

Two-Generation Analysis of Pollen Flow Across a Landscape. IV. Estimating the Dispersal Parameter

Frédéric Austerlitz^{*,1} and Peter E. Smouse[†]

^{*}Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA, Pierroton, F-33611 Gazinet Cedex, France and [†]Department of Ecology, Evolution and Natural Resources, Cook College, Rutgers University, New Brunswick, New Jersey 08901-8551

Manuscript received November 13, 2001

Accepted for publication February 20, 2002

ABSTRACT

The distance of pollen movement is an important determinant of the neighborhood area of plant populations. In earlier studies, we designed a method for estimating the distance of pollen dispersal, on the basis of the analysis of the differentiation among the pollen clouds of a sample of females, spaced across the landscape. The method was based solely on an estimate of the global level of differentiation among the pollen clouds of the total array of sampled females. Here, we develop novel estimators, on the basis of the divergence of pollen clouds for all pairs of females, assuming that an independent estimate of adult population density is available. A simulation study shows that the estimators are all slightly biased, but that most have enough precision to be useful, at least with adequate sample sizes. We show that one of the novel pairwise methods provides estimates that are slightly better than the best global estimate, especially when the markers used have low exclusion probability. The new method can also be generalized to the case where there is no prior information on the density of reproductive adults. In that case, we can jointly estimate the density itself and the pollen dispersal distance, given sufficient sample sizes. The bias of this last estimator is larger and the precision is lower than for those estimates based on independent estimates of density, but the estimate is of some interest, because a meaningful independent estimate of the density of reproducing individuals is difficult to obtain in most cases.

BOTH evolutionary and conservation biologists are interested in the distance of pollen movement because of its role in the establishment of neighborhoods and in connectivity among populations (for a review, see SORK *et al.* 1999). We have elsewhere proposed that one should use measurable differentiation among the inferred pollen clouds of widely spaced females as an assay of the distribution of pollination distances across existing landscapes (SMOUSE *et al.* 2001), using a model we refer to as TwoGener. We propose this model as an alternative to both the current practice of extracting an estimate indirectly from F_{ST} (assuming evolutionary equilibrium) and more direct (but laborious) estimates from parentage analysis, both of which have limitations (SORK *et al.* 1999). We have modeled the relationship between the intraclass correlation of maternal pollen pools, Φ_{fi} , as a function of average pollen dispersal distance (δ), the spatial density (d) of reproductive adults, and the average distance between females (\bar{z}) (AUSTERLITZ and SMOUSE 2001a). We have also (AUSTERLITZ and SMOUSE 2001b) examined the impact of spatially organized genetic structure among the adults on Φ_{fi} . These two efforts have yielded the possibility of ex-

tracting an estimate of the average pollen dispersal distance on the basis of an estimate of Φ_{fi} that is derived from the inferred pollen pool divergence among the complete pairwise array of sampled females.

The TwoGener approach shares a great deal in common with the analysis of adult genetic diversity among populations. In fact, Φ_{fi} is directly analogous to WRIGHT's (1951) F_{st} . With the island model, it is sufficient to obtain a global estimate of F_{st} , from which we can extract an estimate of $N_e m$, assuming evolutionary equilibrium. However, for an "isolation by distance" model (WRIGHT 1943, 1946; MALÉCOT 1948, pp. 54–63; KIMURA and WEISS 1964), the differentiation observed between two populations, gauged by their pairwise F_{st} estimate, is expected to be an increasing function of the distance between them (SLATKIN 1991, 1993). Thus, it is possible to obtain an estimate of both the average and variance of long-term (evolutionary time) dispersal rate, as a function of the distance between pairs of populations (ROUSSET 1997) or individuals (ROUSSET 2000).

Here, we develop a similar estimation procedure for the estimation of contemporaneous pollen flow distance. As we have shown in previous work, there is a direct relation between Φ_{fi} and the average physical distance between pairs of females (\bar{z}), and we can use the same theory to relate a Φ_{fi} -estimate for any particular pair of females to the physical distance between them. In this article, we present several estimators that use

¹Corresponding author: Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA-Domaine de l'Hermitage, Pierroton B. P. 45, F-33611 Gazinet Cedex, France.
E-mail: austerli@pierroton.inra.fr

either the global value of Φ_{ft} , estimated over all sampled females in the population, or the pairwise Φ_{ft} -estimates to gauge the mean pollen dispersal distance.

We first develop estimates that assume that the adult density is independently known in the population. Then, we develop an additional method that allows joint estimation of the density of reproducing adults and the dispersal distance. One might argue that since density can be measured independently in the field, there is no necessity to estimate density from the genetic data, but the real issue is whether we can reliably estimate the true pollination density by direct field observation.

Using computer simulation, we first perform a comparative study on those estimators that assume that adult density is independently known, which is designed to answer several questions. (i) What are the best estimators in terms of bias, standard deviation, and mean squared error; (ii) what is the best strategy for the allocation of experimental effort, in terms of the numbers of mothers and progeny per mother; and (iii) what is the sensitivity of the estimates to the exclusion probability of the set of markers used? We then explore joint estimation, using the same simulation approach. The main question is whether (and how much) we gain by inferring density from the genetic data.

