

# STOCHASTIC MODELS AND STATISTICAL INFERENCE FOR PLANT POLLEN DISPERSAL

Catherine LAREDO \*, Agnès GRIMAUD \*\*

## ABSTRACT

Modelling pollen dispersal is essential to make predictions of cross-pollination rates in various environmental conditions between plants of a cultivated species. An important tool for studying this problem is the “individual pollen dispersal function” or “kernel dispersal”. Various models for airborne pollen dispersal are developed. These models are based on assumptions about wind directionality, gravity, settling velocity and may integrate other biological or external parameters. Some previous approaches have used Brownian Motions with drift for modelling pollen trajectories. However, models for pollen transport used in aerobiology are often based on the Lagrangian Stochastic approach: velocities of pollen grains satisfy stochastic differential equations or Langevin equations and pollen trajectories are obtained by integrating these velocities. New models based on this approach are introduced. A model where the vertical component is driven by an integrated Ornstein-Uhlenbeck process is studied here. Cross-pollination rates data were obtained from large field experiments of maize using the colour of grains as a phenotypic marker of pollen dispersal. We first studied the various individual dispersal functions associated with these models. Second, a thorough statistical framework was developed in order to estimate and compare their performances on data sets. This framework is quite general and can be used to study many other cross-pollination data. Previous and new models were successively analysed using this framework. This new statistical analysis improved significantly former results which had been obtained on the previous models with other statistical methods. The statistical analyses showed that the performances of Lagrange Stochastic models were good, but not better than the previous mechanistic models analysed using this new statistical framework. These results however might be due to some specific environmental conditions in this experiment. Comparisons with the external parameters were quite good, proving that these models can be used in other environmental conditions. All these results show that mechanistic models are good models for predicting short or medium range pollen dispersal and cross-pollination rates.

*Keywords* : Pollen dispersal, field experiment, cross-pollination rates, maize, airborne pollen, meteorological parameters, deconvolution, parametric inference; quasilikelihood, hypothesis testing; individual pollen dispersal function, mechanistic models, Lagrangian Stochastic models, Langevin equations, hitting times.

---

\* Laboratoire M.I.A., INRA, 78350, Jouy-en-Josas, France et PMA Universités Paris 6-Paris 7, UMR 7599. E-mail: catherine.laredo@jouy.inra.fr

\*\* Laboratoire M.I.A., INRA, 78350, Jouy-en-Josas, France et Equipe Select-INRIA Futurs, Université Paris Sud. E-mail: agnes.grimaud@math.u-psud.fr

## RÉSUMÉ

Il est essentiel de modéliser la dispersion du pollen pour prédire dans différentes conditions environnementales les taux de pollinisations croisées entre plantes d'une même espèce. Pour cela, un outil important est la notion de « fonction individuelle de dispersion de pollen » ou « noyau de dispersion ». Pour les plantes anémophiles, nous proposons ici plusieurs modèles de dispersion de pollen transporté par le vent. Ces modèles reposent sur des hypothèses concernant la direction du vent, la gravité, la « vitesse de décollement » du pollen et peuvent intégrer d'autres paramètres externes ou biologiques. Certains modèles étudiés antérieurement décrivent le transport du pollen par des mouvements browniens avec dérive. Cependant, les modèles de transport de pollen utilisés en aérobiologie utilisent en général l'approche stochastique lagrangienne, dans laquelle les vitesses sont modélisées par des équations différentielles stochastiques ou équations de Langevin. Nous proposons de nouveaux modèles provenant de cette approche. Les trajectoires du pollen sont obtenues en intégrant leurs vitesses. Nous étudions en particulier un modèle où la composante verticale est régie par un processus d'Ornstein-Uhlenbeck intégré. Nous disposons de données de pollinisations croisées issues d'expériences en plein champ de maïs, dans lesquelles la couleur des grains est utilisée comme marqueur phénotypique de la dispersion du pollen. Nous étudions les différentes fonctions de dispersion individuelles associées à ces modèles. Pour estimer et comparer leurs performances sur les données, nous développons un cadre statistique approfondi, qui peut être utilisé pour étudier la plupart des données issues de pollinisations croisées. Ceci nous permet d'analyser successivement anciens et nouveaux modèles. Ce nouveau cadre statistique permet d'améliorer significativement les résultats antérieurs obtenus avec les premiers modèles estimés avec d'autres méthodes. Les performances des modèles stochastiques Lagrangiens sont généralement bonnes, mais elles ne sont pas meilleures que celles obtenues avec les premiers modèles analysés dans le cadre statistique introduit ici. Néanmoins, il se peut que ces résultats proviennent de conditions environnementales particulières liées à ces données expérimentales. Les comparaisons des paramètres estimés avec les paramètres physiques ou externes sont très satisfaisantes. L'ensemble de ces résultats montre que les modèles mécanistes sont de bons modèles pour prédire la dispersion du pollen à des distances courtes ou moyennes, ainsi que les taux de pollinisations croisées.

*Mots-clés:* dispersion du pollen, expériences en plein champ, taux de pollinisations croisées, maïs; pollen transporté par le vent, paramètres météorologiques, déconvolution, statistique paramétrique, quasivraisemblance, tests d'hypothèses, fonction de dispersion individuelle, modèles mécanistes, approche stochastique lagrangienne, équations de Langevin, temps d'atteinte.

## 1. Introduction

Gene dispersal through pollen is an important determinant of genetic structure in plant population. It also affects various aspects of population biology, genetic conservation and breeding. The development of Genetically Modified (G. M.) crops has highlighted the need to assess the risks associated with pollen-mediated gene flow. Quantifying pollen-mediated gene flow is indeed vital for evaluating the environmental impact of G. M. crops. Pollen flow is

a main cause of harvest pollution between G. M. and non G. M. fields (see Rieger *et al.* 2002 for oilseedrape, Devos *et al.* 2005 for maize). Quantitative estimations of gene flows are also valuable for managing escape from fields to feral and wild populations and for recommending isolation distances that minimize cross-pollination among fields. Another interest lies in relation with agronomy for assessing the purity of crops and seeds.

Published data on pollen dispersal experiments show that comparison between dispersal data from different field experiments is hardly possible (Treu and Emberlin 2000). Differences in experiment design, genotypes and environmental conditions have contributed to the wide variation in reported cross-pollination rates of gene flows. These limitations are also due to the strong dependence on the shape and extent of both source and receptor plot (Meagher and Vassiliadis 2003, Klein *et al.* 2006). One way to circumvent the dependence of cross-pollination rates on spatial design is to model pollen dispersal via an individual dispersal function (or dispersal kernel). For pollen, an individual dispersal function is defined as “the probability that a pollen grain dispersed by a plant at point  $(0,0)$  pollinates at any point  $(x,y)$ ” (Klein *et al.* 2003 for pollen, Clark *et al.* 1999 for seed dispersal). It is widely accepted that this function depends less strongly on the spatial design than the function describing the cross-pollination rates at various distances. Consequently, the individual dispersal function is useful for forecasting effective dispersal under various spatial configurations.

Another domain where knowing whole dispersal functions is crucial concerns propagation of invasive species. When modelling the space-time evolution of continuously distributed population, several authors (Clark *et al.* 2001, Bolker and Pacala 1997, Lewis 2000) have stressed that various individual dispersal functions (with possibly identical expectation and variance) may generate different colonization patterns. Perthame and Souganidis (2004) have actually proved that the suitably normalized propagation of an invasive species follows an integro-differential equation comprising the individual dispersal function  $\gamma(x, y)$ ,

$$\frac{\partial}{\partial t}u(t, x, y) = \alpha u + \int_{R^2} \gamma(x - x', y - y')u(t, x', y')dx'dy' + f(t, x, y, u) .$$

Function  $u(t; x, y)$  denotes the population density at time  $t$  in location  $(x, y)$ , parameter  $\alpha > 1$  is the growing rate and  $f$  is a regular function modelling competition and local phenomena. This is an alternate model to reaction-diffusion equations which only take into account the  $\gamma(., .)$  mean and variance (see e.g. Weinberger 1978).

Studies of pollen-mediated gene flow have therefore received much attention in recent decades (see Ellstrand 1992 followed by many others). Several methods have been developed to study this individual dispersal function. In some experiments, this function can be directly observed and data consist of noisy observations of the individual dispersal function. This is accurate for studies that focus on the physical dispersal of pollen grains and use pollen traps to observe a pollen density at various distances from a source (Bateman 1947,

McCartney and Lacey 1991). This is also true when paternity analysis is used to measure the “effective pollen dispersal” in plant populations (Adams *et al.* 1992). This method is based on genetic markers and relies on an exhaustive sampling of the males in the vicinity of the sampled females. This method is powerful but often requires a laborious approach so that it is mostly used for studying pollen from trees (Oddou-Muratorio *et al.* 2005) or for long distance dispersal of oilseedrape (Devaux *et al.* 2005 where paternity analysis uses fields as potential parents). In order to cope with the experimental effort required by paternity analysis, an alternate approach based on two-generation analyses, the TWOGENER analysis, attempts at estimating the extent of pollen movement based on the genetic differentiation among the inferred pollen pools of females samples (Smouse *et al.* 2001). The respective properties of these two approaches to estimate pollen-mediated gene flows in plant populations are investigated in Burczyk and Koralewski (2005).

