispersal leads to the clumping of migrant seedlings (e.g. fleshy fruit seeds beneath fruiting or roost trees, Jordano *et al.* 2007) or to the clumping of migrant and local seeds. Thus, although an effective migration rate increasing along life stages can reveal the existence of DDM, interpretation should be cautious when expecting aggregated long-distance dispersal events.

Another classical distinction exists between primary and secondary dispersal (Wang & Smith 2002). With our definition, the basic dispersal kernel cumulates primary and secondary dispersal. Our results thus show that established seedlings analysed with CSM or SEMM provide the basic kernel after secondary dispersal. On the contrary, seeds collected in nets or pots (e.g. Bullock, Shea & Skarpaas 2006) result only from primary dispersal. IM, GSM, CSM and SEMM should all provide the primary dispersal kernel when applied to seeds from physical traps. Finally, GSM applied to established seedling provides the effective dispersal kernel. Thus, experiments that gather seed counts in physical traps (whether genotyped or not), seedling counts and seedling genotypes could estimate three different functions: primary, primary + secondary and effective kernels (e.g. Millerón *et al.* 2013).

Although not devoted to this goal, our simulations offered the opportunity to compare inverse modelling, based only on seedling counts, with methods that exploit molecular markers (CSM, GSM, SEMM). In the absence of DDM, our results confirmed Jones & Muller-Landau (2008): the use of genetic data reduces the bias, the estimation variability and provides more frequent successful estimations (see SSS, GSM and CSM results in Jones & Muller-Landau 2008). Moreover, in the presence of DDM, IM performed even worse since we could not reach a single successful fit for any of 200 simulated data sets. This is because the exponential power model used here was not flexible enough to fit the hump-shaped effective dispersal kernel. Using a log-normal model instead provided more successful fits, but large biases still remained (Figure S4). Although the exponential power model was similarly false for GSM (Figure S4), it estimated realistic values for the mean effective dispersal distance. Choosing a sufficient panel of dispersal kernel families is a well-known challenge of all estimation methods (Bullock, Shea & Skarpaas 2006). Here, methods using the genetic information were more robust to the choice of the kernel family, though the robustness of GSM decreased with inappropriate sampling design (Appendix S2). But the performance of CSM and SEMM would probably be worse in more difficult conditions (e.g. fitting a wrong kernel family on the basic dispersal kernel).

Life stage studies take advantage of differences in patterns of dispersal among successive cohorts to disentangle the roles of seed dispersal and post-dispersal recruitment processes (e.g. Augspurger 1983; González-Martínez *et al.* 2006; Swamy *et al.* 2011; Bontemps, Klein & Oddou-Muratorio 2013). Because EDM and DDM are cumulative over a cohort lifetime, spatial patterns of recruited individuals should become increasingly discordant over time with the spatial patterns of dispersed seeds. Our results question the possibility to conduct life stage studies with CSM or SEMM analyses. These methods

estimate the same basic seed dispersal kernels for cohorts of all ages. Putative differences among kernels obtained from different cohorts should thus result from other environmental/biotic processes (e.g. canopy density closure or annual variations of meteorological conditions). This suggests a new two-step procedure where (i) CSM and SEMM are first used to ensure that the different cohorts result from the same basic dispersal kernel and in that case (ii) GSM and mother–offspring distance distributions retrieved from a parentage analysis (when there is no traps) are used to compare the effective kernels of the different cohorts to conclude about the significance of post-dispersal mortality (Bontemps, Klein & Oddou-Muratorio 2013).

Acknowledgements

We are grateful to Tonya Lander, Arndt Hampe, Annabelle Amm, Cristina García and Gema Escribano-Avila for fruitful discussions and comments on this work. We also thank several reviewers for constructive comments on the earlier versions of the manuscript. This work was supported by grants from the Agence Nationale de la Recherche (ColonSGS: ANR-07-JCJC-0117; EMILE: ANR-09-BLAN-0145-01), the EC-supported Network of Excellence Evoltree (GOCE-016322, BEECH initiative) and the ERA-Net BiodivERsA LINKTREE project (ANR-08-Biodiversa-006-06), part of the 2009 BiodivERsA call for research proposals.