METHODS

The estimators: *Global estimators:* Assume that we have a sample of females from a given population and a sample of offspring from each female, so that we can infer the pollen cloud sampled by each female, using methods provided in SMOUSE *et al.* (2001). Our estimators assume that the pollen dispersal distribution is a bivariate (isotropic) normal distribution, that is, the only distribution for which one can obtain analytic solutions (AUSTERLITZ and SMOUSE 2001a). The model links the estimated differentiation among pollen clouds (Φ_{ft}), the average pairwise distance between females, and the dispersal parameter (σ). This bivariate normal distribution is defined as

$$p(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{z^2}{2\sigma^2}\right\}, \tag{1}$$

where $z^2 = (x^2 + y^2)$ is the squared physical distance from the index female, located at the origin, to the pollinating male, in any direction. The isotropic model is radially symmetric.

The first estimator that can be designed uses the approximate relation between the global Φ_{ft} and σ ,

$$\Phi_{ft} = \frac{1}{4\pi\sigma^2 d} \tag{2}$$

(AUSTERLITZ and SMOUSE 2001a), where d is the real density of reproductive adults in the population. The genetic data used to estimate Φ_{ft} are made up of discrete

multinomial variables, converted to pairwise Euclidean distance measures among pairs of gametes. Those squared genetic distances are generally χ^2 -like, but we have no strong basis for precise distributional assumptions about Φ_{ft} . The estimates constructed here will necessarily be a bit heuristic in their motivation. The relationship in (2) is quite adequate, provided that the average distance between the sampled mothers is $>5\delta$ (AUSTERLITZ and SMOUSE 2001a), where δ is the average distance of pollen dispersal, given by

$$\delta = \sigma \sqrt{\frac{\pi}{2}}. \tag{3}$$

Equation 2 yields a first (very simple) global estimator for σ :

$$\hat{\sigma}_{g1} = \sqrt{\frac{1}{4\pi\Phi_{ft}d}}. \tag{4}$$

However, if the average distance between mothers is $<5\delta$, the relationship between global Φ_{ft} and σ requires refinement (AUSTERLITZ and SMOUSE 2001a):

$$\Phi_{ft} = \frac{1 - \exp\{-z^2/4\sigma^2\}}{8\pi\sigma^2 d - \exp\{-z^2/4\sigma^2\}}. \tag{5}$$

Equation 5 can be transformed into

$$\sigma^2 = \frac{1}{4\pi d} \left[\frac{1 - (1 - \Phi_{ft})\exp\{-z^2/4\sigma^2\}}{2\Phi_{ft}} \right], \tag{6}$$

which yields a solution for σ by numerical iteration. We insert an initial value for σ in the right-hand side (typically $\hat{\sigma}_{g1}$) of (6) and obtain an updated value for σ on the left-hand side of (6), which we insert into the right-hand side of the equation. This procedure is repeated until convergence to a stable value for σ , which is our second estimator, $\hat{\sigma}_{g2}$.

Pairwise estimators: Equation 6 can also be used for $\Phi_{ft}(z)$ between any pair of females, as a function of the distance (z) between them. Denote the pairwise Φ_{ft} observed between the i th and j th females by ϕ_{ij}^{obs} , and denote the physical distance between these females as z_{ij} . Assume that we have sampled n_m mother plants. A simple average of (6) over the $n_m(n_m - 1)/2$ pairs of females yields our first pairwise estimator, denoted $\hat{\sigma}_{p1}$, computed as the solution to the equation

$$\hat{\sigma}_{p1}^2 = \frac{1}{2\pi d n_m (n_m - 1)} \sum_{i < j}^{n_m} \left[\frac{1 - (1 - \phi_{ij}^{obs}) \exp(-z_{ij}^2/4\hat{\sigma}_{p1}^2)}{2\phi_{ij}^{obs}} \right], \tag{7}$$

a solution for which can easily be found, using the same sort of iterative procedure we described above.

Now, σ^2 appears both on the left-hand side of this equation and in the denominator of the exponent on the right-hand side, and it is not entirely obvious whether it would be better to estimate σ or $1/\sigma$, so we

have also evaluated a second pairwise estimator, defining $\beta = 1/\sigma$, and then estimating $\hat{\beta}$ as the solution of

$$\hat{\beta}^2 = \frac{8\pi d}{n_m(n_m - 1)} \sum_{i < j}^{n_m} \left[\frac{2\phi_{ij}^{\text{obs}}}{1 - (1 - \phi_{ij}^{\text{obs}})\exp(-z_{ij}^2\hat{\beta}^2/4)} \right], \quad (8)$$

which yields

$$\hat{\sigma}_{p2} = \hat{\beta}^{-1}. \quad (9)$$

Still another method consists of performing a nonlinear regression to estimate the best fit for σ , using a least-squares criterion. Given any particular estimate of σ , we can predict each of the theoretically expected values $\phi_{ij}^{\text{th}} = \Phi_{\text{ft}}(z_{ij})$ values from the equation

$$\phi_{ij}^{\text{th}} = \frac{1 - \exp(-z_{ij}^2/4\hat{\sigma}^2)}{8\pi\hat{\sigma}^2 d - \exp(-z_{ij}^2/4\hat{\sigma}^2)}. \quad (10)$$

We construct the quadratic criterion $Q(\sigma)$, defined as

$$Q(\sigma) = \sum_{i < j}^{n_m} (\phi_{ij}^{\text{obs}} - \phi_{ij}^{\text{th}})^2, \quad (11)$$

which is minimized for the choice of σ in (10), providing our third pairwise estimator, $\hat{\sigma}_{p3}$.