Unfortunately, direct observations of individual dispersal functions are impracticable for cultivated species such as oilseed rape, maize, sugar beet or wheat. Data consist of cross-pollination rates between fields. In many dispersal studies, pollen dispersal is actually quantified on the basis of the presence of a genetic marker in the progeny of sampled plants. A patch of plants homozygous for a monogenic dominant marker is used as pollen source inside a large patch containing plants homozygous for the recessive allele. Therefore, efficient pollen dispersal is observed whenever the offspring of a sampled plant contains pollination from the marked source (Bateman 1947, Lavigne *et al.* 1998, Nurminiemi *et al.* 1998). The estimation of dispersal functions does no longer require fitting functions to the observed cross-pollination rates data. In these dispersal experiments, two functions are really involved, the backward and the forward pollination functions. The *backward dispersal function* is “the proportion of ovules in recipient plants at a given distance that are fertilized by the source plant”, while the *forward dispersal function* is “the proportion of the source plant pollen that fertilizes ovules at a given distance” (Latta *et al.* 1996). The backward dispersal function describes the pollen cloud composition above a plant and is directly observed from the experiment. The forward dispersal function (called “individual dispersal function” above) is not directly observable since it is impossible to track all the pollen emitted by one plant. Estimating the individual dispersal function hence requires further statistical analyses since retrieving the individual dispersal function from data actually amounts to a non-linear deconvolution problem. Linear deconvolution problems have been largely studied especially in signal or image analysis and many solutions to this problem are available. Non-parametric inference for linear deconvolution has also recently been investigated. A first natural approach for studying pollen dispersal would be a non-parametric approach. However taking pollen competition into account introduces a non-linearity in the deconvolution. To our knowledge, non-parametric inference for non-linear deconvolution problems is a difficult problem which has not been investigated. This is a new difficulty that precludes non-parametric methods. As in Tufto *et al.* (1997), Lavigne *et al.* (1998) and Klein *et al.* (2003), we consider in this paper parametric approaches for studying pollen dispersal.

Models for dispersal functions of wind dispersed particles can roughly be classified in two families (McCartney and Fitt 1985). Originally are the descriptive or empirical parametric families. They have been chosen for their mathematical simplicity and are used to fit the experimental data. Various parametric families depending on the distance from the source have been studied, such as  $r^b e^{-r^\alpha}$  (Austerlitz *et al.* 2004, Klein *et al.* 2006) and variants (Tufto *et al.* 1997 for the bivariate Normal and Weibull distributions). These models allow for investigating how various factors influence the observed cross-pollination rates (spatial design, respective source and recipient field dimensions, tail shape of the individual dispersal function). One weakness of these models is the automatic linking between short and large range behaviour of dispersal functions. In addition, these models are rather descriptive and do not give much insight into how underlying biologically important mechanisms may influence the dispersal amount. These models are also unsuitable for including any physical determinants of the species or any knowledge about the environmental conditions. Finally, effective dispersal depends on many interacting biological and physical phenomena that empirical models are unable to catch. Therefore, predictions using these functions in other situations are debatable. To circumvent this problem, mechanistic models can be used. They are derived from models used in micrometeorology and can include some characteristics of pollen dispersed by wind such as gravity, wind directionality, vegetation density, wind threshold, heterogeneity in fecundity and mating success. They can also be used for pollen dispersed both by wind and insects (Cresswell *et al.* 2005 for oilseedrape). They are built on the scale of a pollen grain by modelling the air flow in which particles are dispersed and the conditions of their emission, transport and deposition. These models involve parameters having a physical or biological interpretation, but they often lead to intricate mathematical expressions with numerous parameters and computationally heavy numerical solutions. A review paper (Aylor *et al.* 2003) integrates in an aerobiological framework all these factors, including pollen transport and successful pollination. Formulation for the transport function differs according to the approaches. One is based on a Gaussian Plume model (Seinfeld and Pandis 1998), another one on a Lagrangian Stochastic model (L. S.). A variant of the L. S. model is the localized-near-field model (Raupach 1989). It is based on the physical processes in the canopy and has been used in Loos *et al.* (2003) for studying two spatially explicit models of maize dispersal. It is generally accepted that the Lagrangian Stochastic is more appropriate than the Gaussian Plume model since it better describes short and medium range behaviour of pollen dispersal (Aylor *et al.* 2003). One explanation might be that only the Lagrangian Stochastic model allows for vertical stochasticity in the trajectory. Involving many parameters often difficult to measure in natural conditions is actually a main drawback of the L. S. approach. In addition, these models have been designed to achieve predictions of dispersal patterns based on measures of physical parameters and not to fit dispersal data.

An intermediate approach has been proposed in Tufto *et al.* (1997), and extended in Klein *et al.* (2003). Only the few major phenomena that account for the dispersal pattern are considered. This allows for building models

simple enough to be fitted to experimental data, but sophisticated enough to contain parameters having a physical meaning. Models in Tufto *et al.* (1997) introduce a wind threshold for emission, models in Klein *et al.* (2003) include the height difference between male and female flowers, mean wind intensity and turbulence. Various pollination times are also investigated. In Klein *et al.* (2003), the fitted model performed reasonably well. Using exponential pollination time was proved to be quite inaccurate in this case, so that other models for pollination times had rather been used. However, the estimated values of the parameters that possessed a physical meaning such as wind intensity were not satisfactory.

Lagrangian Stochastic models are based on the Langevin model (see e.g. Rodean 1996). Velocities of a particle are assumed to satisfy a three-dimensional stochastic diffusion process. The drift and diffusion coefficients are called the Langevin coefficients and are linked to the horizontal and vertical accelerations in a fluid element. Integrating these velocities yields pollen trajectories. An additional term in the vertical component models the settling velocity of pollen grains. In Tufto *et al.* (1997) and Klein *et al.* (2003), a three-dimensional Brownian motion with drift models paths of pollen grains, which is similar to the Gaussian Plume approaches. Lagrangian Stochastic approaches are studied here. Models for pollen trajectories are based on the Langevin equations and a parameter for the settling velocity is added. Models for pollination times already introduced in Klein *et al.* (2003) are kept. A thorough statistical framework is developed in order to estimate and compare the various individual dispersal functions obtained using these different mechanistic models. For this, we reconsider data from a maize pollen dispersal experiment presented in Klein *et al.* (2003). In this experiment, dispersal of a dominant phenotypic marker (blue colored grains) from a patch of plants located at the center of a field of unmarked plants (yellow grains) was monitored. Maize had been chosen as a crop because it is a wind-pollinated outcrossing species and the first transgenic crop species cultivated in Europe. This work is an attempt to introduce environmental conditions in the individual dispersal functions in order to improve the prediction of effective dispersal under various conditions. Since several parameters present in these functions have a physical meaning, these dispersal functions can be used to predict effective dispersal under various spatial designs and other environmental conditions.

This paper is organized as follows. Section 2 describes the experimental conditions, the dispersal data and the meteorological data. The statistical framework and parametric inference are detailed in Section 3. New models are built in Section 4. Estimation results for the different models are compared in Section 5. The last section contains the discussion concerning the statistical methodology and the mechanistic models.

## 2. Experimentation

This experiment has been described in Klein *et al.* (2003).

### 2.1. Maize production for grain

The experimentation was performed during Summer 1998 near Montargis (France). A maize field measuring  $120 \times 120$  m was sown in a production design: 160 rows 0.8 m apart each containing 800 plants 0.15 m apart. A central plot measuring  $20 \times 20$  m was sown with plants producing blue coloured seeds and the rest of the field contained yellow maize of the hybrid variety Adonis. The blue maize was a variety close to Adonis, homozygous for the “blue” allele. The blue colour is coded by the anthocyanin complex, which behaves as a monogenic dominant marker. All plants are homozygous at the loci coding for the seed colour. Checks in the field and on control crosses did not reveal any systematic difference in pollen production and pollen efficiency between plants producing blue or yellow seeds.

The pollen dispersal began on July 18. Both blue and yellow plants flowered almost synchronously: blue maize began blooming on July 19 (male) and July 20 (female). Dispersal lasted 14 days and ended on August 1. The ears were harvested and analysed on 16 October. A total of  $K = 3063$  ears were sampled on a rectangular grid. An amount of 101 rows were sampled (every row for the 72 rows centered on the central plot and every third row elsewhere) and 31 ears on each row (every 4 meters). Sixty-four ears could not be sampled in the West corner of the field. The number of blue grains ( $n_k$ ) on each sampled ear was then determined. The total number  $N$  of grains per ear was considered constant and estimated by counting the total number of grains on 34 randomly chosen ears (mean 394 and standard deviation 65). The proportion of marked pollen grains in the pollen cloud at each sampled plant  $k$  was estimated by dividing the number  $n_k$  of blue grains by the mean number  $N$  of grains per ear.

### 2.2. Meteorological data

We used data for wind direction and intensity collected 10 m high at 3 hours intervals by Meteo France. The meteorological station nearest to the experiment was 70 km West of the field (Orléans). A correction based on a logarithmic model of wind profile (McCartney and Lacey 1991) was used to compute the wind speed at 2 m high (height of maize plants). We then calculated the mean wind speed over the pollination period from wind data between 8 a.m. and 7 p.m., hours during which pollination occurs. In 1999, we placed a meteorological station inside a maize field at the grain production station. This station measured wind speed at a 2 m height. A comparison between data from Meteo France (70 km West of the field) and the local data showed little difference over the 15 days period of dispersal, except for a difference in wind speed due to differences in heights of measures which was well corrected with the logarithmic expression.

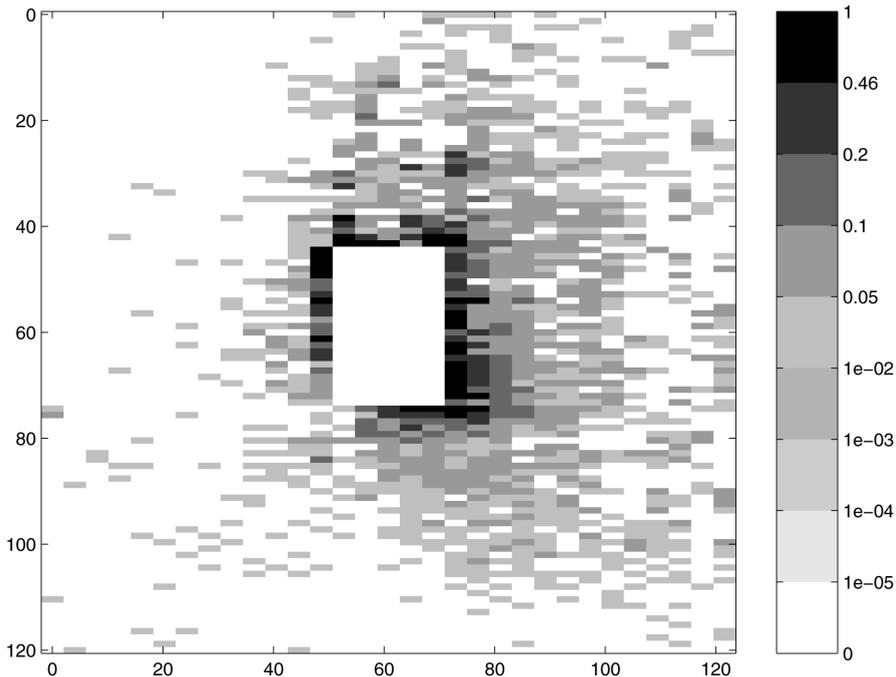


FIG 1. — Experiment in Montargis:  $120 \times 120$  m yellow maize field sown with yellow maize except for the  $20 \times 20$  m central plot sown with blue maize. Observed proportions of blue grains on the 3063 sampled ears (higher densities are brighter).