References

- Amm, A., Pichot, C., Dreyfus, P., Davi, H. & Fady, B. (2012) Improving the estimation of landscape scale seed dispersal by integrating seedling recruitment. *Annals of Forest Science*, **69**, 845–856.
- Augsburger, C. (1983) Recruitment around tropical trees: changes in cohort distance with time. *Oikos*, **40**, 189–196.
- Berger, U., Piou, C., Schiffrers, K. & Grimm, V. (2008) Competition among plants: Concepts, individual-based modelling approaches, and a proposal for a future research strategy. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 121–135.
- Bontemps, A., Klein, E.K. & Oddou-Muratorio, S. (2013) Shift of spatial patterns during early recruitment in *Fagus sylvatica*: evidence from seed dispersal estimates based on genotypic data. *Forest Ecology and Management*, **305**, 67–76.
- Bullock, J.M., Shea, K. & Skarpaas, O. (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology*, **186**, 217–234.
- Burczyk, J., Adams, W.T., Birkes, D.S. & Chybicki, I.J. (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics*, **173**, 363–372.
- Clark, J., Beckage, B., Camill, P., Cleveland, B., HilleRisLambers, J., Lichter, J., McLachlan, J.S., Mohan, J. & Wyckoff, P. (1999a) Interpreting recruitment limitation in forests. *American Journal of Botany*, **86**, 1–16.
- Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. (1999b) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475–1494.
- Coleman, R.D., Gill, G.S. & Rebetzke, G.J. (2001) Identification of quantitative trait loci for traits conferring weed competitiveness in wheat (*Triticum aestivum* L.). *Australian Journal of Agricultural Research*, **52**, 1235–1246.
- Connell, J.H. (1971) On the role of natural enemies in preventing competitive exclusion in some marine animals and rain forests trees. *Dynamics of Populations* (eds P.J. Den Boer & G. Gradwell), pp. 298–312. Centre for Agricultural Publishing and Documentation (PUDOC), Wageningen, The Netherlands.
- Cousens, R., Dytham, C. & Law, R. (2008) *Dispersal in Plants*. Oxford University Press, Oxford, A population perspective.
- File, A.L., Murphy, G.P. & Dudley, S.A. (2012) Fitness consequences of plants growing with siblings: reconciling kin selection, niche partitioning and competitive ability. *Proceedings of the Royal Society Series B*, **279**, 209–218.
- González-Martínez, S.C., Burczyk, J., Nathan, R., Nanos, N., Gil, L. & Alfá, R. (2006) Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology*, **15**, 4577–4588.
- HilleRisLambers, J., Clark, J.S. & Beckage, B. (2002) Density-dependent mortality and the latitudinal gradient in species diversity. *Nature*, **417**, 732–735.
- Hubbell, S.P. (1980) Seed predation and the coexistence of tree species in tropical forests. *Oikos*, **35**, 214–229.
- Janzen, D.H. (1970) Herbivores and number of tree species in tropical forests. *American Naturalist*, **104**, 501–528.
- Jones, F.A. & Muller-Landau, H.C. (2008) Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. *Journal of Ecology*, **96**, 642–652.
- Jones, F.A., Chen, J., Weng, G.-J. & Hubbell, S.P. (2005) A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). *American Naturalist*, **166**, 543–555.
- Jordano, P., García, C., Godoy, J.A. & García-Castaño, J.L. (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3278–3282.
- Klein, E.K. & Oddou-Muratorio, S. (2011) Pollen and seed dispersal inferred from seedling genotypes: the Bayesian revolution has passed here too. *Molecular Ecology*, **20**, 1077–1079.
- Kunstler, G., Chadoeuf, J., Klein, E.K., Curt, T., Bouchaud, M. & Lepart, J. (2007) Tree colonization of sub-Mediterranean grasslands: effects of dispersal limitation and shrub facilitation. *Canadian Journal of Forest Research*, **37**, 103–115.
- Kuparinen, A., Snäll, T., Vänskä, S., O'Hara, B. & R., (2007) The role of model selection in describing stochastic ecological processes. *Oikos*, **116**, 966–974.
- LePage, P.T., Canham, C.D., Coates, K.D. & Bartemucci, P. (2000) Seed abundance versus substrate limitation of seedling recruitment in northern temperate forests of British Columbia. *Canadian Journal of Forest Research*, **427**, 415–427.
- Martin, P.H. & Canham, C.D. (2010) Dispersal and recruitment limitation in native versus exotic tree species: life-history strategies and Janzen-Connell effects. *Oikos*, **119**, 807–824.
- McCanny, S. (1985) Alternatives in parent-offspring relationships in plants. *Oikos*, **45**, 148–149.
- Millerón, M., López de Heredia, U., Lorenzo, Z., Alonso, J., Dounavi, A., Gil, L. & Nanos, N. (2013) Assessment of spatial discordance of primary and effective seed dispersal of European beech (*Fagus sylvatica* L.) by ecological and genetic methods. *Molecular Ecology*, **22**, 1531–1545.
- Moran, E.V. & Clark, J.S. (2011) Estimating seed and pollen movement in a monoeicous plant: a hierarchical Bayesian approach integrating genetic and ecological data. *Molecular Ecology*, **20**, 1248–1262.
- Nathan, R. & Casagrandi, R. (2004) A simple mechanistic model of seed dispersal, predation and plant establishment: Janzen-Connell and beyond. *Journal of Ecology*, **92**, 733–746.
- Nathan, R. & Muller-Landau, H. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution*, **15**, 278–285.
- Nathan, R., Klein, E., Robledo-Arnuncio, J.J. & Revilla, E. (2012) Dispersal kernels: review. *Dispersal and Spatial Evolutionary Ecology* (eds J. Clobert, M. Baguette, T. Benton & J.M. Bullock), pp. 187–210. Oxford University Press, Oxford, UK.
- Oddou-Muratorio, S. & Klein, E.K. (2008) Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Molecular Ecology*, **17**, 2743–2754.
- Ribbens, E., Silander, J.A. & Pacala, S. (1994) Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. *Ecology*, **75**, 1794–1806.
- Robledo-Arnuncio, J.J. & García, C. (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, **16**, 5098–5109.
- Schupp, E.W. & Fuentes, M. (1995) Spatial patterns of seed dispersal and the unification of plant population ecology. *Ecoscience*, **2**, 267–275.
- Schupp, E.W. & Jordano, P. (2011) The full path of Janzen-Connell effects: genetic tracking of seeds to adult plant recruitment. *Molecular Ecology*, **20**, 3953–3955.
- Steinitz, O., Troupin, D., Vendramin, G.G. & Nathan, R. (2011) Genetic evidence for a Janzen-Connell recruitment pattern in reproductive offspring of *Pinus halepensis* trees. *Molecular Ecology*, **20**, 4152–4164.
- Stoyan, D. & Wagner, S. (2001) Estimating the fruit dispersion of anemochorous forest trees. *Ecological Modelling*, **145**, 35–47.
- Swamy, V., Terborgh, J., Dexter, K.G., Best, B.D., Alvarez, P. & Cornejo, F. (2011) Are all seeds equal? Spatially explicit comparisons of seed fall and sapling recruitment in a tropical forest. *Ecology letters*, **14**, 195–201.