Although the true value of Φ_{ft} is nonnegative, ϕ_{ij}^{obs} is an estimate that can be negative. For $\hat{\sigma}_{p2}$ and $\hat{\sigma}_{p3}$, we set all the negative estimates to zero. This cannot be done for $\hat{\sigma}_{p1}$ because the ϕ_{ij}^{obs} 's appear in the denominator of a fraction, so negative values have been removed from the summation for $\hat{\sigma}_{p1}$. This is likely to create an upward bias for this estimate because it ignores some of the pairs of mothers with the lowest differentiation in their pollen cloud.

For all the methods above, we have to impose an external estimate of d that is obtained from field measures of the density of reproducing individuals. We can also obtain such an estimate for d from the genetic data themselves, by optimizing (11) simultaneously for σ and d , providing a final set of estimates, $\hat{\sigma}_{d1}$ and \hat{d}_1 . All nonlinear regression estimates ($\hat{\sigma}_{p3}$, $\hat{\sigma}_{d1}$, and \hat{d}_1) are obtained via the Levenberg-Marquardt method (LEVENBERG 1944; MARQUARDT 1963), which is implemented in Mathematica 4.0 (WOLFRAM 1999), and for which a C source code is given in PRESS *et al.* (1988). Both a C program and a DOS-executable version that performs all the calculations described above are available from F. Austerlitz, on request.

Simulations: We assessed bias, standard deviation, and square root of mean-squared error ($\sqrt{\text{MSE}}$) for each of the estimators above, by repeated simulation. We simulated a population according to the method described in SMOUSE *et al.* (2001), constructing a population of 10,000 adult, monoecious, self-fertile individuals, distributed randomly across the landscape, using a bivariate uniform distribution, a square of size 100 units \times 100 units, folded into a torus; to avoid border effects, left (resp. top) and right (resp. bottom) extremities

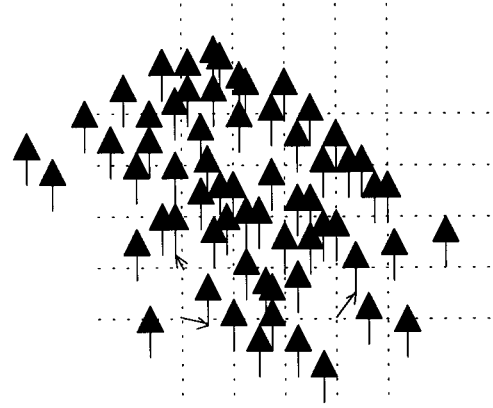


FIGURE 1.—Schematic representation of the sampling of the mothers. As indicated by the arrows, we sample the closest mother to each gridpoint.

were connected. By construction, density was $d = 1.0$. All individuals were assigned a genotype of n_L independently segregating loci. Each locus had n_A equiproportional alleles. We can compute their multiple-locus exclusion probability (E) as a measure of the information content available from the genetic battery (CHAKRAVARTI and LI 1983; JAMIESON 1994). For a single locus with equiproportional alleles, this probability is

$$E(\text{single locus}) = \frac{(n_A - 1)(n_A^3 - n_A^2 - 2n_A + 3)}{n_A^4} \quad (12)$$

(JAMIESON and TAYLOR 1997) and thus for n_L independent loci

$$E = 1 - \left(1 - \frac{(n_A - 1)(n_A^3 - n_A^2 - 2n_A + 3)}{n_A^4} \right)^{n_L}. \quad (13)$$

The genotype of each individual was created by randomly drawing its two alleles at each locus from the population allele frequency array.

We drew the sample of n_m mother plants on a square grid with spacing 1.0 between consecutive points. Thus the grid was approximately of size $\text{Int}(\sqrt{n_m}) \times \text{Int}(\sqrt{n_m})$ plus the remaining individuals on a line below, where $\text{Int}(x)$ denotes the integer part of a real number x . In fact, the sampled mother was the closest individual to each grid point (Figure 1). We also drew, at random on the whole landscape, a complementary subset of n_s individuals, which we used, along with the mothers, to provide an estimate of allelic frequencies in the population (SMOUSE *et al.* 2001). From each of the n_m mothers, we drew n_o offspring. The coordinates of the father for each offspring were drawn at random from the pollen dispersal distribution, and we then assigned the adult nearest those coordinates as the father. The offspring's genotype was constructed by randomly drawing one allele for each locus from the maternal genotype and one

allele from the paternal genotype. Using the maternal and offspring genotypes, we then estimated the ϕ values and the various estimates of σ (and also d_1).