### 3. Statistical framework and parametric inference

#### 3.1. Statistical model

In the dispersal experiment, a patch of plants homozygous for a dominant marker (blue-coloured grains) is grown in a field of plants homozygous for the absence of marker (yellow-coloured grains). The backward dispersal function  $\mu(x, y)$  is defined as the proportion of pollen from the marked source in the pollen cloud at point  $(x, y)$ . It is the result of the individual dispersal of pollen grains from all the plants, both marked and non-marked. The data consist of the  $K$  counts of blue grains on the ears located at all sampling points  $(x_k, y_k)$ . They are noisy observations of the backward dispersal function. Hence, using this function is more intuitive and most used in biology. We assume that (a) all plants in the field disperse their pollen according to the same dispersal function, (b) both plant types produce the same amount of pollen, (c) both pollens are equally efficient. These assumptions are quite reasonable for this experiment, but there exist situations where they are no longer verified: presence of a discontinuity in the experimental set-up for (a); cross-pollination between G. M. plants and non G. M. ones or heterogeneity

in fecundity for (b); lag in the flowering time between source and recipient plants or heterogeneity in mating success for (c).

Since pollen production in maize is prolific, pollination occurs on a plant at  $(x, y)$  according to the pollen cloud composition above it. Let  $B$  (resp.  $A$ ) denote the set of indexes corresponding to blue (resp. yellow) plants and let  $\{(x_k, y_k), k \in K\}$  be the positions of these plants in the field. The following equation links the backward dispersal  $\mu(x, y)$  and the individual (or forward) dispersal function  $\gamma(x, y)$ ,

$$\mu(x, y) = \frac{\sum_{k \in B} \gamma(x - x_k, y - y_k)}{\sum_{k \in A} \gamma(x - x_k, y - y_k) + \sum_{k \in B} \gamma(x - x_k, y - y_k)} \quad (1)$$

The total number of grains on each ear is assumed constant equal to  $N$ . Denote by  $N_k$  the random variable describing the number of blue grains on an ear located at  $(x_k, y_k)$ , and  $n_k$  its realization. The  $K$  random variables  $N_k$  are noisy observations of  $(\mu(x_k, y_k), k \in K)$ . A first natural assumption for the  $(N_k, k \in \{1, \dots, K\})$  is thus the Binomial assumption:  $N_k \sim \text{Bin}(N, \mu(x_k, y_k))$ . Using now that pollen is overabundant, pollination at  $(x_k, y_k)$  does not depend on pollination at neighbouring locations. Hence, the random variables  $(N_k)$  can be assumed independent and the expression for the likelihood is,

$$L(\gamma(\cdot, \cdot); n_1, \dots, n_K) = \prod_{k=1}^K C_N^{n_k} \mu(x_k, y_k)^{n_k} (1 - \mu(x_k, y_k))^{N - n_k} . \quad (2)$$

Recovering  $\gamma(x, y)$  from the observations  $(n_k, k \in K)$  is a non-linear deconvolution problem which is difficult to solve in a non-parametric statistical framework. Therefore, we propose to use parametric families for modelling individual dispersal functions  $(\gamma_\theta(x, y), \theta \in \Theta, (x, y) \in \mathbb{R}^2)$ . This leads to associated parametric families for  $\mu, \mu_\theta(x, y)$ . One possibility is to introduce parameters via the descriptive or empirical families described in the introduction. Another possibility is to derive parametric families from mechanistic approaches (Section 4).

### 3.2. Quasilikelihood methods and parametric Inference

Statistical estimation was performed in Klein *et al.* (2003) using either least square methods or the likelihood (2) associated with the Binomial model. Studying the residuals of the statistical analysis suggests that the Binomial model is too specific a model. Many reasons can explain this. First, it is an approximate statistical model and the Binomial likelihood is not sufficiently robust to deviations from this model. Second, eventual correlations between genotypes on an ear are ignored by the Binomial model. Third, data only consisted of the number of blue grains instead of the joint observation (number of blue grains, total number of grains) on the ears. We just assumed that all the ears possessed the same total number  $N$  of grains. At last, presence of unidentified covariates might be important. For all these reasons, we prefer to

use here the framework of quasilielihood methods, where the precise form of the  $N_k$  distribution can be relaxed. Let us set

$$N_k = N \mu_\theta(x_k, y_k) + \varepsilon_k, \quad k \in K. \quad (3)$$

We just assume now that the random variables  $(\varepsilon_k)$  are independent and satisfy

$$E(\varepsilon_k) = 0; \quad \text{Var}(\varepsilon_k) = N \sigma^2 v_\theta(x_k, y_k), \quad (4)$$

$$\text{with } \sigma^2 = (1 + d(N - 1)) \text{ and } v_\theta(x_k, y_k) = v(\mu_\theta(x_k, y_k)). \quad (5)$$

Parameter  $d$  is an overdispersion parameter (Collet 1991), which partly takes into account correlation between genotypes at  $(x_k, y_k)$ ;  $v(\cdot, \cdot)$  is a function depending on  $(x_k, y_k)$  via  $\mu(\cdot)$ . Note that this model includes the Binomial model setting  $d = 0$  and  $v(\mu) = \mu(1 - \mu)$ .

*Remark.* — Considering a prescribed number  $N$  of grains on each ear might have biased the analysis. We could have built a model taking into account these variations which are not negligible (see Section 2.1). We chose not to model this since these numbers will be observed in future experiments.

Assume that parameter  $\theta$  has dimension  $p$ . The  $p$  quasilielihood equations associated with the statistical model defined in (3), (4), (5) are

$$U_i(\theta; n_1, n_2, \dots, n_K) = \sum_{k=1}^K \frac{\partial \mu}{\partial \theta_i}(\theta; x_k, y_k) \frac{n_k - N \mu_\theta(x_k, y_k)}{v(\mu_\theta(x_k, y_k))}, \quad i = 1, \dots, p. \quad (6)$$

Heuristically, quasilielihood equations stand for the score functions  $\frac{\partial \text{Log} L}{\partial \theta_i}$  obtained when maximizing the loglikelihood. Usually, an overdispersion parameter is added and function  $v(\cdot)$  is the one associated with the Binomial model. However, modelling the variance of the  $(\varepsilon_k)$  in such a way has a drawback here because  $n_k/N$  is less than 0.2 for almost all the observations. As a consequence, the quasilielihood equations (6) do not penalize as expected. This led us to propose another form for  $\text{Var}(\varepsilon_k)$ , which results in adding a new parameter  $b$  for modelling  $v(\cdot)$

$$v(x_k, y_k) = v(b; \mu_\theta(x_k, y_k)). \quad (7)$$

The associated quasilielihood equation is

$$U_{p+1}(\theta, b) = \sum_{k=1}^K \frac{\partial v}{\partial b}(b; \mu_\theta(x_k, y_k)) \frac{(n_k - N \mu_\theta(x_k, y_k))^2 - N \sigma^2 v(b; \mu_\theta(x_k, y_k))}{v^2(b; \mu_\theta(x_k, y_k))}. \quad (8)$$

The  $p$  quasilielihood equations are obtained substituting  $v(\mu_\theta(x, y))$  by  $v(b; \mu_\theta(x, y))$  and  $U_i(\theta; n_1, \dots, n_K)$  by  $U_i(\theta, b; n_1, \dots, n_K)$  in (6). The estimators  $(\hat{\theta}_K, \hat{b}_K)$  are solutions of the  $(p + 1)$  equations,

$$U_i(\theta, b; n_1, \dots, n_K) = 0, \quad i = 1, \dots, p + 1. \quad (9)$$

The estimator  $\hat{\sigma}_K^2$  is defined using the residual variance. Various forms for  $v$  have been tested. We obtained that setting  $v(b, \mu) = b + \mu$  was a good compromise: slope at 0 is kept and too much weight is not assigned to almost null observations, which is consistent with the data. The asymptotic properties of quasilielihood estimators are well known (McCullagh and Nelder 1989, Huet *et al.* 1996). They are consistent and asymptotically normal at the usual rate  $\sqrt{K}$ .

Details for this section are given in Grimaud (2005).

## 4. Building new models

Successful pollination depends on various elements: rate of production and release of pollen grains, atmospheric transport, losses of pollen in the air due to deposition, survival of pollen during the flight, deposition and competition on receptive silks and fertilization. Each stage can be described with varying detail. In order to provide a framework for evaluating current data and posing additional questions, models are necessarily built using many simplified and idealized assumptions.

A two-step modelling approach is used here. A first model describes the transport of pollen grains and a second one pollination times. A common framework is taken in order to hang together transport and effective pollination of pollen grains. Trajectories of pollen grains ( $P_t = (X_t, Y_t, Z_t), t \geq 0$ ) are random paths, and successful pollination times  $T_F$  are random times such that  $T_F$  and  $(X_t, Y_t)$  are independent. Within this framework, the random variable  $(X_{T_F}, Y_{T_F})$  possesses a density on  $\mathbb{R}^2$ , so that we can give the following definition for the individual dispersal function  $\gamma(x, y)$ .

DEFINITION 1. — *The individual dispersal function  $\gamma(x, y)$  is the joint probability density on  $\mathbb{R}^2$  of the random variable  $(X_{T_F}, Y_{T_F})$ .*

Similar ways of defining  $\gamma(x, y)$  have been used in Tufto *et al.* (1997) with exponential pollination times and Klein *et al.* (2003) with other pollination times.

