Spatial models of pollen dispersal in the forage grass meadow fescue

January 20, 1997

Minna Nurminiemi*, Jarle Tufto†, Nils-Otto Nilsson‡ and Odd Arne Rognli*

* Department of Biotechnological Sciences, Agricultural University of Norway, PO Box 5040, N-1432 Aas, Norway
† Norwegian Institute for Nature Research, Tungasletta 2, 7005 Trondheim, Norway.
‡ Hilleshög AB, Box 302, S-261 23 Landskrona, Sweden.

Corresponding author: Jarle Tufto, Norwegian Institute for Nature Research, Tungasletta 2, 7005 Trondheim, Norway; E-mail: jarle.tufto@nina.nina.no; Phone: +47 73 58 06 46; Fax: +47 73 91 54 33.

Running title: Analysis of pollen dispersal

Keywords: Meadow fescue, gene flow, pollen, Brownian motion, maximum likelihood, gravity, asynchrony, protandry, protogyny.

In press in Evolutionary Ecology

Summary

Several bivariate probability distributions generated by different underlying dispersal mechanisms are fitted to the observed frequencies of an isozyme marker gene using a maximum likelihood approach. The pollen dispersal data was generated using two experimental populations of meadow fescue (*Festuca pratensis* Huds.), homozygous for different allozymes at the (*Gpi-2*) locus, arranged in a circular donor-acceptor field design. The contribution of all plants in the populations depending on differences in plant position, fecundity, and flowering time are taken into account when fitting the different models. Several approximate likelihood-ratio tests are done between alternative nested models, and a wind threshold model with bimodality in the wind direction is selected. The evolutionarily important variances and expectations of the gene displacement under the selected model are calculated. It is also shown that the underlying probability distribution is significantly more than exponentially leptokurtic. By fitting a distribution of deposition in all three dimensions to the data, taking into account differences in plant height, separate estimates of additional physical parameters are obtained, showing that gravity and vertical random movements are more important than intervening vegetation in limiting pollen dispersal in meadow fescue. According to the model, plants with a high seed yield, contribute over-proportionally with pollen to neighbouring plants.

Introduction

Semi-domesticated species, such as perennial grasses, are often major components of large and important ecosystems. At the same time, natural populations found in these ecosystems are important sources for the continuous development of new cultivars. In the near future we may also see the development of transgenic cultivars of forage grasses (Wang *et al.* 1992; Fuettere and Potrykus 1995). Cultivars of perennial grasses are often well adapted to local climatic conditions. Being wind-pollinated, they easily disperse pollen into surrounding natural populations. This can increase both the risk and the effects of gene flow on natural populations.

Pollen traps (e.g. Bateman 1947c; Lindgren *et al.* 1991; Dafni 1992; Timmons *et al.* 1995), morphological markers (Bateman 1947b; Chaudhary *et al.* 1980), isozymes (Bos *et al.* 1986; Yazdani *et al.* 1989; Adams and Birkes 1991; Xie and Knowles 1994), chloroplast DNA (Arnold *et al.* 1992), and transgenes (Scheffler *et al.* 1993; Kareiva *et al.* 1994; Morris *et al.* 1994) have been used frequently in studies of gene flow in plant populations. Many of these studies only report the frequency of the marker gene at different distances from the donor field, or the contamination by foreign pollen into some breeding population. Utilizing information from these studies is difficult, however, since the observed frequencies of the marker gene are typically influenced not only by the distance to the donor field, but also by other factors such as the size, form and density of the donor and acceptor fields, geographic or vegetative barriers, placement in relation to predominant wind direction, and varietal differences in time of flowering (Griffiths 1950), factors which are often not reported.

Of particular interest in studies of gene flow in spatially continuous plant populations is the bivariate distribution $f(x, y)$ of the reproductive events associated with the pollen originating from each individual plant in the population. To a good approximation, the most important properties of this distribution can be summarized through its variances and expectations. The variances determine the opportunities for local adaptations (Slatkin 1973; Nagylaki 1978b). If dispersal is asymmetric, clinal variation will also depend on the expectations of the displacement of genes from one generation to the next (Nagylaki 1978a). The variances of the dispersal distribution also determine the speed of the wave-front formed when new advantageous genes enter geographically structured populations (Fisher 1937). More exact solutions to this problem, based on the entire distribution, are also available (Weinberger

1982).

In the present investigation, several possible spatial distributions $f(x, y)$ of the reproductive events, based on different underlying pollen dispersal mechanisms, and mixtures of such distributions, are fitted to marker gene frequency data by maximum likelihood, using and further developing methods given in Tufto *et al.* (1997). We test several alternative models, taking into account the differences in plant positions, plant height, flowering time, and fecundity. The variances and expectations of the dispersal distributions of different alternative models are then calculated.

We analyze a data set for the forage grass meadow fescue (*Festuca pratensis* Huds.), generated using non-transgenic plant material, genetically marked for naturally occurring allozymes at the *Pgi-2* locus. Meadow fescue was selected since it is an important diploid forage grass species, one that is cross-pollinated with a high degree of self-incompatibility (Lundqvist 1961, 1962). Its reproductive biology is thus representative for the majority of temperate grass species.

Materials and methods

Plant materials

The local cultivar 'Løken', which originates in Valdres in southern Norway (61°7'N, 9°4'E, 550 m a.s.l.), was selected as the plant material. Two populations were established by selecting plants within 'Løken' homozygous for different alleles (allozymes) at the *Pgi-2 (Phosphoglucoisomerase-2)*, E.C. 5.3.1.9) locus. In Norwegian fescue, four different alleles are present at this locus. One population, homozygous for the *b*-allele, served as donor plants (*Pgi-2/bb*). The other population, homozygous for the *c*-allele (*Pgi-2/cc*), served as acceptor plants. Alleles *b* and *c*, with typical frequencies of around 0.4, are the most frequent alleles at this locus. Donor and acceptor plants were checked twice for *Pgi-2* genotype before being planted in the experimental field in June 1995.