We used the bivariate normal pollen dispersal distribution for simulations, and we tested the impact of various design criteria, conducting 1000 simulation runs for each parameter set and computing the mean, standard deviation, and square root of the mean-squared error for all estimators. Our reference setting included density $d = 1$, axial dispersal variance $\sigma = 1$, number of sampled mothers $n_m = 20$, number of other adults sampled to estimate allelic frequencies $n_s = 30$, number of offspring per mother $n_o = 20$. For the genetic markers, we chose a setting that would correspond to microsatellite markers: number of loci studied $n_L = 5$, number of alleles per locus $n_A = 10$. We then varied each parameter separately, to assess its impact on the various estimators.

RESULTS

Best estimators: A plot of the pairwise Φ_{it} 's against the distance between the sampled mothers shows clearly that these pairwise values are dispersed around the theoretical curve as expected and that this dispersion decreases when more offspring per mother are sampled (Figure 2). Table 1 shows two of the estimators that are strongly biased, the global estimator ($\hat{\sigma}_{g1}$), which does not take into account the average distance between mothers, and ($\hat{\sigma}_{p1}$), the pairwise estimate based on (7). Recall that for the latter, we were forced to ignore negative estimates, which could only bias the estimate upward. Among the methods that treat density as external information, these two estimates also show the greatest variance. The combination of large bias and large variance yields a very large mean-squared error. We henceforth dispense with $\hat{\sigma}_{g1}$ and $\hat{\sigma}_{p1}$, since their behavior is always inferior.

Number of mothers vs. number of progeny: The better global estimator ($\hat{\sigma}_{g2}$), which allows for the average distance between females, and the best pairwise estimators ($\hat{\sigma}_{p2}$ and $\hat{\sigma}_{p3}$), which allow for the variation among those pairwise distances, provide very similar results. They all show very small bias, which is positive for $\hat{\sigma}_{g2}$ and negative for $\hat{\sigma}_{p2}$ and $\hat{\sigma}_{p3}$. The variance for $\hat{\sigma}_{p2}$ is less than that for $\hat{\sigma}_{p3}$, but the bias is a bit larger; on balance, $\hat{\sigma}_{p3}$ has a lower MSE than $\hat{\sigma}_{p2}$, which is itself better than $\hat{\sigma}_{g2}$, at least for low sampling effort. This remains true when the number of mothers is increased, but when the number of offspring is increased, $\hat{\sigma}_{p2}$ becomes slightly better.

Increasing the number of sampled mothers (n_m) has a stronger impact on the estimators than increasing the number of offspring per mother, yielding a greater increase of precision and a greater reduction of bias (Table 2). Since the product $N = n_m n_o$ is the size of the experimental effort invested in the study, it is generally

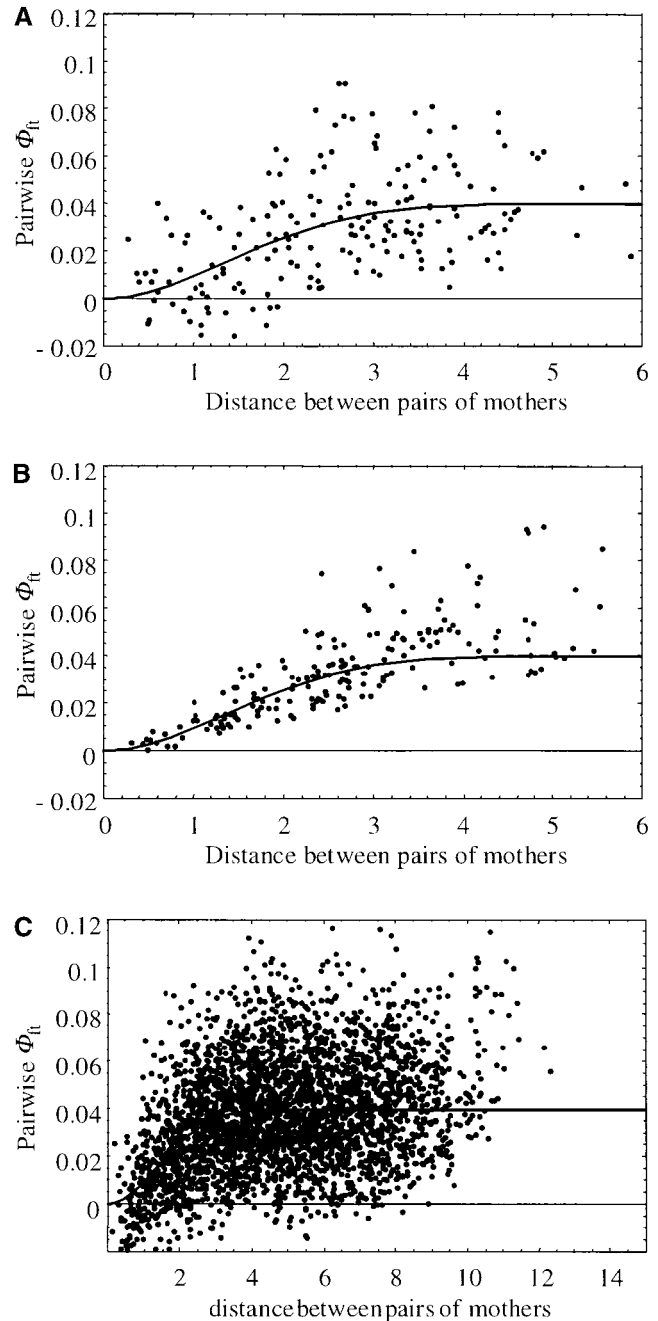


FIGURE 2.—Pairwise Φ_{it} 's obtained in a simulation. The theoretical curve is also given (solid line). (A) Density $d = 1$, axial dispersal variance $\sigma = 1$, number of sampled mothers $n_m = 20$, number of offspring sampled per mother $n_o = 20$, number of other adults sampled to estimate allelic frequencies $n_s = 30$, number of loci studied $n_L = 5$, number of alleles per locus $n_A = 10$ (denoted in text as reference situation); (B) higher numbers of offspring ($n_o = 80$); (C) higher number of mothers ($n_m = 80$), all other parameters being equal.