### 4.1. Modelling transport of pollen grains

Dispersal behaviour of small particles such as pollen grains has been studied for many decades, and models with various degrees of complexity have been developed (McCartney *et al.* 1991, Squire *et al.* 1997). According to Aylor *et al.* (2003), improved aerobiological models are needed for getting a better understanding of wind-pollinated species such as maize. In micrometeorology, pollen trajectories are often obtained using the Lagrangian Stochastic approach, which can incorporate the relevant physical characteristics of pollen. This study intends to take better into account the physical characteristics of maize pollen together with the environmental conditions, in order to improve predictions of effective dispersal under various conditions.

#### 4.1.1. Previous models

The simplest stochastic model for modelling trajectories ( $P_t$ ) of pollen grains is a three-dimensional Brownian Motion with drift,

##### Model 1

$$\begin{cases} dX_t = f_x dt + \tau dB_t^1, & X_0 = 0, \\ dY_t = f_y dt + \tau dB_t^2, & Y_0 = 0, \\ dZ_t = f_z dt + \tau_z dB_t^3, & Z_0 = 0. \end{cases} \quad (10)$$

The three Brownian motions ( $B_t^i$ ) related to each coordinate are assumed independent; the expressions  $dX_t$ ,  $dY_t$  and  $dZ_t$  are the infinitesimal displacements during time  $dt$ ;  $f_x$  and  $f_y$  are horizontal drift parameters and model horizontal mean wind velocities and  $f_z$  is the negative vertical drift due to gravity. Parameters  $\tau$  and  $\tau_z$  are the diffusion coefficients related to horizontal and vertical atmospheric turbulence. The terms  $dB_t^i$  describe infinitesimal random perturbations affecting the trajectory during  $dt$ . These models have been investigated in Tufto *et al.* (1997) and Klein *et al.* (2003) coupled with pollination times. This approach for pollen trajectories is actually close to the Gaussian Plume model since  $(X_t, Y_t)$  has the same marginal distribution.

#### 4.1.2 Lagrangian Stochastic models

Aylor *et al.* (2003) proposed to use Lagrangian Stochastic models for wind dispersed pollen trajectories. It relies on two Langevin equations used to increment the wind and vertical air velocities ( $U, V$ ) surrounding a pollen grain. These velocity increments are integrated to yield particle displacements in the along wind direction ( $X$ ) and the vertical direction ( $Z$ ) over a time step  $dt$ . Each segment of the trajectory is affected by a deterministic part and a random part of the local fluid flow field. Let  $a(U)$  (resp.  $\tilde{a}(W)$ ) denote the horizontal (resp. vertical) acceleration of the fluid element, then

$$\begin{cases} dU_t = a(t, U_t)dt + b(t, U_t)dB_t^1 \\ dW_t = \tilde{a}(t, W_t)dt + \tilde{b}(t, W_t)dB_t^2. \end{cases} \quad (11)$$

Integrating the velocities  $U_t$  and  $W_t$  yields the positions  $X_t$  along the wind direction and  $Z_t$  along the vertical direction,

$$\begin{cases} X_t = \int_0^t U_s ds, \\ Z_t = \int_0^t (W_s - v_S) ds. \end{cases} \quad (12)$$

The Langevin coefficients  $a(\cdot), b(\cdot), \tilde{a}(\cdot), \tilde{b}(\cdot)$  depend on time  $t$  and on the velocities at time  $t$ ,  $U_t$  and  $W_t$ . Thus,  $(U_t)$  and  $(W_t)$  are two independent diffusion processes with drift coefficient  $a(\cdot)$  (resp.  $\tilde{a}(\cdot)$ ) and diffusion coefficient  $b(\cdot)$  (resp.  $\tilde{b}(\cdot)$ ). The vertical diffusion ( $W_t$ ) is assumed centered (i.e.  $\forall t, E(W_t) = 0$ ) and the additional term  $v_S$  in the definition of  $Z_t$  is the settling velocity of pollen grains in still air. The order of magnitude of  $v_S$  is

0.2 – 0.3  $ms^{-1}$ . In this set-up, trajectories  $(P_t)$  are integrated diffusion processes. Let us stress that dispersal is described in these models according to one direction (the dominant wind direction) since  $U_t$  is one dimensional. We study below a model which derives from the Langevin equations but uses a two-dimensional approach for the horizontal components.

**Model 2:** Model 1 is modified by means of a more precise description of its vertical component using the Langevin equation. The vertical position  $Z_t$  is obtained as an integrated Ornstein-Uhlenbeck process.

$$\begin{cases} Z_t = \int_0^t W_s ds \\ dW_t = (d_z - \beta W_t)dt + \eta dB_t^3, \end{cases} \quad (13)$$

while

$$\begin{cases} dX_t = f_x dt + \tau dB_t^1, & X_0 = 0, \\ dY_t = f_y dt + \tau dB_t^2, & Y_0 = 0. \end{cases} \quad (14)$$

Here  $\beta > 0$ ,  $\eta > 0$  and  $d_z < 0$ . The vertical stochasticity of the trajectory is more precisely described than in Model 1. If  $d_z = 0$ ,  $V_t$  is a positive recurrent diffusion process with stationary distribution  $\mathcal{N}(0, \frac{\eta^2}{2\beta})$ . The additional term  $d_z$  models the vertical drift downside.

#### 4.2. Modelling pollination times

This second step may be described with varying detail since it comprises various functions such as rates of deposition of viable grains, fertilization efficiency, competition rates. To keep models tractable, a simple framework is used here. Pollination times are positive random variables  $T_F$  independent of  $(X_t, Y_t)$ . Various models for pollination times are related to a rather general family of distributions, the Generalized Inverse Gaussian distributions (see e.g. Jorgensen 1982). A distribution is GIG  $(\alpha, \rho, \eta)$  with  $\rho > 0$ ,  $\eta > 0$  if its density on  $\mathbb{R}^+$  is

$$f_{GIG}(\alpha, \rho, \eta; t) = \frac{1}{I(\alpha, \rho, \eta)} t^{-\alpha} e^{-\rho t - \frac{\eta}{t}} \mathbb{I}_{t>0}. \quad (15)$$

The term  $I(\alpha, \rho, \eta)$  is a normalizing constant with a precise form, since it is a special function (Abramovitz and Stegun 1972). These distributions generalize the classical Inverse Gaussian distribution  $IG(\mu, \lambda)$

$$IG(\mu, \lambda)(t) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right) \mathbb{I}_{t>0} = GIG\left(\frac{3}{2}, \frac{\lambda}{2\mu^2}, \frac{\lambda}{2}\right) \quad (16)$$

The link between these two definitions is  $\alpha = 3/2$ ,  $\lambda = 2\eta$ ,  $\mu^2 = \frac{\eta}{\rho}$ .

Pollen is emitted by male flowers and received by female flowers which are lower. Let  $h < 0$  denote the difference in heights between female and male flowers. We consider below several models for pollination times.

**Exponential time: Model BR-IG.** This model has been proposed when vegetation is overabundant (Tufto *et al.* 1997, Klein *et al.* 2003). Given an exponential distribution  $T_e = \mathcal{E}(\lambda)$  independent of  $(P_t)$ , the pollination time  $T_F$  is obtained as the conditional distribution of  $T_e$  given  $Z_{T_e} = h$ . Then  $T_F$  is distributed according to the Generalized Inverse Gaussian distribution

$$GIG\left(\frac{1}{2}, \lambda + \frac{f_z^2}{2\tau_z}, \frac{h^2}{2\tau_z^2}\right). \quad (17)$$

**First hitting time of level h: BR-IG.** Assume that  $Z_0 = 0$  and  $h < 0$ . The pollination time is defined as  $T_F = T_h = \inf\{t > 0, Z_t = h\}$ . If  $(Z_t)$  is a Brownian motion with negative drift ( $f_z < 0$ ),  $T_F$  follows an Inverse Gaussian distribution,

$$IG\left(\frac{h}{f_z}, \frac{h^2}{\tau_z^2}\right) = GIG\left(\frac{3}{2}, \frac{f_z^2}{2\tau_z^2}, \frac{h^2}{2\tau_z^2}\right). \quad (18)$$

**Generalized Model: BR-GIG.** Pollination times  $T_F$  follow  $GIG(\alpha, \rho, \eta)$  distributions as defined in (15).

These models have been proposed in Klein *et al.* (2003). Models BR-EX ( $\alpha = \frac{1}{2}$ ) and BR-IG ( $\alpha = \frac{3}{2}$ ) are specific cases of the BR-GIG model ( $\alpha$  arbitrary).

### 4.3. Computation of the individual dispersion function $\gamma_\theta(x, y)$

We use a result on Brownian motion stopped at  $GIG$  random times (Barndorff-Nielsen 1997 or Jorgensen 1982). This requires to define the Generalized Hyperbolic Distributions (GHD). The Generalized Hyperbolic Distribution with parameters  $(\chi, \psi, \alpha)$  is the distribution on the positive half-line with density function for  $\chi > 0, \psi > 0, \alpha \in R$ ,

$$\left\{2\left(\frac{\chi}{\psi}\right)^{\alpha/2} K_\alpha(\sqrt{\chi\psi})^{-1}\right\} x^{\alpha-1} \times \exp\left\{-\frac{1}{2}(\chi x^{-1} + \psi x)\right\} \quad (19)$$

The notation  $\mathcal{K}_\alpha(s)$  is the modified Bessel function of the third kind (Abramovitz and Stegun 1972). We use in the sequel multivariate extensions of these distributions.