Experimental design

The experimental field was established on a local farm at the village of Ås (59°40'N 10°48'E), being planted in a homogeneous, spring-sown oat-field.

The panicles of the meadow fescue plants extended about 10 cm above the top of the canopy of oats.

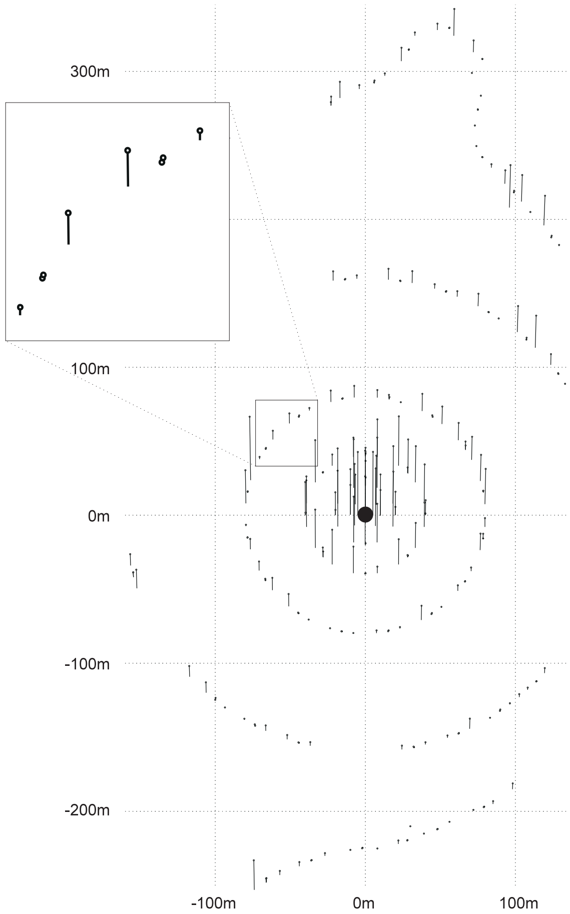

Figure 1: Observed frequency of the donor allele at each sampled plant in the experimental plot, represented by the length of the line at each respective plant.

The donor field consisted of 451 plants uniformly distributed in a circular plot 9.6 m in diameter. The plants in the plot were spaced approximately 40 cm apart.

The acceptor field surrounding this donor field consisted of 244 plants in 6 concentric circles at distances of 5, 10, 20, 40, 80, and 160 m from the center of the donor field (Fig. 1). Additional acceptor plants, placed at varying distances (200-300 m) from the donor field, were located along the border of surrounding vegetation, which consisted of forests and cereal fields. The plants were spaced 16, 8, 0.3, 8, and 16 m apart along the circumference of the 7 circles. Plants spaced 0.3 m apart will be referred to as 'double' and other plants as 'single' (Fig. 1). Double and single acceptor plants were used to investigate the effect of pollen from neighbouring acceptor plants on the donor-pollen's success in fertilizing the acceptor plants.

Feral populations of meadow fescue were growing along the roadsides and along paths leading to the cultivated fields. These plants were observed every second day, visible panicles being cut down prior to anthesis so as to keep gene flow from these populations into the experimental field at a minimum. This was done up to 500 m from the center of the experiment.

The time of anthesis, t_i, was recorded when fully exserted anthers were observed on at least three inflorescences of the plant (Griffiths 1950). Plant height, z_i, measured from the basis to the middle of the panicles, was recorded during the seed harvest in 1995. The seed yield, m_i, was measured as gram per plant.

Electrophoresis and isozyme analysis

A minimum of 88 seeds (if available) from every acceptor plant were germinated in a greenhouse, the seedlings being used to obtain plant sap samples. These samples were analyzed for allelic composition at the *Pgi-2* locus by electrophoresis, using methods described by M. Nurminiemi and O. A. Rognli (in prep). The number of seeds that were pollinated by the donor pollen out of the total number of seeds sampled, could then be determined from the number of heterozygous *bc* (*Pgi-2/bc*) seedlings. Altogether, 24 672 seedlings of the 244 acceptor plants, were analyzed.

Dispersal models

Several plausible spatial distributions can be derived from different sets of assumptions about the underlying process of dispersal. Here we use three different models described in greater detail in Tufto *et al.* (1997). The objective of these models is in part to test our understanding of the underlying mechanisms against real data, and in part to obtain accurate estimates of such important evolutionary parameters as the dispersal variances and expectations. Also, modeling these distributions with some accuracy allows more unbiased estimates of other interesting parameters to be obtained than would otherwise have been possible.

The normalizing constants of the various dispersal distributions below have been omitted since, as will become clear in the next subsection, they are irrelevant for the type of data considered in this paper. We will also express (x, y) through polar coordinates (r, θ), whenever this is convenient.

In the first model, we let the radial distance $R = \sqrt{X^2 + Y^2}$ at deposition follow the Weibull distribution. The corresponding bivariate distribution then takes the form

$$f(x, y) \propto r^{b-2} e^{-ar^b}. \tag{1}$$

When the shape parameter $b = 2$, model (1) equals the binormal distribution, and when $b = 1$ the distance R to deposition is exponentially distributed, as is sometimes hypothesized in the literature (e.g. Kareiva *et al.* 1994). Model (1) is similar to a model suggested by Bateman (1947a) and Wright (1969), but their alternative formulation have fewer interpretations (Tufto *et al.* 1997).

The second model assumes that the movement of each individual pollen grain is a three-dimensional Brownian motion and that deposition on the reproductive and non-reproductive parts of the vegetation are equally probable at any point in time and space. Assuming that all reproductive events occur in a thin flower layer only, the conditional bivariate probability distribution of these reproductive events generated by the movement process is

$$f(x, y) \propto \frac{1}{r} e^{a^2(\tau_x x + \tau_y y) - \sqrt{a^2 + a^4\tau_x^2 + a^4\tau_y^2}\, r}. \tag{2}$$

The flexibility of this model lies in the fact that the expectations $E(X) = \tau_x$ and $E(Y) = \tau_y$, as a result of wind directionality, can differ from zero. Due to the assumptions made, the distribution is also strongly leptokurtic, which is often found to be the case in empirical studies of air-borne pollen dispersal.