better to increase the number of mothers, rather than the number of offspring per mother. SMOUSE *et al.* (2001) showed that for any fixed number of mothers, n_m , the optimal value of n_o is $1/\Phi$, suggesting a strategy of setting $n_o \approx 1/\Phi_{it}$ and increasing the number of

TABLE 1
Impact of the number (n_0) of offspring sampled per mother on the various global and pairwise estimators that use density as an independently known value

n_0	$\hat{\sigma}_{g1}^a$			$\hat{\sigma}_{g2}^b$			$\hat{\sigma}_{p1}^c$			$\hat{\sigma}_{p2}^d$			$\hat{\sigma}_{p3}^e$		
	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$
20	0.156	0.147	0.214	0.035	0.087	0.093	0.325	0.318	0.455	-0.049	0.077	0.092	-0.023	0.085	0.088
40	0.143	0.130	0.193	0.028	0.074	0.079	0.150	0.170	0.227	-0.038	0.071	0.081	-0.025	0.076	0.080
80	0.145	0.115	0.185	0.031	0.064	0.071	0.076	0.119	0.141	-0.030	0.061	0.068	-0.022	0.068	0.071
160	0.139	0.116	0.181	0.027	0.064	0.069	0.034	0.086	0.092	-0.031	0.062	0.069	-0.026	0.068	0.072

Bias, standard deviation, and square root of mean-squared error ($\sqrt{\text{MSE}}$) were computed over 1000 replicates. All other parameters are set at the reference values density $d = 1$, axial dispersal variance $\sigma = 1$, number of sampled mothers $n_m = 20$, number of other adults sampled to estimate allelic frequencies $n_s = 30$, number of loci studied $n_l = 5$, number of alleles per locus $n_A = 10$.

^a Global estimate of σ , ignoring the average distance between mothers (\bar{z}), using (4).

^b Global estimate of σ , allowing for the average distance between mothers (\bar{z}), using (6).

^c Pairwise estimate of σ , using (7).

^d Pairwise estimate of σ , using (8) and (9).

^e Pairwise estimate of σ , from nonlinear regression, (10) and (11), and assuming that density (d) is known independently.

mothers, n_m , as much as possible. Under these conditions, $\hat{\sigma}_{p3}$ is a slightly better estimator than is either $\hat{\sigma}_{g2}$ or $\hat{\sigma}_{p2}$.

Number of loci vs. number of alleles: The best allocation of laboratory effort is a matter of some concern, given the substantial cost of lab assay. One cannot design the loci to order, of course, but some loci are more polymorphic than others, and one can choose among those available. The MSE decreases with an increase in the exclusion probability (E). As we pointed out in SMOUSE *et al.* (2001), one needs enough exclusion probability from the assay battery to make the enterprise profitable, but beyond a certain level of exclusionary information ($E > 0.99$), greater genomic sampling is not very helpful: An asymptotic value is reached for MSE when E becomes very high (see Table 3). Given the cost realities of laboratory assay, the best strategy would seem to be five loci with 10–20 alleles each. The pairwise estimates are considerably better than the global estimate for genetic batteries with low exclusion probability (for which case, the global estimate Φ_{fit} can be negative, which precludes estimation of $\hat{\sigma}_{g1}$ and $\hat{\sigma}_{g2}$). The difference becomes minimal for higher exclusion probabilities. In almost all cases, $\hat{\sigma}_{p3}$ remains slightly better than $\hat{\sigma}_{p2}$, but the gap decreases with increasing genetic resolution.

Joint estimation of σ and d : The estimate $\hat{\sigma}_{d1}$ shows much more variance than any estimate obtained when only σ is estimated with an extraneously imposed density (d), which is assumed to be correct (Table 4). The estimator of density \hat{d}_1 shows some bias and considerable variance, even when many offspring are sampled. The primary value of joint estimation is that it frees us from the assumption that raw stem count is a reasonable estimator of reproductive density. Again, it is better to increase the number of mothers, rather than the number of offspring, and combinations of 80 or 160 mothers and 20 offspring per mother seem to produce results on the joint estimation that can be trusted with reasonable confidence. Everything else being equal, greater replication is required for joint estimation than for estimation of σ alone. MSE also decreases faster with increasing exclusion probability than was true for the density-specified estimates. For instance, for $\hat{\sigma}_{d1}$, $\sqrt{\text{MSE}}$ is 0.355 for 5 loci and 0.208 for 20 loci (Table 5). For joint estimation, both precision and accuracy improve with increasing exclusion probability, and $\sqrt{\text{MSE}}$ continues to decrease with increasing E , even when it becomes very close to one.