PROPOSITION 1 (Barndorff-Nielsen, 1997). — *Assume that  $(X_t, Y_t)$  is a Brownian Motion with drift on  $\mathbb{R}^2$  (see (10)), and that  $T_F$  is distributed according to a  $GIG(\alpha, \rho, \eta)$  independent of  $(X_t, Y_t)$ . Then the distribution of  $(X_{T_F}, Y_{T_F})$  is a GHD on  $\mathbb{R}^2$  with density*

$$f_{GHD}(\theta; x, y) = \frac{\nu^{1-\alpha} \delta^2(p/q(x, y))^{\frac{\alpha}{2}} \mathcal{K}_\alpha(\sqrt{pq(x, y)})}{2\pi \mathcal{K}_{1-\alpha}(\nu)} - \left(\frac{f_x x + f_y y}{\tau^2}\right), \text{ where} \quad (20)$$

$$\nu = 2(\rho\eta)^{1/2}, \quad \delta = \frac{1}{\tau(2\eta)^{1/2}}, \quad p = 2\eta \left( 2\rho + \frac{f_x^2}{\tau^2} + \frac{f_y^2}{\tau^2} \right), \quad q(x, y) = 1 + \delta^2(x^2 + y^2). \quad (21)$$

Recall that parameter  $h$  is negative (a pollen grain starts from a male flower -height 0- and stops on lower female flowers),  $f_z$  is also negative; other parameters are positive. According to pollination times, we get several parametric distributions for  $\gamma(\cdot, \cdot)$ .

**Model BR-EX.** The distribution of  $T_F$  is  $\mathcal{E}(\lambda)$  and  $\gamma_\theta(x, y)$  is the GHD associated with  $\alpha = 1/2$ ,  $\rho = \lambda + \frac{f_z^2}{2\tau_z}$ , and  $\eta = \frac{h^2}{2\tau_z^2}$ .

**Model BR-IG.** The distribution of  $T_F$  is the distribution of the first hitting time of level  $h$ . Hence  $\alpha = 3/2$ ,  $T_F$  follows an Inverse Gaussian distribution, and  $\gamma_\theta(x, y)$  is a Normal Inverse Gaussian distribution (NIG) which is a specific case of GHD distributions

$$\gamma_\theta(x, y) = \frac{\delta^2 e^{\lambda z} (q(x, y)^{-1/2} + p^{1/2})}{2\pi q(x, y)} e^{-\sqrt{pq(x, y)}} e^{\delta(\lambda_x x + \lambda_y y)} \quad \text{with} \quad (22)$$

$$\lambda_z = \nu = \frac{f_z h}{\tau_z^2}, \quad \delta = \frac{\tau_z}{\tau |h|}, \quad \lambda_x = \frac{f_x h}{\tau \tau_z}, \quad \lambda_y = \frac{f_y h}{\tau \tau_z}. \quad (23)$$

**Model BR-GIG.**  $T_F$  follows a GIG( $\alpha, \rho, \eta$ ) with  $\alpha$  arbitrary and  $\gamma_\theta(x, y)$  is the GHD distribution defined in (20) with parameters

$$\nu = 2\sqrt{\rho\eta}, \quad \delta = \frac{1}{\tau\sqrt{2\eta}}, \quad \lambda_x = \frac{f_x}{\delta\tau^2}, \quad \lambda_y = \frac{f_y}{\delta\tau^2}. \quad (24)$$

These models for pollen dispersal have been used in Klein *et al.* (2003), Grimaud and Larédo (2006).

**Model BR-OU.** Paths of pollen grains are described in equations (13) and (14). The vertical position ( $Z_t$ ) is obtained by integrating  $W_t$  defined in (13). Pollination times are defined as  $T_h = \inf(t, Z_t \leq h)$ . Four parameters are now present in this modelling:  $h < 0$ ,  $\beta > 0$ ,  $\eta > 0$ ,  $d_z < 0$  (parameters in  $(W_t)$ ). Set  $b = \frac{\beta h}{\eta_z} + \frac{d_z}{\eta_z \beta}$ , and  $c = -\frac{d_z}{\eta_z \beta}$  and define three functions associated with these parameters,

$$g_\theta(t) = \frac{1}{2}(2\beta t + 4e^{-\beta t} - 3 - e^{-2\beta t}), \quad f_\theta(t) = b + c(\beta t + e^{-\beta t}), \quad a_\theta(t) = f_\theta \circ g_\theta^{-1}(t).$$

The following results holds.

**PROPOSITION 2.** — *The hitting time  $T_h$  of  $(Z_t)$  is equal in distribution to  $g_\theta^{-1}(\nu)$ , where  $\nu$  is the first hitting time of a standard Brownian motion starting from 0 through the curved boundary  $a_\theta(t)$ .*

The proof is given in Grimaud and Larédo (2006) and relies on the computation of  $T_h$  using a random time change. The vertical component  $Z_t$  can be split into two terms:  $Z_t = f_\theta(t) + M_t$  where  $(M_t)$  is a martingale satisfying  $\langle M \rangle_t = g_\theta(t)$ . It follows that  $T_h = g_\theta^{-1}(\nu)$  with  $\nu = \inf\{s > 0, \tilde{B}_s \geq a_\theta(s)\}$ , where  $(\tilde{B}_s)$  is a Brownian motion and  $a_\theta(\cdot)$  is the explicit function defined above. Therefore an approximation of the distribution of  $\nu$  is obtained using a result due to Durbin (1992) for the first passage of a curved boundary:

$$r_\theta(t) = \frac{1}{\sqrt{2\pi t}} \left( \frac{a_\theta(t)}{t} - a'_\theta(t) \right) \exp \left( -\frac{a_\theta(t)^2}{2t} \right) \quad (25)$$

Studying precisely  $a_\theta$  leads to quite good approximations of  $r_\theta(t)$ . Using the definition of Inverse Gaussian distributions given in (16), we obtain

COROLLARY 1. — Set  $\mu_1 = \frac{2\beta^2 h + d_z}{\beta d_z}$ ,  $\mu_2 = \frac{2\beta^2 h - d_z}{\beta d_z}$  and  $\lambda = \frac{h^2}{\tau_z^2}$ . Clearly  $\mu_1 > \mu_2$  and the following holds for the distribution  $\phi_\theta(t)$  of  $T_h$ ,

$$IG(\mu_1, \lambda)(t) \leq \phi_\theta(t) \leq IG(\mu_2, \lambda).$$

The proof uses the definition of  $IG(\mu, \lambda)$  as the the first hitting time distribution for a standard Brownian motion starting from 0 of a line  $\sqrt{2\lambda} - \frac{\sqrt{2\lambda}}{\mu}t$ . Hence, the individual dispersal function satisfies  $\gamma_\theta(x, y) \simeq NIG(\theta; x, y)$ , where  $\theta$  is obtained substituting  $f_z$  (Model BR-IG) by  $\frac{d_z}{\beta}$  (Model BR-OU) and  $\tau_z^2$  (BR-IG) by  $\frac{\eta_z^2}{\beta^2}$  (BR-OU). Note that this result agrees with the rough approximation deriving from  $E(Z_t) = f_z t$ ,  $\text{Var}(Z_t) = \tau_z^2 t$  in the BR-IG model and  $E(Z_t) = \frac{d_z}{\beta} t$ ,  $\text{Var}(Z_t) = \frac{\eta_z^2}{\beta^2} (t + \frac{\beta}{4} t^2)$  in the BR-OU model.

Biologists are most concerned with the tail behaviour of dispersal kernels. Ignoring anisotropy or studying the tail in the dominant wind direction, we get

COROLLARY 2. — Set  $r = (x^2 + y^2)^{1/2}$ . Then, for  $r$  large enough, the tail behaviour of  $\gamma(\cdot, \cdot)$  in the dominant wind direction satisfies:  $\gamma(r) \sim C r^{-(\alpha+1/2)} e^{-br}$  with  $b > 0$  and  $\alpha > -1/2$ .

This is easily obtained using some classical approximations for Bessel functions. The tail behaviour of  $\gamma$  is therefore quite similar for all models.

*Remark.* — Several published papers in Ecology argue that kernel dispersals have to possess fatter tails than these ones in order to describe what is observed in landscapes (Oddou-Muratorio *et al.* 2005, Devaux *et al.* 2005). Models are built here to fit short and medium range dispersal, they do not describe long-range dispersal.

## 5. Statistical analyses for individual dispersal functions

Various models (BR-EX, BR-IG, BR-GIG, BR-OU) associated with different pollen grain trajectories and pollination times were studied. Parameters of these different models were estimated from data obtained in the experiment described in Section 2.

### 5.1. Estimation results

Three different statistical methods were used: Binomial likelihood (BIN), quasilielihood with overdispersion (QUASI) and linear variance quasilielihood with overdispersion (LIN). The last two methods differ in the modelling of errors variance. Since  $\sigma^2 = (1 + d(N - 1))$ , the expression of the last two methods is

$$\text{QUASI : Var } \varepsilon_k = N \sigma^2 \mu(x_k, y_k) (1 - \mu(x_k, y_k)) \quad (26)$$

$$\text{LIN : Var } \varepsilon_k = N \sigma^2 (b + \mu(x_k, y_k)) \quad (27)$$

Results are detailed in Tables 1, 2 and 3. Model BR-EX led to results that were always bad for any method. This model is clearly unsuitable for this experimental data set and results are not presented here. Table 1 displays parameter estimations for Model BR-IG and the three statistical methods, Table 2 parameter estimation for Model BR-GIG and the three methods. Results for Model BR-OU are shown in Table 3. A more complete presentation of all the results obtained using various models is available in Grimaud (2005).

Drawings of individual dispersal functions are given in Figures 2 and 3. Figure 2 displays the three individual dispersal functions obtained estimating the parameters of the BR-IG, BR-GIG and BR-OU models using the linear variance quasilielihood. In Figure 2 (a) these dispersal functions are given according to the dominant wind direction, in Figure 2 (b) they are drawn according to the orthogonal dominant wind direction. Figures 3 (a) and (b) display the individual dispersal function obtained for one model and the three statistical methods according to the dominant wind direction for (a) and the orthogonal one for (b).