If we in addition assume that the velocity and direction of the wind change over time and that there is some wind velocity threshold at which pollen is released from the anthers and taken by the wind, we arrive at a third model,

$$f(x, y) \propto \frac{1}{r^{3/2}} e^{-\lambda_0 r + \kappa \cos(\theta - \theta_0)}, \tag{3}$$

which under certain conditions is even more leptokurtic than (2). The parameter θ_0 represents the 'mean' wind direction, $1/\kappa$ represents the 'variance' of the wind direction, and λ_0 is a scale parameter.

To account for a possible bimodality in the distribution of the wind direction, mixtures of the above models can be constructed, for example from (3). If we assume that one wind direction θ_0 is dominant during a proportion α of the time, and that another wind direction θ_1 is dominant in the remaining proportion $1-\alpha$, and in addition assume that the other parameters of (3) are equal for each component of the mixture, we arrive at a fourth more complex model

$$f(x,y) \propto \frac{1}{r^{3/2}} e^{-\lambda_0 r} \left(\alpha e^{\kappa \cos(\theta-\theta_0)} + (1-\alpha) e^{\kappa \cos(\theta-\theta_1)} \right). \tag{4}$$

Note that (3) is a special case of (4) when $\alpha = 1$.

If the height z_i of each plant has been measured, the complete trivariate distribution of the deposition in space

$$f(x,y,z) \propto \frac{1}{r^{3/2}} e^{-\frac{(z-\mu_r r)^2}{2\sigma_r^2 r} - \lambda_r r} \left(\alpha e^{\kappa \cos(\theta-\theta_0)} + (1-\alpha) e^{\kappa \cos(\theta-\theta_1)} \right), \tag{5}$$

corresponding to the underlying process that generates (4), may be relevant to consider (see derivation in Appendix A). This distribution involves three additional parameters: the ratio μ_r of the vertical to the horizontal (threshold) movement velocity, the rate σ_r^2 at which the variance of Z increases with r, and the probability λ_r of pollen deposition per unit distance. We shall see that these additional parameters can be estimated if a sufficient amount of data is available. Note that (4) becomes a special case of the more general model (5) when σ_r^2 tends to infinity. The parameters of the various distributions are summarized in the Table 1.

Table 1: List of symbols

Parameters	
Weibull model (1)	
a	Scale parameter
b	Shape parameter
Brownian motion model (2)	
a	Scale parameter
τ_x	Asymmetry in the x-direction
τ_y	Asymmetry in the y-direction
Wind threshold model (3)	
λ_0	Scale parameter

$1/\kappa$	Dispersion of wind direction
θ_0	Mean wind direction
Bimodal wind threshold model (additional parameters) (4)	
θ_1	Second wind direction
α	Proportion of θ_0
Trivariate, bimodal wind threshold model (parameters replacing λ_0) (5)	
λ_r	Probability per unit distance of deposition in $r, r+dr$ at any height z
μ_r	Ratio μ_z/μ_0 of the vertical to the horizontal threshold movement velocity
σ_r	Rate of increase in $\mathrm{Var}(Z)$ with distance r
Asynchrony term (6)	
σ^2_{mf}	Within-individual variance in floret flowering times
δ	Degree of protandry/protogyny
Fecundity term (6),(B.2)	
β	Exponent of power-function relationship between pollen production and seed yield.
Covariates	
x_i, y_i, z_i	Position of plant i in space
t_i	Time of anthesis of plant i
m_i	Plant i's seed yield
Response variables	
M_i	Number of seeds collected from plant i pollinated by genotype $k = 1$
N_i	Total number of seeds collected from i

Model fitting

Provided that we can express the probability of the data as a function of the unknown parameters, models such as these can be fitted to data that can include several non-point pollen sources and any number of plant genotypes. This can be done numerically by starting with some initial parameter values and changing these repeatedly until the likelihood is maximized.

Let (x_i, y_i, z_i) be the known position of plant i, possibly including its height z_i, and let $k_i = 1$ for plants of genotype *bb* and $k_i = 2$ for plants of genotype *cc*. To calculate the amount of pollen of genotype k that arrives at a particular plant i, the variable contribution of each plant $j = 1, 2, \ldots, n$ in the population has to be taken into account. The contribution to a recipient plant i from a single plant j will depend, first of all, on its position in space $(x_i - x_j, y_i - y_j, z_i - z_j)$ relative to the recipient plant i.

Secondly, if some variable m_j related to pollen production has been measured, say seed yield, this can also be taken into account by assuming that pollen production is related to seed production according to some particular function, e.g. the power function m_j^β. Under the hypothesis that the proportion of resources allocated to pollen is independent of the total amount of resources used for reproduction by a particular plant, β can be expected to be close to 1.

Thirdly, the contribution of plant j may also depend on the degree of synchrony in the flowering times t_i and t_j of plants i and j, and on the effects of protandry, acting through the function $h(t_i, t_j)$ described in appendix B. This function involves two parameters: the amount of overlap in time σ^2_{mf}, and δ representing the amount of protogyny or protandry.

Taking these factors into account the contribution of all the plants of genotype k to a particular plant i becomes

$$g_{ik} = \sum_{\substack{j=1,2,\ldots,n,\\ j \neq i, k_j = k}} m_j^\beta h(t_i, t_j) f(x_i - x_j, y_i - y_j, z_i - z_j). \tag{6}$$

Note that the summation, for a self-incompatible species in which each individual plant cannot be fertilized by its own pollen, is over plants for which $i \neq j$ only. No account is taken of the rare occurrence of non-reproductive events due to different, but incompatible genotypes. This is a sensible assumption since the number of incompatible alleles is expected to be very high in populations of this species.