Any estimate of σ that is based on an extraneous estimate of density, d , will be biased by error in that estimate of d . For instance, if parametric d and σ are both 1.0, a biased field estimate of d , say $\hat{d}_t = 0.8$, will yield an estimate of σ that has expectation $(0.8)^{-1/2} = 1.118$, producing a bias of +0.118. If \hat{d}_t is 0.5, this bias will be of 0.414, much higher than the bias of $\hat{\sigma}_{d1}$. Thus,

TABLE 2
Impact of sampling effort

Sampling effort		$\hat{\sigma}_{g2}^a$			$\hat{\sigma}_{p2}^b$			$\hat{\sigma}_{p3}^c$		
n_m	n_o	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$
20	20	0.035	0.087	0.093	-0.085	0.065	0.107	-0.017	0.089	0.090
40	20	0.045	0.074	0.086	-0.065	0.054	0.085	-0.019	0.068	0.071
80	20	0.027	0.060	0.066	-0.052	0.044	0.068	-0.021	0.052	0.056
160	20	0.006	0.042	0.042	-0.042	0.033	0.054	-0.021	0.038	0.044
20	40	0.028	0.074	0.079	-0.038	0.071	0.081	-0.025	0.076	0.080
20	80	0.031	0.064	0.071	-0.030	0.061	0.068	-0.022	0.068	0.071
20	160	0.027	0.064	0.069	-0.031	0.062	0.069	-0.026	0.068	0.072
40	80	0.043	0.060	0.073	-0.029	0.051	0.059	-0.021	0.055	0.059
80	40	0.028	0.052	0.059	-0.031	0.041	0.052	-0.021	0.045	0.049

Mean and standard deviation of the best global and pairwise estimators that use density as an independently known value for various sampling efforts, *i.e.*, combinations of number of mothers (n_m) and number of offspring (n_o). All other parameters are set at the reference values density $d = 1$, axial dispersal variance $\sigma = 1$, number of other adults sampled to estimate allelic frequencies $n_s = 30$, number of loci studied $n_L = 5$, number of alleles per locus $n_A = 10$.

^a Global estimate of σ , allowing for the average distance between mothers (\bar{z}), using (6).

^b Pairwise estimate of σ , using (8) and (9).

^c Pairwise estimate of σ , from nonlinear regression, (10) and (11), and assuming that density (d) is known independently.

deciding whether to use $\hat{\sigma}_{d1}$ or $\hat{\sigma}_{p3}$ involves a tradeoff between one's trust in the field estimate of density (\hat{d}_f) and that quantity of genetic information available.

DISCUSSION

We describe an effective means for estimating the pollen distribution function, assuming a bivariate normal distribution. Provided that density is known independently, this study shows that it is possible to design estimators that are minimally biased and that have enough precision to provide a trustworthy estimator of the average distance of pollen dispersal. Concerning global estimates, we have shown that it is important to take into account the average pairwise distance between mothers, as a way of removing the potential bias that occurs when the mothers are sampled at distances that are too close. Among the pairwise estimates, the one that exhibits the lowest MSE is the nonlinear regression. The behavior of both global and pairwise estimates is satisfactory even with 20 mothers and 20 offspring per mother, but the best way to decrease mean-squared error is to increase the number of sampled mothers, rather than the number of offspring per mother. Increasing exclusion probability (E), by increasing either the number of loci or choosing more polymorphic loci, also decreases MSE, but there is nothing much to be gained by increasing this exclusion probability above 0.99, and for many purposes, a genetic battery that yields $E > 0.90$ is quite adequate to the task.

This study was motivated by the idea that pairwise analysis could be used to extract better information from the genetic data. The results suggest that this

is indeed the case, especially for lower values of E , for which pairwise estimates yield an appreciable reduction in MSE; for very high values of E , the reduction is minimal. The failure to achieve greater gains may be traceable to a feature that is common to all pairwise analyses. The availability of multiple measures provides the impression that they should increase resolution, but each of those measures is extracted from a much smaller sample. Moreover, the collection of all pairwise Φ_{ij} -values is far from being an independent set of estimates (Φ_{jk} is not independent of Φ_{ij} and Φ_{ik}). In practice, the collection of pairwise information provides a modest improvement on a single global average, which—while it ignores the detail—has the virtue of being based on a substantially larger sample size; the pairwise strategy is better, but only mildly so. Since an iterative pairwise estimate is neither time- nor labor-intensive to obtain, particularly in view of the field and laboratory scope of such studies, it would always seem preferable to compute one.