For all models, we get better standard errors with ‘‘Binomial +overdispersion’’ type variance than with the Binomial variance. This confirms that the Binomial assumption is too stringent and does not lead to good results. Moreover, a better description for  $v(\cdot)$ , (linear variance model), yielded better residuals. The main wind direction (given by the vector  $(\lambda_x, \lambda_y)$ ) was consistently estimated with all methods and different models. Parameters for models BR-IG and BR-GIG differed according to the statistical method. For the BR-IG model, this led to individual dispersal functions that differed from the ones obtained in Klein *et al.* (2003) (see Figures 3 (a)-(b)). The individual dispersal functions of Klein *et al.* (2003) are very sharp at the origin. Elsewhere, they are always below the two other ones (up to ten meters from the source). Comparing now the three individual dispersal functions of the BR-IG, BR-GIG, BR-OU models, heavier tails were obtained by the BR-OU

in the dominant wind direction, and by the BR-IG for the orthogonal one. The BR-OU dispersion is sharper than the two other ones at the origin. The three estimated individual dispersal functions (BR-IG, BR-GIG, BR-OU) obtained using quasilielihood coupled with linear variance are presented in Figure 2. There are differences when looking in the dominant wind direction (Figure 2-(a)). The function associated with the BR-GIG model has the thinnest tail, while the dispersal function associated with BR-OU presents the heaviest tail with respect to the dominant wind direction, and the thinnest tail in the opposite direction. It also decreases faster than the other dispersal functions at short distances. The BR-IG dispersal function decreases less rapidly near the origin and has tails almost comparable with the BR-OU dispersal function.

## 5.2. Testing hypothesis

The main interest in introducing the GIG family of distributions for pollination times is the ability both to estimate  $\alpha$  and to test with accuracy the assumption “ $\alpha = 3/2$ ” against “ $\alpha \neq 3/2$ ”, because we have nested statistical models. Indeed, the case “ $\alpha = 3/2$ ” is particularly interesting since it leads to easier computations for  $\gamma_\theta$ .

The estimation of parameter  $\alpha$  in the model corresponding to the GHD ( $\alpha$ ) (model BR-GIG) is  $\hat{\alpha} = 1.4$  using the linear variance quasilielihood method. Now, model BR-IG corresponds to  $\alpha = 3/2$ . It is nested in Model BR-GIG, and we tested the hypothesis “ $H_0 = 3/2$ ” against “ $H_1 \neq 3/2$ ” (the pollination time follows a  $GIG(\alpha, \rho, \eta)$ , with  $\alpha$  unknown). We used likelihood and quasilielihood test statistics for heteroscedastic non linear models (Huet *et al.* 1996). This statistics  $S_K$  converges as  $K \rightarrow +\infty$  to a  $\chi^2(1)$ . The quantile  $q_{0.05}$  at level 5 % is equal to 3.84. Results for this test were different according to the statistical method. For the Binomial likelihood,  $S_K = 5.4$ . Hence, “ $H_0$ ” is rejected at level 5%. The quasilielihood method coupled with overdispersion and linear variance leads to  $S_K = 2.26$ , so “ $H_0$ ” is accepted at level 5%. Hence, we inferred that modelling pollination times as hitting times of level  $h$  by the vertical component of pollen trajectories was quite satisfactory.

## 5.3. Residuals and selection criteria

The standardized residuals associated with the statistical analysis were studied. Let  $\hat{\theta}$  denote an estimator of the true parameter value  $\theta_0$  and consider  $\mu(\hat{\theta}, x, y)$  the associated backward dispersal function. Residuals are then defined by

$$R_k = \frac{n_k - N\mu(\hat{\theta}, x_k, y_k)}{\sqrt{N\sigma^2 v(\mu(\hat{\theta}, x_k, y_k))}}$$

Displaying these residuals  $R_k$  on the experimental field shows that better results are obtained using the NIG distribution for  $\gamma_\theta(x, y)$  with parameter  $\theta$  estimated by the linear variance quasilielihood. Let us now compare the Model BR-OU with the first three models. We are no longer in a nested framework and have to use selection criteria. The Akaike criterion cannot

be used since we do not know the likelihood (the  $(N_k)$  distributions are not specified). The Mallows  $C_p$  criterion cannot be used either since the errors variances have to be constant in this case. However, Hurvich and Tsai (1995) have proposed a criterion for quasilielihoods which is a modification of the Akaike criterion,

$$AIC_c = \log(\sigma^2) + \frac{2(p+1)}{n-p-2}$$

The number of observations is  $n$  ( $n = 3063$  here) and parameter  $\theta$  has dimension  $p$ . The  $AIC_c$  score is slightly lower for the BR-IG model (5.99) than for the BR-OU model (6.04) using Binomial+overdispersion method. For the linear variance quasilielihood, the  $AIC_c$  score is 3.23 for the BR-IG model and 3.94 for the BR-OU model.

Joining all the results, we have finally obtained that the model that fitted best these experimental data was the BR-IG model, and that we had rather use a quasilielihood method with linear variance modelling and overdispersion for analysing this data set. The model derived from a Lagrangian Stochastic approach did not lead to better results in this case. A possible explanation is as follows. In this experiment, the Lagrangian Stochastic model we have used actually results in a similar dispersal function to the one obtained with the BR-IG model. This is not an intrinsic feature of the L. S. model and it is due here to the precise values of the estimated parameters. Other environmental conditions as for instance stronger winds might lead to other conclusions. This clearly has to be investigated on other experiments.

TABLE 1. — Parameters estimations for Model BR-IG and the three statistical methods. Pollen trajectories  $(X_t, Y_t, Z_t)$  Brownian motion with drift; pollination time  $T_F \sim \text{IG}$ .

Parameters	Quasilielihood		Linear Quasilielihood		Bin Likelihood	
	Estimation	Std error	Estimation	Std error	Estimation	Std error
$\delta$	0.5176	0.0837	0.5177	0.0225	0.499	0.024
$\lambda_x$	- 0.0056	0.0509	-0.0096	0.0065	-0.007	0.013
$\lambda_y$	0.1808	0.0244	0.1914	0.0143	0.165	0.018
$\lambda_z$	0.0561	0.0835	0.0669	0.0204	0.027	0.128
$a$	-	-	$1.175 \cdot 10^{-5}$	$3.589 \cdot 10^{-7}$	-	-
$\sigma^2$	145.9		9.261		-	
	[138.7 ; 156.7]		[8.801 ; 9.755]		-	
$AIC_c$	5.986		3.2299		-	

TABLE 2. — Parameters estimations for Model BR-GIG and the three statistical methods. Pollen trajectories  $(X_t, Y_t, Z_t)$  Brownian motion with drift; pollination time  $T_F \sim \text{GIG}$ .

Parameters	Quasilikelihood		Linear Quasilikelihood		Bin Likelihood	
	Estimation	Std error	Estimation	Std error	Estimation	Std error
$\delta$	0.5985	0.2334	0.5850	0.0571	0.500	0.026
$\lambda_x$	-0.0046	0.0217	-0.0082	0.0057	-0.007	0.014
$\lambda_y$	0.1555	0.0746	0.1683	0.0205	0.165	0.019
$\lambda_z$	0.0799	0.0573	0.0853	0.0151	0.047	0.087
$\alpha$	1.3986	0.1804	1.4133	0.0461	1.49	0.150
$a$	-	-	$1.166 \cdot 10^{-5}$	$3.534 \cdot 10^{-7}$	-	-
$\sigma^2$	155.6 [147.9 ; 163.8]		9.263 [8.808 ; 9.752]		-	
$AICc$	6.051		3.2308		-	

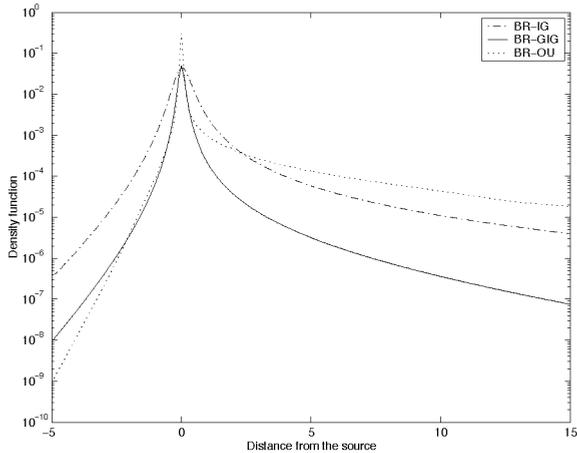
TABLE 3. — Parameters estimations for Model BR-OU. Pollen trajectories  $(X_t, Y_t)$  Brownian motion with drift,  $(Z_t)$  modelled via its velocity (Langevin Stochastic model); pollination time  $T_F$  hitting time of female flowers height  $h$ .

Parameters	Quasilikelihood		Linear Quasilikelihood	
	Estimation	Std error	Estimation	Std error
$r_x$	-0.1822	0.3221	-0.1903	0.1010
$r_y$	1.4359	0.4796	1.3340	0.1620
$\lambda$	0.1571	0.0179	0.1820	0.0078
$b_z$	0.4683	0.0532	0.4417	0.0163
$c_z$	0.1208	0.0647	0.1299	0.0216
$a$	-	-	$8.948 \cdot 10^{-6}$	$4.803 \cdot 10^{-8}$
$\sigma^2$	153.6 [ 146.03; 161.76]		18.81 [17.88 ;19.82 ]	
$AICc$	6.0384		3.9391	

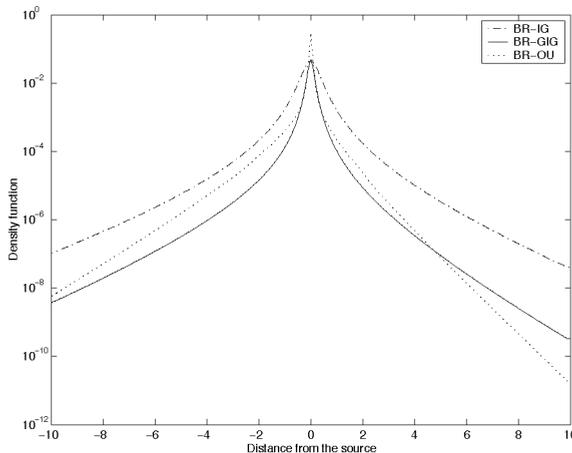
#### 5.4. Comparison with the external parameters

In order to investigate whether these approaches could be used in other situations, we compared the values of the parameters present in the models to the values that could be obtained by other means. In Klein *et al.* ((2003), results were promising in the sense that predicted/observed comparisons were good, but residuals from the statistical analysis (either least-square estimators or Binomial Maximum Likelihood estimators) were unsatisfactory. Moreover, the parameters values estimated from the field experiment differed from those calculated from external physical parameters. The differences of the calculated

INFERENCE FOR POLLEN DISPERSAL



(a)

