# QUANTIFYING OUTCROSSING PROBABILITIES OF GENETICALLY MODIFIED PLANTS

## DEVELOPMENT OF A PREDICTIVE MODEL

M. W. Smith-Kleefsman
F. J. Weissing
R. Bijlsma

Evolutionary Genetics / Theoretical Biology
Centre for Ecological and Evolutionary Studies
University of Groningen

February 2005

Results of literature study commissioned by the Commission on Genetic Modification (COGEM), an advisory body of the Ministry of Spatial Planning, Housing and the Environment, The Netherlands (VROM).

Advisory Committee:

| | |
|---|---|
| Dr. J. C. M. den Nijs (chairman) | Subcommission Agriculture of the COGEM |
| Ing. M. Berendsen | Subcommission Agriculture of the COGEM |
| Dr. ir. W. A. Brandenburg | Wageningen University |
| Em. prof. dr. W. van Delden | Subcommission Agriculture of the COGEM |
| Dr. ir. M. M. C. Gielkens | Bureau Genetically Modified Organisms (GGO) |

# CONTENTS

# EXECUTIVE SUMMARY

## DESCRIPTION OF THE PROBLEM

As an advisory body of the Dutch Ministry of Spatial Planning, Housing and the Environment (VROM), the Netherlands Commission on Genetic Modification (COGEM) has to evaluate, among other things, requests for permission to cultivate genetically modified (GM) crops in the natural environment. Many of these crops are able to cross with wild populations of the same species or one that is closely related. To reduce outcrossing probabilities, the modified crops are grown some distance away from possible recipient populations. The COGEM aims to develop a mathematical model that estimates outcrossing probabilities of GM crops with recipient populations in relation to the separation distance, to arrive at scientifically motivated and justified procedural rules. This literature study evaluates which components should be included in a mathematical model that estimates outcrossing probabilities and discusses possible modes to develop such model. In this report, we only focus on dispersal and outcrossing by means of pollen.

## LITERATURE SURVEY

The process that leads to outcrossing can be divided into three steps. *(A)* First, viable pollen must reach the stigma of a compatible specimen, which is usually, but not necessarily of the same species as the GM crop. The main processes and mechanisms that affect pollen dispersal are considered. *(i)* Self-fertilising species can be expected to lack specific (long-distance) dispersal mechanisms. Most self-fertilising species, however, show appreciable levels of outcrossing, but they are expected to have lower chances of outcrossing than typically outcrossing species, since pollen of the former species, being at low concentration in target populations, has to compete with much selfing pollen. Therefore, differences between outcrossers and selfers should be a component of the model. *(ii)* Pollen grains are not able to disperse themselves actively, but need dispersal vectors. In our biogeographic region, insects and wind seem to be the predominant dispersal vectors. Pollen flow by insects as well as by wind can be highly variable among species, among plant populations and over time, depending on the weather, on population characteristics and on the environment. Distances travelled by pollen of wind-pollinated species are generally larger than the distances travelled by insect-dispersed pollen, but the shape of wind and insect pollination curves relating pollination probability (or relative pollen density) to dispersal distance is generally highly similar. A large fraction of the pollen lands close to the donor plant and only a small fraction disperses further, some of which may travel large distances. From a modelling perspective, this similarity between the dispersal curves for wind-and insect-dispersed pollen has the great advantage that the same mathematical approach can be used for both. *(iii)* During the dispersal process, a certain percentage of the dispersing pollen will have lost viability. These pollen do not contribute to the fertilisation process. It is questionable whether, on the basis of current knowledge, pollen viability rates of crops growing in the field can be predicted.

*(B)* After landing on a suitable stigma, the pollen has to succeed in fertilising an ovule. Therefore, it has to compete with other pollen that has landed on the same stigma. The competitive ability of GM pollen may be different from that of pollen

from the target population. Some important mechanisms that influence fertilisation success are reduced viability, time of arrival, the exact place where the pollen lands and the presence of self-incompatibility mechanisms. Since the fertilisation process is very complicated that is affected by many mechanisms, the process needs to be incorporated the model in a simplified mode to be workable.

*(C)* When hybrid seed is developed successfully, the modified DNA of the GM species must be incorporated into the gene pool of the receiving species (introgression). Whether this will happen, is in the beginning mainly dependent on stochastic processes that determine whether the modified DNA will establish in the receiving population. Later on, when the modified DNA is established in sufficient numbers of plants to render stochastic effects negligible, deterministic processes will determine the persistence of the modified DNA. Introgression is a very complicated process on which little information is available; therefore, modelling of this part is not considered at present.