The proportion of genotype $k = 1$ (donor pollen) out of the total amount of pollen arriving at plant i then becomes

$$p_i = \frac{g_{i1}}{g_{i1} + g_{i2}}. \tag{7}$$

Let M_i be the number of seeds pollinated by genotype $k = 1$ out of the total number of seeds N_i collected from plant i. To calculate the total probability of the data, it is reasonable to assume that M_i at each sampled plant i, follows a binomial distribution with parameters p_i and N_i. The likelihood function for the data is then

$$L(\boldsymbol{\theta}|M_1,\ldots,M_n) = \prod_{i=1}^{n} \binom{N_i}{M_i} p_i^{M_i}(1-p_i)^{N_i-M_i} \tag{8}$$

The value of the parameter vector $\boldsymbol{\theta}$ that maximizes (8) can be found numerically using standard methods. Approximate values of the standard errors and correlations between the maximum likelihood estimates of the parameters can be found by parametric bootstrapping (Efron and Tibshirani 1993) by simulating binomial data (Press *et al.* 1986) from the fitted model.

In count data of this sort, the variance of the response variable M_i is often larger than the expected binomial variance. This extra variance can be due to dependencies between the individual bernoulli events. If pollen grains arrive at each plant in clusters, for example, this would increase the variance of the response variable. The amount of such over-dispersion is usually quantified as the factor σ^2 by which the response variable M_i exceeds the binomial variance. The binomial assumption can, however, still be used when fitting the model, but the extra variation must be accounted for when the standard errors are calculated (McCullagh and Nelder 1989, p. 124), as well as in the model selection process (see below).

Model selection

The objective of the model selection process is to find the simplest model that best explains all significant structures in the data. In the present case we have, in addition to choosing between five competing dispersal models, the opportunity to decide whether or not to include β and σ^2_{mf} (and possibly δ) in the model. This provides a total of $5\times 2\times 3 = 30$ possible models for the data. As an informal model selection criterion, we borrow some methodology from the theory of generalized linear models (McCullagh and Nelder 1989).

In generalized linear models the deviance of a fitted model H_0 is defined as two times the difference between the maximum achievable log likelihood and the log likelihood of the fitted model. Under H_0, with r residual degrees of freedom and n_0 parameters in the model, the deviance $D(H_0)$ is distributed asymptotically or approximately as $\chi^2_{r-n_0}$. For two nested models, if n_0 is the number of parameters of the null model H_0, and n is the number of parameters of the extended more general model H, the change in deviance, $D(H_0) - D(H)$, is approximately $\chi^2_{n-n_0}$ distributed.

However, if the true model has not yet been achieved, or if there is over-dispersion in the data, these tests tend to become inflated, that is, they tend to detect significant structures in the data too frequently. In such cases, it is recommended to use the test statistic

$$\frac{(D(H_0) - D(H))/(n - n_0)}{\hat{\sigma}^2}, \tag{9}$$

where $\hat{\sigma}^2$ is the amount of over-dispersion estimated from the standardized residuals of the fitted model, and the deviances are calculated as before. This statistic is approximately $F_{n-n_0,r-n_0}$ distributed (SAS Institute Inc. 1996, p. 268). For large values of (9), H_0 is rejected in favor of the more general model H.

It should be stressed that these tests can only be used for models that are nested. For non-nested models, that is, for cases in which the less dimensional model is never a special case of the more dimensional one, e.g. model (1) and (3), no formal tests exist. As we shall see, the deviance can in such cases in fact increase with the model dimensionality.

It should also be emphasized that the models here are nonlinear, which may make the use of (9) questionable.

Results

The observed frequencies of donor pollen at each acceptor plant are presented in Figure 1. The frequencies of donor allele b in acceptor offspring are given as bars at the position of the respective acceptor plants. It can be seen clearly in Figure 1 that the main pattern in the data is that pollen dispersal decreases with distance, that the dispersal distribution is strongly leptokurtic, and that there is a large effect of the distance to the nearest neighbouring plants. The expected frequencies of the donor allele predicted by the selected model are shown in Figure 2. Clearly, the model predicts the main patterns in the data very well.

Figure 3 shows a subset of models in the neighbourhood of the model that we believe explains the most important structures in the data. Several interesting points should be noted.

The shape parameter of the fitted Weibull model (1) was estimated to $\hat{b} = 0.65$, which is close to $\hat{b} = 0.60$ for the maize data of Bateman (1947c)

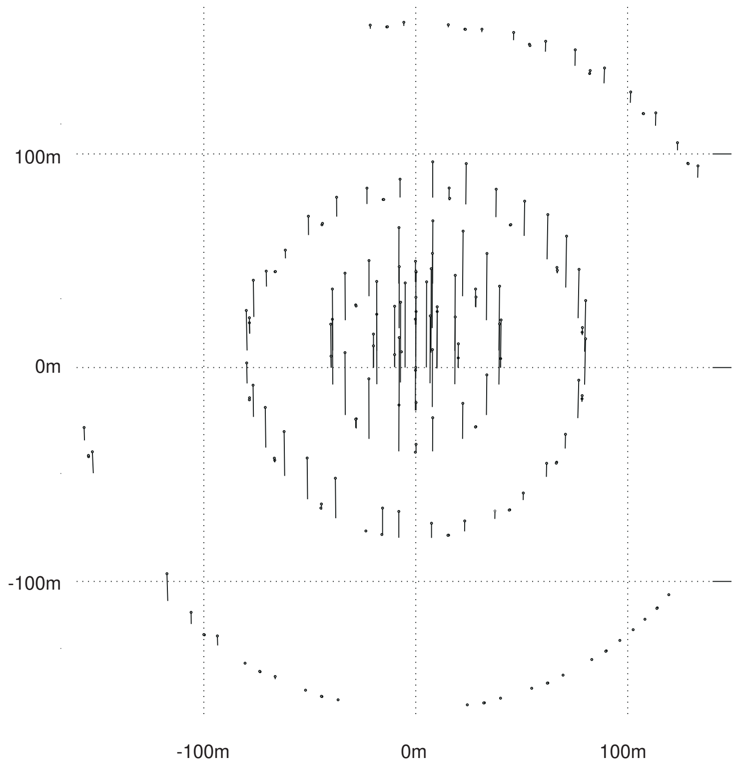

Figure 2: Expected frequency of the donor allele p_i at each sampled plant in the experimental field, predicted from the selected bimodal wind threshold model.

estimated by Tufto *et al.* (1997). The data was significantly more leptokurtic than would be expected if the radial distance R follows an exponential distribution. This was tested by fitting the reduced model $b = 1$ to the data.