In practice, estimating adult reproductive density is a serious problem. For example, not all adults reproduce during a given year, and phenology is variable even within a year, so that not all parents can mate at the same time. There is also variation in male fecundity within a population, as a function of age and size differences, genetic differences, and microenvironmental factors. All of these factors contribute to a reduction in the effective density of reproducing individuals. In some cases, it can be nearly impossible to count all adult plants belonging to a given species, in particular in tropical forests. Finally, and especially in the case of forest trees, populations can cover very wide areas, and

TABLE 3
Impact of the number of loci (n_t) and number of alleles (n_a) on the best global and pairwise estimators that use density as an independently known value

Loci and alleles		Exclusion probability				$\hat{\sigma}_{g^2}^d$				$\hat{\sigma}_{p2}^b$				$\hat{\sigma}_{p3}^c$			
n_t	n_a	E	$-\text{Log}(1 - E)$	Bias	SD	$\sqrt{\text{MSE}}$	$(\Phi_t \leq 0)^d$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$	
2	2	0.340	0.42	0.130	0.416	0.436	59	-0.123	0.152	0.196	-0.068	0.192	0.204				
5	2	0.646	1.04	0.093	0.265	0.281	5	-0.095	0.113	0.147	-0.041	0.142	0.147				
3	3	0.750	1.39	0.075	0.230	0.242	3	-0.090	0.110	0.142	-0.042	0.134	0.141				
5	3	0.901	2.31	0.050	0.159	0.166	0	-0.080	0.100	0.128	-0.033	0.122	0.127				
3	5	0.934	2.71	0.053	0.147	0.156	0	-0.069	0.100	0.121	-0.027	0.117	0.120				
5	4	0.970	3.50	0.044	0.131	0.138	0	-0.068	0.095	0.117	-0.028	0.111	0.115				
4	5	0.973	3.62	0.042	0.123	0.130	0	-0.065	0.092	0.112	-0.027	0.105	0.109				
5	5	0.989	4.52	0.041	0.118	0.125	0	-0.059	0.093	0.110	-0.024	0.107	0.109				
5	10	0.99964	7.92	0.035	0.087	0.094	0	-0.012	0.092	0.093	-0.017	0.089	0.090				
5	20	0.999989	11.42	0.033	0.076	0.083	0	-0.044	0.073	0.085	-0.024	0.077	0.080				
10	10	0.9999987	15.83	0.036	0.078	0.086	0	-0.041	0.076	0.086	-0.019	0.079	0.081				
20	10	0.99999999	31.67	0.037	0.076	0.084	0	-0.036	0.076	0.084	-0.018	0.078	0.080				

All other parameters are set at the reference values density $d = 1$, axial dispersal variance $\sigma = 1$, number of sampled mothers $n_m = 20$, number of offspring sampled per mother $n_p = 20$, number of other adults sampled to estimate allelic frequencies $n_s = 30$. The highest values of exclusion probability are experimentally unrealistic, but they provide the asymptotic limit, *i.e.*, the lowest value of MSE that can be reached for that parameter set.

^a Global estimate of σ , allowing for the average distance between mothers (\bar{z}), using (6).

^b Pairwise estimate of σ , using (8) and (9).

^c Pairwise estimate of σ , from nonlinear regression, (10) and (11), and assuming that density (d) is known independently.

^d Number of simulations (out of 1000) for which the estimated global Φ_t was negative and thus for which $\hat{\sigma}_{g^2}$ could not be computed.

TABLE 4
Joint pairwise estimates of d and σ for various sampling efforts

Sampling effort		\hat{d}_1			$\hat{\sigma}_{d1}$		
n_m	n_o	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$
20	20	0.151	0.726	0.742	0.016	0.266	0.267
40	20	-0.097	0.325	0.340	0.061	0.165	0.176
80	20	-0.116	0.201	0.232	0.051	0.105	0.117
160	20	-0.120	0.133	0.179	0.048	0.076	0.089
20	40	-0.052	0.419	0.422	0.043	0.183	0.188
20	80	-0.068	0.385	0.390	0.045	0.161	0.167
20	160	-0.079	0.348	0.357	0.038	0.134	0.140

All other parameters are set at the reference values density $d = 1$, axial dispersal variance $\sigma = 1$, number of other adults sampled to estimate allelic frequencies $n_s = 30$, number of loci studied $n_L = 5$, number of alleles per locus $n_A = 10$. \hat{d}_1 and $\hat{\sigma}_{d1}$ are obtained from nonlinear regression using (10) and (11).

density can be nonhomogeneous across the landscape, making any estimation of the effective density approximate and subject to uncertainty. Simply using the number of individuals above a given size class can yield a serious overestimate of d and a serious underestimate of σ . The estimator \hat{d}_1 can be contrasted with the more usual stem count; any discrepancy becomes an indicator of the extent to which various forms of heterogeneity have impacted on “effective male reproductive density,” d_e , which is likely to be less than the actual stem count. We must keep in mind, however, that very large data sets are going to be required for reliable joint estimation: large numbers of adults and offspring per adult, along with a high-resolution (exclusion probability) genetic battery. The increasing availability of numerous highly polymorphic loci at reasonable cost, however, provides some hope that we can apply the method effectively in real situations.

This study shows that, at least for bivariate normal pollen dispersion, the relationship between pollen dispersal distance and Φ_{ft} can be used to extract a useful estimate of the decay parameter, σ . The fact that we

can establish at least a numerical relation between Φ_{ft} and dispersal distance for any dispersal function (AUSTERLITZ and SMOUSE 2001a) suggests that this estimation procedure can be extended to a wider array of dispersal functions. Since we can also account for genetic structure among adults in the population (AUSTERLITZ and SMOUSE 2001b), we will also be able to design estimates that take that information into account. All such extensions will also require extensive testing. These are matters that we will leave for future work.