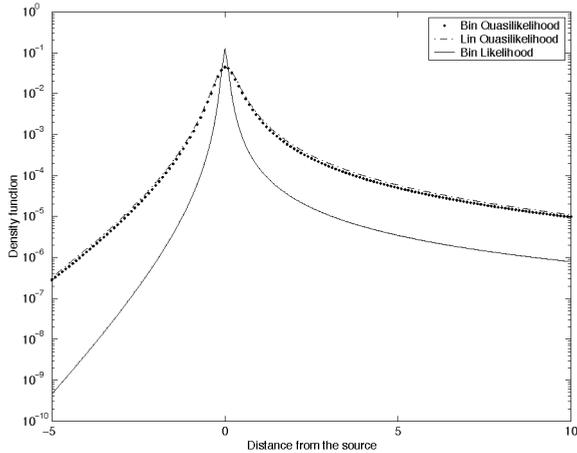


(b)

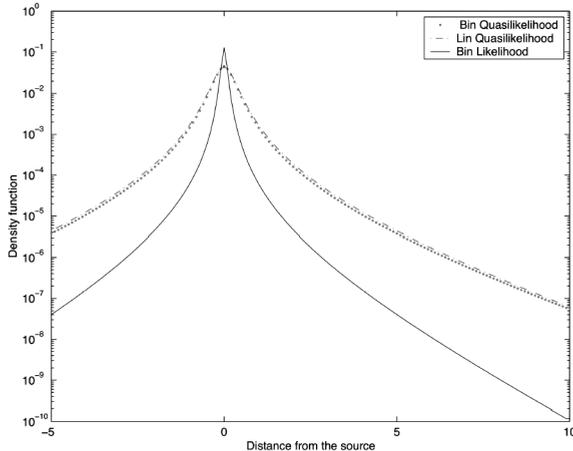
FIG 2. — Individual dispersal functions on  $\mathbb{R}$  for Models BR-IG (dashed points), BR-GIG (solid line) and BR-OU (points) following the dominant wind axis (a) (distance to the source between -5 and 15 m) and the orthogonal direction (b) (distance between -10 and 10 m) using Linear Quasilikelihood.

parameters between minimal, median or maximal cases were also large. Our knowledge about turbulence parameters (vertical variance  $\tau_z$  and horizontal variance  $\tau$ ) is very poor for this experiment. This has consequences on the values of parameters  $\delta, \lambda_x, \lambda_y, \lambda_z$ . Moreover, the correspondence between the parameters estimated from the dispersal pattern and those calculated from independent measures of physical parameters was rough. This was to be expected since getting simultaneously these meteorological data had not been planned in this experiment, and comparisons actually relied on inaccurate data.

INFERENCE FOR POLLEN DISPERSAL



(a)



(b)

FIG 3. — Individual dispersal functions on  $\mathbb{R}$  for Model BR-IG following the dominant wind axis (a) and the orthogonal direction (b) (distance from the source between -5 and 10 m) using successively Binomial likelihood (solid line), Quasilikelihood (points) and Linear Quasilikelihood (dashed points).

The new estimations obtained here re-analysing this experiment significantly improved the fits with the external parameters. Results are given in Table 4. Wind physical parameters are  $f_x = -0.056$  and  $f_y = 0.998$ . They are respectively estimated by  $-0.074$  and  $1.74$  in Klein *et al.* (2003), and by  $-0.061$  and  $1.22$  here. Turbulence coefficients are poorly known in this experiment. The vertical turbulence or diffusion coefficient  $\tau_z$  is approximately 1.18. It was estimated by 2.37 and by 1.65 here. The gap is even larger for the horizontal turbulence. It is close to 1.33 and was estimated by 5.70 in Klein *et al.* (2003)

and 3.51 here. The BR-OU model led to parameters that were consistent with the physical parameters, but not better than the ones obtained with the BR-IG model. Although not optimal, these results seem more reliable than the previous ones. They are moreover positive with respect to the ability to keep in the model some a priori knowledge. Models studied here rely only on a mean wind intensity during all the pollination period. For this experiment, we had just this knowledge. These models could easily include a variable wind intensity so that these results are promising concerning the ability of getting predictions in other situations.

TABLE 4. — Comparison between physical parameters and estimated parameters for the BR-IG model and the three methods; for the BR-GIG model and the Linear Quasilikelihood.

Parameters				BR-IG	BR-IG	BR-IG	BR-GIG
	Min	Mean	Max	Bin Likelihood	Bin QL	Lin QL	Lin QL
Vertical Drift, $f_z (m.s^{-1})$		0.183					
Height difference, $h (m)$		0.831					
Horizontal drift: $f_x (m.s^{-1}) f_y (m.s^{-1})$		-0.056		-0.074	-0.042	-0.061	-0.0362
		0.998		1.74	1.37	1.22	0.742
Vertical variance, $\tau_z (m.s^{-1})$	0.35	1.175	2	2.37	1.65	1.51	1.33
Horizontal variances, $\tau_x = \tau_y (m.s^{-1})$	0.65	1.325	2	5.70	3.83	3.51	2.75

## 6. Discussion and Conclusion

Two problems were addressed here. The first one was to build an accurate statistical framework to analyse dispersal data obtained from these kinds of experiments (cross-pollination rates between fields). The second one was to investigate whether mechanistic approaches were suitable to analyse these data. We considered models previously studied in Klein *et al.* (2003) and built new models derived from the Lagrangian Stochastic approach used in micrometeorology. We illustrated the process of estimating and comparing individual dispersal functions.

### 6.1. Statistical methodology

We have concentrated on the estimation of the individual dispersion function from the noisy observation of the backward dispersal function (1) since cross-pollinations strongly depend on the shapes, sizes and positions of a marked pollen source and a non-marked one. When competition is taken into account, this leads to a non-linear deconvolution problem. This deconvolution problem is impossible to study with classical deconvolution methods. It also precludes a non parametric approach. We solved the deconvolution problem by building parametric functions for the individual dispersion function (see Tufto *et al.* 1997, Klein *et al.* 2003 for a similar approach).

Results obtained in Klein *et al.* (2003) were promising since predicted/observed comparisons were good, but residuals from the statistical analysis (either least-square estimators or Maximum Likelihood estimators associated with the Binomial model) were not satisfactory. We reconsidered the experiment studied in Klein *et al.* (2003) and reanalysed the data here using a statistically more accurate framework. We fitted two-dimensional models to two-dimensional data. This is better in many regards. The model accuracy is strongly improved since all data are used to fit the dispersal model. Two-dimensional dispersal functions clearly show that a better insight is gained and that looking at various directions for dispersal function is important. The Lagrangian Stochastic approach (Aylor *et al.* 2003) and other models (McCartney and Fitt 1998) model dispersal only in the dominant wind direction.

Our results stressed the importance of the statistical method for analysing the data. We compared the Binomial likelihood method to two other quasilielihood methods. Assumptions only concern the mean and the variance of the observations and assumptions on the precise distribution of the observations are no longer required to get good estimation results. On the other hand the Binomial likelihood is quite sensitive to deviations from the statistical model. We obtained results that differed from results of Klein *et al.* (2003) just using such methods. Improvements were also obtained by modelling the error variances. Observed data display at most locations very low proportions of marked grains ( $\mu_k = n_k/N$  is generally lower than 0.2). This results in theoretical small variances in Binomial models ( $N\mu_k(1 - \mu_k)$ ) especially when there is no marked pollen observed, leading to statistical analyses that implicitly take more into account these last data. Analysing residuals confirmed that really good predictions were obtained with these models for cross-pollination rates among fields. The statistical analysis is performed using short and medium range data. Hence, even if individual dispersal functions are defined on  $\mathbb{R}^2$  and thus theoretically might give predictions at large distances, it is hardly possible to expect good predictions at long distances using these models. Although the models studied here do not possess heavy tails, there is no contradiction with other results where heavier tails of individual dispersal functions were obtained (Oddou-Muratorio *et al.* 2005, Devaux *et al.* 2005).

## 6.2. Mechanistic approaches for pollen dispersal

We proposed new models derived from aerobiological models. These models were estimated with the above statistical framework, and compared to previous ones. Our goal was to take better into account some physical phenomena that affect pollen dispersal. They are detailed in Aylor *et al.* (2003) and McCartney and Fitt (1998). There is an interplay between physics and biology, and between space and time. Managing this interplay is quite difficult, especially when trying (1) to keep the main important physical phenomena and (2) to obtain tractable models that can be estimated from data.

Trajectories of pollen grains can be formulated in several ways. One way to do it is based on the Gaussian Plume model. An alternate way is based on a Lagrangian Stochastic particle trajectory model. According to Aylor *et al.* (2003), the Gaussian Plume model lacks sufficient detail to look at short and medium range transport, which are best approached using a L.S. pollen trajectory model. We kept this last approach in this work. It differs with the models previously investigated (Tufto *et al.* 1997, Nurminieni *et al.* 1998, Klein *et al.* 2003) in which pollen trajectories are Brownian motion with drift. This is similar to the Gaussian Plume model. Actually, the Lagrangian Stochastic approach relies on the Langevin equations for modelling velocities. Vertical and horizontal velocities satisfy stochastic differential equations having for drift and diffusion coefficients the Langevin coefficients. Then, integrating these velocities yields pollen trajectories, which are in this approach integrated diffusion processes.

In this work, we investigated the contribution of models deriving from these aerobiological approaches. We considered trajectories for the vertical component obtained by modelling velocity with an Ornstein-Uhlenbeck diffusion process. This aimed to take better into account the stochasticity of the vertical component. An additional term modelled the settling velocity. Since we have to deal with effective pollination, many mechanisms are involved and the approaches used in micrometeorology are usually very detailed (Aylor *et al.* 2003). To keep models tractable, especially for statistical purposes, we rather chose a simplified mechanism for pollination times: the first hitting time of silk levels for a particle starting from a male flower. For instance, in a maize plant 3 meters high, male flowers (tassels) are higher than female flowers (cobs) with a vertical distance of 1-1.3. This is a natural way to link together pollen trajectories and pollination times without introducing many additional parameters. We finally obtained that the best model for this experiment was the BR-IG model. We do not know whether this is a general or a specific result due to the precise values of the parameters. We expect that different results could be obtained for other environmental conditions, and we do not exclude that conclusions could be different for other experiments. We can however highlight that we have proposed a general framework to study pollen dispersal, for both getting various parametric individual dispersal functions and performing accurate statistical analysis.