It is also rather surprising that the two-parametric Weibull model, judged by its maximum likelihood, appeared to have a better fit than the three-parametric Brownian motion and the unimodal wind threshold models, yet this is what might readily happen with non-nested models. The explanation for the reduced fit is also simple: whereas the Weibull model has some flexibility in adjusting to kurtosis in the data, the flexibility of the other two three-parametric models in representing wind directionality is lost, since there appears to be two dominant wind directions. This was the reason for constructing the bimodal wind threshold model, which yielded very signifi-

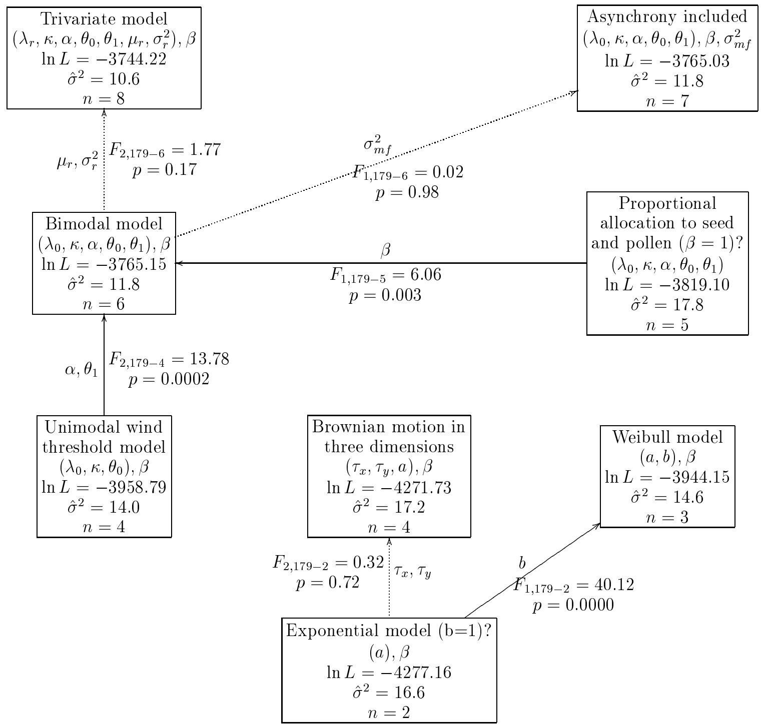

Figure 3: Some possible models for the data. The arrows point towards the more general models, that is, models that include the less general ones as special cases. Significance and non-significance are indicated by solid and dotted lines, respectively. No tests have been made between non-nested models. The parameters estimated are shown for each model.

cantly improved fit to the data. The parameter estimates under this model are given in Table 2A.

Having arrived at this model, we tried further extentions by entering additional covariates, such as flowering times t_i through introducing the $h(t_i, t_j)$ term in (6), and plant height z_i through using (5) instead of (4) as the dispersal model. None of these extensions of the overall model resulted in any significant improvement in the fit to the data. It is encouraging, however, that the fitted trivariate model appeared to give reasonable estimates of the additional parameters μ_r, σ_r^2, and λ_r (see Table 2B). In the less general model (4), these effects were only estimated jointly through the combined parameter $\lambda_0 = \lambda_r + \mu_r^2/2\sigma_r^2$. The separate estimate of $\lambda_r = 0.0007$ was

Table 2: Parameter estimates $\pm$ standard errors of (A) the selected bimodal wind threshold model (4), and (B) the extended trivariate model (5). The standard errors were calculated by parametric bootstrapping by simulating binomial data from the fitted model, and then multiplied by the estimated amount of over-dispersion $\hat{\sigma}$ following McCullagh and Nelder (1989) p. 126.

(A) Bimodal model		(B) Trivariate model	
Parameter	Estimate	Parameter	Estimate
β	1.41 ± 0.11	β	1.40 ± 0.10
κ	1.94 ± 0.23	κ	1.86 ± 0.21
α	0.58 ± 0.05	α	0.60 ± 0.06
θ_0	0.67 ± 0.11	θ_0	0.69 ± 0.09
θ_1	-2.78 ± 0.10	θ_1	-2.75 ± 0.11
λ_0	0.02 ± 0.01	λ_r	0.0007 ± 0.0002
		μ_r	0.06 ± 0.01
		σ_r	0.28 ± 0.06
EX	$0.76\,m$		$-0.23\,m$
EY	$4.04\,m$		$4.33\,m$
$SD(X)$	$31.10\,m$		$29.26\,m$
$SD(Y)$	$28.16\,m$		$27.28\,m$
$\rho(X,Y)$	0.36		0.14

considerably smaller than the estimate of $\mu_r^2/2\sigma_r^2 = 0.022$ (Table 2B), indicating that gravity and vertical random movements are more important than intervening vegetation in limiting pollen dispersal.

We also tested whether β in the power function relationship between pollen production and seed yield significantly differed from the expected value of 1, as indicated by the estimate of $\beta = 1.41$, through fitting the reduced model $\beta = 1$ to the data, keeping the rest of the model unchanged. This test turned out to be significant, that is, we can conclude that each individual plant's pollen production scales over-proportionally with seed yield.

The fit of the selected model (4) was also examined by plotting the standardized residuals against different covariates. As can be seen in Figure 4A, there is some discrepancy between the observed and the predicted gene frequencies at distances of 160 m from the donor field.