This approach is essential since a good estimation of contemporary gene flow is essential to understand the evolutionary processes that occur at the scale of a landscape (SORK *et al.* 1999). It is the only way to infer the consequence of various processes, which are often recent and man induced: fragmentation, loss of pollinators, and extinction of local populations. Thus, only reliable inference of this instantaneous gene flow will yield the possibility of predicting future changes for many species. TwoGener estimation should be useful in that context, because it allows us to gauge pollen flow, without typing all potential fathers in the popula-

TABLE 5
Joint pairwise estimates (\hat{d}_1 , $\hat{\sigma}_{d1}$) of d and σ from nonlinear regression, (10) and (11), for various numbers of loci (n_L), and number of alleles (n_A)

Loci and alleles		Exclusion probability		\hat{d}_1			$\hat{\sigma}_{d1}$		
n_L	n_A	E	$-\text{Log}(1 - E)$	Bias	SD	$\sqrt{\text{MSE}}$	Bias	SD	$\sqrt{\text{MSE}}$
5	5	0.989	4.52	0.214	1.154	1.174	0.072	0.348	0.355
5	10	0.9996	7.92	0.151	0.726	0.742	0.016	0.266	0.267
5	15	0.999953	9.96	-0.052	0.557	0.559	0.080	0.247	0.260
5	20	0.999989	11.42	-0.098	0.463	0.473	0.083	0.226	0.241
10	10	0.9999987	15.83	-0.058	0.490	0.493	0.075	0.243	0.255
20	10	0.9999999	31.67	-0.058	0.435	0.439	0.057	0.200	0.208

All other parameters are set at the reference values density $d = 1$, axial dispersal variance $\sigma = 1$, number of sampled mothers $n_m = 20$, number of offspring sampled per mother $n_p = 20$, number of other adults sampled to estimate allelic frequencies $n_s = 30$.

tion and even without having an estimate of adult density. Thus, it makes the method less labor intensive than alternative methods such as paternity analysis, and it can be carried out over a broader stretch of landscape and during several years.

The authors thank A. J. Irwin, A. Kremer, V. L. Sork, and S. Oddou-Muratorio for helpful commentary on the manuscript. P.E.S. was supported by the New Jersey Agricultural Experiment Station (USDA/NJAES-17106), McIntire-Stennis grant USDA/NJAES-17309, as well as by National Science Foundation grant BSR-0089238. Most simulations were performed on the UNIX machines of the Institut National de la Recherche Agronomique (Bordeaux and Jouy-en-Josas, France).

LITERATURE CITED

- AUSTERLITZ, F., and P. E. SMOUSE, 2001a Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{it} , pollen dispersal and inter-females distance. *Genetics* **157**: 851–857.
- AUSTERLITZ, F., and P. E. SMOUSE, 2001b Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure. *Genet. Res.* **78**: 271–280.
- CHAKRAVARTI, A., and C. C. LI, 1983 The effect of linkage on paternity calculation, pp. 411–422 in *Inclusion Probabilities in Parentage Testing*, edited by R. H. WALKER. American Association of Blood Banks, Arlington, VA.
- JAMIESON, A., 1994 The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Anim. Genet.* **25** (Suppl. 1): 37–44.
- JAMIESON, A., and S. C. TAYLOR, 1997 Comparisons of three probability formulae for parentage exclusion. *Anim. Genet.* **28**: 397–400.
- KIMURA, M., and G. H. WEISS, 1964 The stepping stone model of population structure and the decrease of gametic correlation with distance. *Genetics* **49**: 561–576.
- LEVENBERG, A., 1944 A method for the solution of certain nonlinear problems in least squares. *Q. Appl. Math.* **2**: 164–168.
- MALÉCOT, G., 1948 *Les mathématiques de l'hérédité*. Masson, Paris.
- MARQUARDT, D. W., 1963 An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* **11**: 431–441.
- PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY, 1988 *Numerical Recipes in C. The Art of Scientific Computing*, Ed. 2. Cambridge University Press, Cambridge, UK.
- ROUSSET, F., 1997 Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- ROUSSET, F., 2000 Genetic differentiation between individuals. *J. Evol. Biol.* **13**: 58–62.
- SLATKIN, M., 1991 Inbreeding coefficients and coalescence times. *Genet. Res.* **58**: 457–462.
- SLATKIN, M., 1993 Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.
- SMOUSE, P. E., R. J. DYER, R. D. WESTFALL and V. L. SORK, 2001 Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**: 260–271.
- SORK, V. L., J. NASON, D. R. CAMPBELL and J. F. FERNANDEZ, 1999 Landscape approaches to the study of gene flow in plants. *Trends Ecol. Evol.* **14**: 219–224.
- WOLFRAM, S., 1999 *The Mathematica Book*. Wolfram Media/Cambridge University Press, Champaign, IL/Cambridge, UK.
- WRIGHT, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- WRIGHT, S., 1946 Isolation by distance under diverse systems of mating. *Genetics* **31**: 39–59.
- WRIGHT, S., 1951 The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

Communicating editor: O. SAVOLAINEN