- Tautenhahn, S., Heilmeier, H., Jung, M., Kahl, A., Kattge, J., Moffat, A. & Wirth, C. (2012) Beyond distance-invariant survival in inverse recruitment modeling: A case study in Siberian *Pinus sylvestris* forests. *Ecological Modelling*, **233**, 90–103.
- Wang, B.C. & Smith, T.B. (2002) Closing the seed dispersal loop. *Trends in Ecology & Evolution*, **17**, 379–386.

Received 22 March 2013; accepted 20 August 2013
Handling Editor: Justin Travis

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1 Technical aspects concerning the simulation of datasets.

Appendix S2 Performance of the Spatially Explicit Mating Model (SEMM) approaches for a wider range of population parameters.

Appendix S3 Performance of the Spatially Explicit Mating Model (SEMM) approach for uneven allelic frequencies.

Figure S1. Effective dispersal kernels under different regimes of density-dependent mortality (DDM).

Figure S2. Effective dispersal kernels under different regimes of Environmentally-Dependent Mortality.

Figure S3. Performance of four estimation methods of the mean dispersal distance based on seedlings collected in 49 seedling traps.

Figure S4. Performance of four estimation methods of the mean dispersal distance based on seedlings collected in 49 seedling traps.

Figure S5. Performance of the estimates of (A) the mean seed dispersal distance δ_s , (B) the shape parameter b_s and (C) the migration rate m_s for a random sampling of seedlings within the 1ha central plot.

Figure S6. Performance of the estimates of (A) the mean seed dispersal distance δ_s , (B) the shape parameter b_s and (C) the migration rate m_s for 500 seedlings sampled within 49 traps, for various molecular markers sets and for three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S1. Mean effective dispersal distance (MEDD), average mortality rate (AMR) and effective migration rate (EMR) for various sets of parameters.

Table S2. Mean dispersal distance, δ_s , estimated with GSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S3. Mean dispersal distance, δ_s , estimated with CSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM = low/high density-dependence mortality).

Table S4. Shape parameter, b_s , estimated with GSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S5. Shape parameter, b_s , estimated with CSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S6. Seed immigration density, Im_s , estimated with GSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S7. Seed migration rate, m_s , estimated with CSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S8. Effect of DBH on female fecundity, ρ_s , estimated with GSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).

Table S9. Effect of DBH on female fecundity, ρ_s , estimated with CSM for various sets of population parameters and three levels of density-dependent mortality (NoM, no Mortality; LowDDM/HighDDM, low/high density-dependence mortality).