We also plotted the residuals for single and double plants separately, that is, plants located 8 m and 0.3 m from their nearest neighbours, respectively. As can be seen in Figure 4B and in comparing Figure 2 with Figure 1, the model essentially predicts these small-scale effects of pollen competition correctly, which is encouraging.

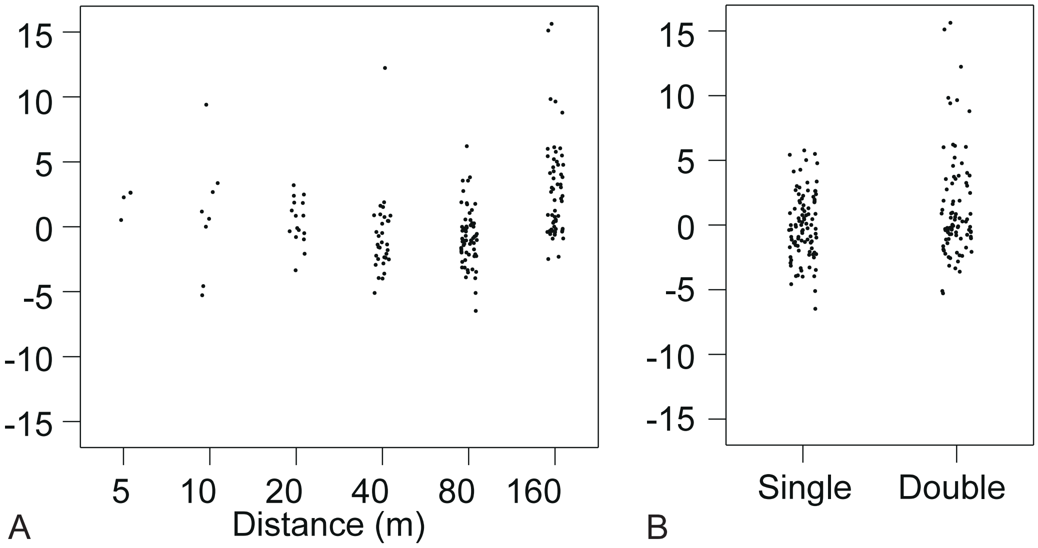

Figure 4: Jittered scatter plots of standardized residuals against a) distance from the donor field, and b) single vs. double plants.

The residuals of the model were also plotted against the direction relative to the donor field, and against the expected values predicted under the fitted model. No systematic departures from the model were apparent in any of these plots.

Discussion

Although the models described in this paper may seem overly complex, this is no appreciable problem if large amounts of data are available. From a model fitting point of view, the complexity of the models presents no particular difficulties when the likelihood is maximized numerically. Rather, this gives us flexibility in writing models that are based to a greater extent on biological assumptions rather than models limited by the requirements of mathematical tractability, such as is the case with more traditional statistical regression models.

Many studies typically report only the estimated frequencies of the marker gene at different distances from the donor field, ignoring the problem of translating these frequencies to a probability distribution $f(x, y)$. According to Assuncão and Jacobi (1996), this translation is easy to perform and "so standard in the literature that its assumptions rarely are discussed." Those authors, however, provide no advice on to how to carry out this translation. In our approach, information on the density and the distribution of both the donor and the recipient plants is taken into account, explicit assumptions are made concerning the underlying probability distribution $f(x, y)$ that determines the observed frequencies of the marker gene, and inferences regarding this distribution are in turn made from the observed frequencies, using basic statistical principles.

There is good agreement generally between the observed frequencies of the marker gene and expected values predicted by the fitted selected model. Certain discrepancies were evident, however, at distances 160 m from the donor field. There are two possible causes for this. First, since the $Gpi - 2$ alleles employed within the experimental plot were also present in the surrounding areas, plants located at the outermost circles of the plot inevitably experienced a minute amount of gene flow from the outside. Secondly, some of the discrepancy may be due to a possible inability of our model to predict all the leptokurtosis present in the underlying dispersal distribution. It seems difficult to distinguish these possibilities, however, without further information concerning the allele frequencies and plant densities in surrounding areas. Such data have been collected and will be presented in a forthcoming paper (M. Nurminiemi and O. A. Rognli, in prep)

An assumption in the models used here which is specific to airborne pollen, is that the density of the vegetation is independent of the height z above the ground (Tufto *et al.* 1997). This could make the applicability of these models questionable. However, in fitting the full trivariate model (5), we found that, in meadow fescue, gravity and vertical random movements were more important than intervening vegetation in limiting pollen dispersal. The fact that the density of the vegetation depends on z may thus not be particularly important. Our results suggest that models that ignore this dependency may still predict the spatial pattern of pollination events with sufficient accuracy.

We have not discussed the implications of our models for experimental design. It would be desirable to have as much information available as possible on gene frequencies and plant densities in areas surrounding the experiment. Although it seems sensible to also arrange for one of the plant genotypes being located in a small dense plant field surrounded by the other genotype distributed in a larger, less dense plant field, so as to be able to extract as much information as possible from the data on the details of $f(x, y)$, only simulation studies can give us information about the efficiency of different experimental designs. To be able to fit the model, however, it is not necessary that plants are arranged in any particular way.

Naturally occurring isozymes used in this study should behave to a large extent in the same way as transgenes during the dispersal phase of the life history. Results from experiments using isozymes as genetic markers may thus be used to predict the fate of transgenes being introduced into the environment through the deliberate release of transgenic cultivars. Estimates of the amount of dispersal from one generation to the next is an important part of the information needed to make such predictions. It should be emphasized, however, that the invasiveness of transgenes released in the environment also will depend on whether the transgene will give its carrier a selective advantage in the wild. Since many of the genetic modifications now being tested in agriculture are developed with the aim of producing better-performing organisms in an agricultural setting, it can be expected that some modifications will be potentially advantageous. Only a few studies, e.g. Crawley *et al.* (1993), have compared the performance of both transgenic and non-transgenic plant varieties in a large variety of environmental settings, even though the kind of information obtained from such studies perhaps is of greater importance in the risk assessment of transgenic organisms in general.