### 6.3. Extensions and future work

In this work, we examined the Lagrangian Stochastic approach in the vertical direction. A first modification to previous models was to introduce the well known dependence of wind speed and turbulence on the height above the ground. We evaluated this method on one experiment. We used here a specific model for the Langevin equation: the Ornstein-Uhlenbeck model ( $a(v) = d_z - \beta v$ ,  $b(v) = \eta$ ). Other models coming from micrometeorology or using external wind velocity measures should be studied. This approach clearly has to be applied to other experiments to conclude about it. A model where the two horizontal components are driven by Langevin stochastic equations (with either simple Langevin coefficients or coefficients that are defined using external velocities) should be considered. The vertical component would be defined either by a Brownian motion with drift or by the Langevin equation. Studying these models on just one experiment is not sufficient and other experiments should be analysed in order to investigate such models.

Studying pollen dispersal in heterogeneous environments is a crucial problem which requires further attention. Various studies have described cross-pollination rates from the observation of the backward function. This possesses the same drawbacks as before (dependence on the spatial design, shape and size of fields, dependence on the heterogeneous environments). The individual dispersal function approach however presents difficulties, chief among which is the fact that it can no longer be assumed that plants disperse their pollen according to the same individual dispersal function. This function has to depend on the respective positions of source and recipient plants. For oilseed rape, pollen dispersal has been studied with this approach in presence of discontinuities (Milhem *et al.* 2006). Results were consistent with previous experimental results concerning the backward dispersal function. This study remains a preliminary approach to the problem of heterogeneous environments which certainly requires further work.

## References

- ABRAMOWITZ M. and STEGUN A.I. (1972) *Handbook of Mathematical Functions*. Dover Publications, New-York.
- ADAMS W.T., GRIFFIN A.R. and MORAN G.F. (1992) Using paternity analysis to measure effective pollen dispersal in plant populations. *American Naturalist*, **140**, 762-780.
- AUSTERLITZ F., DICK C.W., DUTECH C., KLEIN E.K., ODDOU-MURATORIO S., SMOUSE P.E. and SORK V.L. (2004) Using genetics markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937-954.
- AYLOR D.E., SCHULTES N.P. and SHIELDS E.J. (2003) An aerobiological framework for assessing cross-pollination in maize. *Agricultural and Forest Meteorology*, **119**, 111-129.
- BARNDORFF-NIELSEN O.E. (1997) Normal Inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. of Statistics*, **24**, 1-13.
- BATEMAN A.J. (1947) Contamination of seed crops. II. Wind pollination. *Heredity*, **1**, 235-246.

- BOLKER P. and PACALA S.W. (1997) Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, **52**, 179-197.
- BURCZYK J. and KORALEWSKI T.E. (2005) Parentage versus two-generation analyses for estimating pollen-mediated gene flow in plant populations. *Molecular Ecology*, **14**, 2525-2537.
- CLARK J.S., LEWIS M. and HORVATH L. (2001) Invasion by extremes: population spread with variation in dispersal and reproduction. *American Naturalist*, **157**, 537-554.
- CLARK J.S., SILMAN N., KERN R., MACKLIN E. and HILLERISLAMBERS J. (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475-1494.
- COLLETT D. (1991) *Modelling binary data*. Chapman and Hall, London.
- CRESSWELL J.E., DAVIES T.W., PATRICK M.A., RUSSELL F., PENNEL C., VICOT M. and LAHOUBI M. (2004) Aerodynamics of wind pollination in a zoophilous flower, *Brassica Napus*. *Functional Ecology*, **18**, 861-866.
- DEVAUX C., LAVIGNE C., FALENTIN-GUYOMARC'H H., VAUTRIN S., LECOMTE J. and KLEIN E.K. (2005) High diversity of oilseed rape pollen clouds over an agro-ecosystem indicates long-distance dispersal. *Molecular Ecology*, **14**, 2269-2280.
- DEVOS Y., REHEUL D. and DE SCHRIJVER A. (2005) The co-existence between transgenic and non-transgenic maize in the European Union: a focus on pollen flow and cross-fertilization. *Environ. Biosafety Res.*, **4**, 71-87.
- DURBIN J. (1992) The first-passage density of the Brownian motion process through a curved boundary. *J. Appl. Prob.*, **29**, 291-304.
- ELLSTRAND N.C. (1992) Gene flow by pollen: implications for plant conservation genetics. *Oikos*, **63**, 77-86.
- GRIMAUD A. (2005) Modélisation stochastique et estimation de la dispersion du pollen de maïs. Estimation dans des modèles à volatilité stochastique. *Ph.D Thesis, Université Paris 7- Denis Diderot*.
- GRIMAUD A. and LARÉDO C. (2006) Parametric models for Corn Pollen dispersal using diffusion processes and statistical estimation. *Submitted to publication*. 25 p.
- HUET S., BOUVIER A., GRUET M.A. and JOLIVET E. (1996) *Statistical tools for nonlinear regression*. Springer-Verlag, New-York.
- HURVICH C.M. and TSAI C.L. (1995) Model selection for extended quasi-likelihood in small samples. *Biometrics*, **51**, 1077-1084.
- JORGENSEN B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture Notes in Statistics. Springer-Verlag.
- KLEIN E.K., LAVIGNE C., FOUÉILLASSAR X., GOUYON P.H. and LARÉDO C. (2003) Corn Pollen Dispersal: Quasi-mechanistic models and Field Experiments. *Ecological Monographs*, **73**, 131-150.
- KLEIN E.K., LAVIGNE C., PICAULT H., RENARD M. and GOUYON P.H. (2006) Pollination dispersal of oilseedrape: estimation of the dispersal and effects of field dimension. *Journal of Applied Ecology*, **43**, 141.
- LATTA R.G., LINHART D., FLECK D. and ELLIOT M. (1996) Direct and indirect estimates of seeds versus pollen movement within a population of Ponderosa Pine. *Evolution*, **52**, 61-67.

- LAVIGNE C., KLEIN E.K., VALLEE P., PIERRE J., GODELLE B. and RENARD M. (1998) A method to determine the mean dispersal pollen of individuals growing within a large pollen source. *Theoretical and Applied Genetics*, **96**, 886-896.
- LEWIS M.A. (2000) Spread rate for a nonlinear stochastic invasion *J. Math. Biol.*, **41**, 430-454.
- LOOS C., SEPPELT R., MEIER-BETHKE S., SCHIEMANN J. and RICHTER O. (2003) Spatially explicit modelling of transgenic maize pollen dispersal and cross-pollination. *Journal of Theoretical Biology*, **225**, 241-255.
- MCCARTNEY H.A. and FITT B.D.L. (1985) Construction of dispersal models. Pages 107-143. in C.A. Dilligan editor. *Mathematical modeling of crop diseases*. Academic Press, London.
- MCCARTNEY H. A. and LACEY M.E. (1991) Wind dispersal of pollen from crops of oilseedrape (*Brassica Napus* L.). *Journal of Aerosol Science*, **22**, 476-477.
- MCCULLAGH H.A. and NELDER J.A. (1989) *Generalized Linear Models*. 2nd Edition. Chapman and Hall, London.
- MEAGHER T.R. and VASSILIADIS C. (2003) Spatial geometry determines gene flow in plant populations. *Genes in the Environment* (eds R. Hails, J. Beringer and H.C.J. Godfray), 76-90. Blackwell Science, Oxford
- MILHEM H., H. PICAULT, M. RENARD and HUET S. (2006) Effect of a crop discontinuity on the oilseed rape pollen dispersal function. *Submitted to publication*. 17 p.
- NURMINIENI M., TUFTO J., NILSSON N.O., ROGNLI O.A. (1998) Spatial models of pollen dispersal in the forage grass meadow fescue. *Evolutionary Ecology*, **12**, 487-502.
- ODDOU-MURATORIO S., KLEIN E.K. and AUSTERLITZ F. (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441-4452.
- PERTHAME B. and SOUGANIDIS P.E. (2004) Front propagation for a jump model arising in spatial ecology. Preprint DMA-UMR8553, Ecole Normale Supérieure. Paris. 14 p.
- RAUPACH (1989) Applying Lagrangian fluid mechanics to infer scalar source distributions from concentration profiles in plant canopies. *Agric. Forest Meteorol.*, **47**, 85-108.
- RIEGER M.A., LAMOND M., PRESTON C., POWLES S.B., and ROUSCH R.T. (2002) Pollen mediated-movement of herbicide resistance between commercial canola fields. *Science*, **296**, 2386-2388.
- RODEAN H.C. (1996) Stochastic Lagrangian models of turbulent diffusion. *American Meteorological Society*, **26**, No 48.
- SEINFELD J.H. and PANDIS S.N. (1998) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Wiley, New York.
- SMOUSE P.E., DYER R.J., WESTFALL R.D. and SORK V.L. (2001) Two-generation analysis of pollen flow across a landscape .I. Male gamete heterogeneity among female. *Evolution*, **55**, 260-271.
- SQUIRE G.R., BURN D. and CRAWFORD J.W. (1997) A model for the impact of herbicide tolerance on the performance of oilseedrape as a volunteer weed. *Annals of Applied Biology*, **131**, 315-338.

INFERENCE FOR POLLEN DISPERSAL

- TREU R. and EMBERLIN J. (2000) Pollen dispersal in the crop Maize (*Zea mays*), Oil seed rape (*Brassica Napus*), Potatoes (*Solanum Tuberosum*), Sugar beet (*Beta vulgaris*) and Wheat. <http://www.soilassociation.org>
- TUFTO J., ENGEN S. and HINDAR K. (1997) Stochastic Dispersal Processes in Plant Populations. *Theoretical Population Biology*, **52**, 16-26.
- WEINBERGER H.F. (1978) Asymptotic behaviour of a model in population genetics. *In Chadam J.M.*. Lecture Notes in Mathematics: Nonlinear partial differential equations and applications, **648**, 47-96.