## MODELLING APPROACH

An overview of the proposed mathematical that estimates probabilities of pollen to land in populations of compatible species and to achieve fertilisation in such population is given in Figure 4 (pg 20). This proposed model consists of three modules, but only the first two are programmed at present.

*(A)* The first module addresses the question: how does pollen, originating from a GM source population, disperse over the landscape? In this module, pollen dispersal of a source population is simulated. Different equations are pre-programmed that can be used to describe the dispersal pattern of a species. These are a negative exponential equation, an inverse power law, and a uniform distribution. The user can also choose to estimate the dispersal process using two equations, one describing the first part of the curve (i.e. describing the dispersal pattern of the pollen that lands close to the dispersing individual), the other one describing the tail of the curve (i.e. describing the dispersal pattern of the pollen that travels farther away). If the dispersal pattern is described by another known curve than the ones that are pre-programmed, the user can enter this equation using the custom function. In this module, data on pollen viability can be added; pollen survival can be described by a negative cumulative normal distribution, or by an equation added by the user.

*(B)* The second module addresses the question: what is the expected frequency of seeds in a target population that is fertilised by pollen originating from a given GM source population? In this module, the percentage of seeds that is the result of a cross between pollen from the source population and ovules from the target populations is calculated, thus giving an estimation of the contamination level of the target population with DNA from the GM source population. Relative competitive ability of the pollen is included in this module.

## CONCLUSIONS

The model will be helpful for estimating the separation distances required to reduce contamination levels with modified DNA to acceptably low levels. The outcome of the model will largely depend on the parameter values entered by the user. The user should be aware of two types of uncertainty associated with this. One type of uncer-

tainty is whether the parameter values used have been estimated correctly. The other uncertainty is caused by variation in parameter values due to stochastic processes. The model, however, only handles 'standard' situations: it does not reckon with fluctuations in time and/or space. The process of pollen dispersal, however, appeared to be highly variable.

We recommend that in the future the COGEM aims to extend and refine the present model, for example *(i)* by including more complex situations, like estimation of gene flow at landscape level, and/or *(ii)* by including stochasticity, allowing to estimate possible deviation from the contamination levels found. By adding modules and functions like these, the procedure to estimate outcrossing probabilities of GM populations with cultivated or wild relatives will be continually improved.

# CHAPTER 1. INTRODUCTION

## 1.1 DESCRIPTION OF THE PROBLEM

As an advisory body of the Dutch Ministry of Spatial Planning, Housing and the Environment (VROM), the Netherlands Commission on Genetic Modification (COGEM) has to evaluate, among other things, requests for permission to cultivate genetically modified[1] (GM) crops in the natural environment. By means of an environmental risk analysis (Box 1), the COGEM estimates the risk involved in cultivating GM plants. This risk can be defined as a function of the effect of cultivating GM plants and the likelihood of the effect, i.e. the likelihood to outcross with a wild or cultivated relative. Many GM crops, as well as non-modified crops, do cross with wild populations of the same species or one that is closely related (for a review see Groot *et al.* 2003). This is not surprising, since crops were developed from wild species only a few thousand generations ago and are therefore expected to have relatively high genetic similarity with their wild relatives (Ellstrand *et al.* 1999). Hybridisation is possibly even more likely to occur between two crop species. Groot *et al.* (2003) conclude that crop-wild and crop-to-crop gene flow can be expected in nearly all cultivated crops that are grown in their reproductive phase.

To reduce outcrossing probabilities, the modified crops are grown some distance away from possible recipient populations. At present, the choice of such distances is largely based on an ad-hoc procedure, simply doubling the separation distances advised by the Dutch General Inspection Service[2]. To arrive at scientifically motivated and justified procedural rules, the COGEM aims to develop a mathematical model that estimates outcrossing probabilities of GM crops with recipient populations in relation to the separation distance. Such model would clearly not constitute a complete ERA, since a full analysis should include estimation of the impact on the environment of a potential outcrossing event as well as the probability it occurs. This literature study evaluates which components should be included in a mathematical model that estimates outcrossing probabilities and discusses possible modes to develop such model.