An MS-DOS program in Pascal, for doing the model fitting, and accompanying source code and documentation, are available on the world-wide-web at `http://www.nina.no/jtu/dispersal/`.

Acknowledgements

We wish to thank Einar Berg, Øyvind Jørgensen, Aslaug Berg and Torleiv Veum for excellent technical assistance, Astrid Myhre and Hilde Gunn Opsahl-Ferstad for help in the selection of plant material, and Kjetil Hindar and Steinar Engen for comments on the manuscript. We gratefully acknowledge the financial support provided by the Norwegian Research Council.

References

Adams, W. T. and Birkes, D. S. (1991) Estimating mating patterns in forest tree populations. In S. Fineschi, M. E. Malvolti, F. Cannata and H. H. Hattemer, eds., *Biochemical Markers in the Population Genetics of Forest Trees*, pp. 157–172. SPB Academic Publishing bv, Hague.

Arnold, M. L., Robinson, J. J., Buckner, C. M. and Bennett, B. (1992) Pollen dispersal and interspecific gene flow in Lousiana irises. *Heredity* **68**, 399–404.

Assuncão, R. and Jacobi, C. M. (1996) Optimal sampling design for studies of gene flow from a point source using marker genes or marked individuals. *Evolution* **50**, 918–923.

Bateman, A. J. (1947a) Contamination in seed crops. III. Relation with isolation distance. *Heredity* **1**, 303–336.

Bateman, A. J. (1947b) Contamination of seed crops. I. Insect pollination. *J. Genet.* **48**, 257–75.

Bateman, A. J. (1947c) Contamination of seed crops. II. Wind pollination. *Heredity* **1**, 235–246.

Bos, M., Harmens, H. and Vrieling, K. (1986) Gene flow in *Plantago* I. Gene flow and neighbourhood size in *P. lanceolata*. *Heredity* **56**, 43–54.

Chaudhary, H. R., Jana, S. and Acharya, S. N. (1980) Outcrossing rates in barley populations in the Canadian prairies. *Can. J. Genet. Cytol.* **22**, 353–360.

Crawley, M. J., Hails, R. S., Rees, M., Kohn, D. and Buxton, J. (1993) Ecology of transgenic oilseed rape in natural habitats. *Nature* **363**, 620–623.

Dafni, A. (1992) *Pollination Ecology: A Practical Approach.*. Oxford University Press, Oxford.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman & Hall, London.

Fisher, R. A. (1937) The wave of advance of advantageous genes. *Ann. Eugen. Lond.* **7**, 355–369.

Fuettere, J. and Potrykus, I. (1995) Transformation of poaceae and gene expression in transgenic plants. *Agronomie* **15**, 309–319.

Griffiths (1950) The liability of seed crops of perennial ryegrass (*Lolium perenne*) to contamination by wind-borne pollen. *J. Agric. Sci.* **40**, 19–36.

Kareiva, P., Morris, W. and Jacobi, C. M. (1994) Studying and managing the risk of cross-fertilization between transgenic crops and wild relatives. *Mol. Ecol.* **3**, 15–21.

Lindgren, D., Yazdani, R., Lejdebro, M.-L. and Lejdebro, L.-G. (1991) The spread of conifer pollen from a point source. Do pollen fly like birds or fall like stones. Tech. Rep. 10, Swedish Univ. of Agric. Sci. Dept. of For. Genet. and Plant Physiol.

Lundqvist, A. (1961) Self-incompatibility in *Festuca pratensis* Huds. *Hereditas* **47**, 542–562.

Lundqvist, A. (1962) The nature of the two-loci incompatibility system in grasses II. Number of alleles at the incompatibility loci in *Festuca pratensis* Huds. *Hereditas* **48**, 169–181.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. Chapman & Hall, London.

Morris, W. F., Kareiva, P. M. and Raymer, P. L. (1994) Do barren zones and pollen traps reduce gene escape from transgenic crops? *Ecol. Appl.* **4**, 157–165.

Nagylaki, T. (1978a) Clines with asymmetric migration. *Genetics* **88**, 813–827.

Nagylaki, T. (1978b) Random genetic drift in a cline. *Proc. Natl. Acad. Sci. USA* **75**, 423–426.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986) *Numerical Recipes in Pascal*. Cambridge University Press, Cambridge.

SAS Institute Inc. (1996) *SAS/STAT Software: Changes and Enhancements Through Release 6.11*. Cary, NC.

Scheffler, J., Parkinson, R. and Dale, P. (1993) Frequency and distance of pollen dispersal from transgenic oilseed rape (*Brassica napus*). *Transgenic Res.* **2**, 356–364.

Slatkin, M. (1973) Gene flow and selection in a cline. *Genetics* **75**, 733–756.

Timmons, A., O'Brien, E., Charters, Y., Dubbels, S. and Wilkinson, M. (1995) Assessing the risks of wind pollination from fields of genetically modified *Brassica napus* ssp. *oleifera*. *Euphytica* **85**, 417–423.

Tufto, J., Engen, S. and Hindar, K. (1997) Stochastic dispersal processes in plant populations. *Theor. Pop. Biol. (in press)* .

Wang, Z., Takamizo, T., Iglesias, V., Osusky, M., Nagel, J., Potrykus, I. and Spangenberg, G. (1992) Transgenic plants of tall fescue (*Festuca arundinacea* Schreb.) obtained by direct gene transfer to protoplasts. *Bio/Technology* **10**, 691–696.

Weinberger, H. F. (1982) Long-time behaviour of a class of biological models. *SIAM J. Math. Anal.* **13**, 353–396.

Wright, S. (1969) *Evolution and the Genetics of Populations. Volume 2: The Theory of Gene Frequencies*. University of Chicago Press, London.

Xie, C. Y. and Knowles, P. (1994) Mating system and effective pollen immigration in a Norway spruce (*Picea abies* L. Karst) plantation. *Silvae Genetica* **43**, 48–52.

Yazdani, R., Lindgren, D. and Stewart, S. (1989) Gene dispersion within a population of *Pinus sylvestris*. *Scand. J. For. Res.* **4**, 295–306.

A Trivariate model

If it is assumed that the two horizontal infinitesimal variances are small compared with the wind velocity, and if some threshold wind velocity μ_0 exist at which pollen is released from the anthers, then $R = \mu_0 T$ and the unconditional distribution of R will, like T, be approximately exponential (Tufto *et al.* 1997). The simultaneous distribution of R and Z is then

$$f_{R,Z}(r,z) = f_{Z|R=r}(z) f_R(r). \tag{A.1}$$

Using equation (A.3) from Tufto *et al.* (1997), it follows that

$$\begin{aligned} f_{X,Y,Z}(x,y,z) &= \frac{1}{r} f_{R,Z}(r,z) f_\Theta(\theta) \\ &= \frac{1}{r} f_{Z|R=r}(z) f_R(r) f_\Theta(\theta) \\ &\propto \frac{1}{r^{-3/2}} e^{-\frac{(z-\mu_z r/\mu_0)^2}{2\sigma_z^2 r/\mu_0} - \frac{\lambda}{\mu_0} r} \left(\alpha e^{\kappa \cos(\theta-\theta_0)} + (1-\alpha) e^{\kappa\cos(\theta-\theta_1)}\right). \end{aligned} \tag{A.2}$$

Only the combined parameters

$$\mu_r = \frac{\mu_z}{\mu_0}, \qquad \sigma_r^2 = \frac{\sigma_z^2}{\mu_0}, \qquad \text{and } \lambda_r = \frac{\lambda}{\mu_0}. \tag{A.3}$$

will be identifiable from the data.

It should be noted that the variances and expectations of X and Y in this model no longer depend on the fitted distribution $f(x,y,z)$ only, but also on the distribution of the heights of neighbouring flowers. This is due to the fact that pollen also disperses in the z dimension. If the flower layer is thin, the probability that a deposition will result in a pollination event therefore decreases rapidly as each pollen grain disperses out of the flower layer, leading to the variances of X and Y dropping.

For a fitted model, the calculations of these variances and expectations can be carried out as follows. If we assume that plant heights are normally distributed with variance σ_0^2, then $\mathrm{Var}(Z|R=0) = \sigma_0^2$. Since $\mathrm{Var}(Z)$ increases linearly with R, it follows that $Z|R=r$ is normally distributed with an expectation of $\mu_r r$ and a variance of $\sigma_0^2 + \sigma_r^2 r$. The probability that a pollen grain deposited at a distance R from the source result in pollination is then proportional to the product of the probability of deposition and that of pollination given deposition at height z, integrated over all possible values of z,

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{z^2}{2\sigma_0^2}} \frac{1}{\sqrt{2\pi(\sigma_0^2+\sigma_r^2 r)}} e^{-\frac{(z-\mu_r r)^2}{2(\sigma_0^2+\sigma_r^2 r)}} dz \\ &= \frac{1}{\sqrt{2\pi(2\sigma_0^2+\sigma_r^2 r)}} e^{-\frac{(\mu_r r)^2}{2(2\sigma_0^2+\sigma_r^2 r)}}. \end{aligned} \tag{A.4}$$

This expression assumes its maximum value of $\frac{1}{2\sqrt{\pi}\sigma_0}$ when $r = 0$. An acceptance-rejection method for generating a single realization of X, Y corresponding to a pollination event is then as follows:

1. Generate $R \sim \text{Exp}(\lambda_r)$.
2. Generate $Z \sim \text{Normal}(-\mu_r R, \sigma_0^2 + \sigma_r^2 R)$.
3. Generate $U \sim \text{Uniform}(0, 1)$.
4. If $U > \frac{e^{-\frac{(\mu_r R)^2}{2(2\sigma_0^2+\sigma_r^2 R)}}}{\sqrt{1+\frac{\sigma_r^2 R}{2\sigma_0^2}}}$ then go to 1.
5. Let $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$.

The variances and expectations can then be calculated from, say 1000 such realisations.

B Effects of asynchrony and protandry/protogyny

Consider the contribution of plant j to a recipient plant i, and assume that the times t_i and t_j at which the respective plants began flowering have been measured. Assume that plant j distributes its pollen over time according to a normal distribution with a variance σ_m^2 around the mean t_j. Assume also that the times at which the individual stigmas on plant i are receptive to pollination are normally distributed over time with variance σ_f^2 and mean $t_i + \delta$. The parameter δ thus represents the degree of protandry, that is, the delay between the dehiscence of the anthers and the time at which the stigmas become receptive.

It follows from this that the contribution of plant j to the recipient plant i is proportional to the product of plant i's susceptibility to pollination and the amount of pollen emitted from plant j at time t, integrated over time

$$\begin{aligned} h(t_i, t_j) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_f} e^{-\frac{(t-t_i-\delta)^2}{2\sigma_f^2}} \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{(t-t_j)^2}{2\sigma_m^2}}\, dt \\ &= \frac{e^{-\frac{(t_i+\delta-t_j)^2}{2(\sigma_f^2+\sigma_m^2)}}}{\sqrt{2\pi(\sigma_f^2+\sigma_m^2)}}. \end{aligned} \tag{B.1}$$

Note that this implies that the contribution of plant j to plant i is at its maximum when $t_j = t_i + \delta$, whereas plant i contributes most to plant j when $t_j + \delta = t_i$, that is, the interaction between plant i and j is asymmetric unless $t_i = t_j$ or the degree of protandry $\delta = 0$.

Noting that the parameters σ_m and σ_f in (B.1) can be combined into a single parameter, for example the average of the variances $\sigma_{mf}^2 = \frac{1}{2}(\sigma_m^2 + \sigma_f^2)$, and substituting σ_{mf} into (B.1) we get

$$h(t_i, t_j) = \frac{e^{-\frac{(t_i+\delta-t_j)^2}{4\sigma_{mf}^2}}}{\sqrt{4\pi}\sigma_{mf}}. \tag{B.2}$$