## 1.2 OUTLINE OF THE REPORT

The process that leads to outcrossing can be divided into three steps (fig. 1). (A) First, viable pollen must reach the stigma of a compatible specimen[3] (chapter 2). Two types of potential recipient species can be distinguished, namely non-modified crops cultivated by farmers and related wild populations. Both will be discussed in this report. The ability to arrive at a suitable stigma depends on species- specific characteristics concerning dispersal. In this report, we only focus on dispersal and

[1] Genetic modification can be defined as any change in the genetic constitution of a living organism (here plants) that has been brought about by joining together *in vitro* genes from different sources or genes that have in some way been modified *in vitro*. Genetic engineering and recombinant DNA techniques are synonymous with genetic modification.

[2] The separation distances used by the Dutch General Inspection Service (NAK) are based on reducing gene flow into crops grown for seed production, to keep contamination within the quality demands of the Inspection Service.

[3] This compatible specimen is usually, but not necessarily of the same species as the GM crop.

outcrossing by means of pollen, since *(i)* pollen exchange, unlike seed exchange, directly results in genetic exchange between the populations involved, and *(ii)* farmers cultivating non-modified crops, especially organic farmers, are interested in contamination levels of their seeds, which is a direct result of pollen exchange.

(B) After landing on a suitable stigma, the pollen has to succeed in fertilising an ovule (chapter 3). Therefore, it has to compete with other pollen that has landed on the same stigma.

(C) When hybrid seed is developed successfully, the modified DNA of the GM

**BOX 1. ENVIRONMENTAL RISK ANALYSIS**

At present, the COGEM applies a precautionary principle to assess requests for permission to cultivate genetically modified crops. Which precautions are taken is mainly based upon the amount of information available. When little information is present, only small field experiments may be carried out. When additional information is available about effects on the environment and the transferred elements, larger experiments are permitted, but only when the effects on the environment are expected not to be deleterious. In table B1.1, the information requirements for different classes of field experiments are given.

If the environmental risk analysis shows that the deliberate release may result in adverse effects, either the request will be refused or, more commonly, risk-management measures will be imposed by demanding (extra) constraints. Mostly, the second option is chosen, which is called risk management. Risk management cannot exclude all possible risks, but it aims to minimise them. One risk management option is to enforce separation distances or increase those that have already been proposed. From the point of view of environmental safety, it is important to know the effectiveness of such a regulation. A mathematical model may give better insight into whether, at the required separation distance, outcrossing probabilities are sufficiently reduced, and may indicate whether adjustments are required.

**Table B1.1.** Guidelines used by the COGEM for the evaluation of requests concerning field experiments with GM plants. Five different classes are distinguished (COGEM 1999).

| Class | Size of field | Max. nr of locations | Spread | Information requirements |
|---|---|---|---|---|
| 1 | 1 ha per location | 1 | - Prevented by removal of inflorescence or doubling of the NAK separation distance (see footnote page 1).<br>- The effects of the genetic modification are properly monitored by observing the experimental field(s). | - genetic element involved<br>- donor involved<br>- suspected function(s) of the genetic element after expression |
| 2 | 1 ha per location | 5 | As class 1 | As class 1 +<br>possible effects of the expression based on former experiments |
| 3 | total of 5 ha | 10 | - No prevention. The genetically modified organisms are kept separated in the field.<br>- The effects of the genetic modification are properly monitored by observing the experimental field(s). | As class 2 +<br>no reasons to suspect deleterious effects of the genetically modified organisms, its offspring or after transfer to other organisms |
| 4 | total of 10 ha | 10 | As class 3 | As class 3 +<br>the map of constructs used for the modification, showing the combinations of (regulation) sequences that are expressed and showing other selection elements |
| 5 | no maximum | no maximum | As class 3 | As class 4 +<br>- molecular characteristics of the transferred elements<br>- complete performed assessment concerning the safety for the environment, public health and animal feed |

species must be incorporated into the gene pool of the receiving species (introgression, chapter 4). Whether this will happen, is firstly dependent on stochastic processes that determine whether the modified DNA will establish in the receiving population or not. Later on, when the modified DNA is established in sufficient numbers of plants to render stochastic effects negligible, deterministic processes will determine the persistence of the modified DNA.

Different approaches can be taken to simulate the outcrossing process (chapter 5). In this report we will present what we believe is the best way of modelling the pollen dispersal and fertilisation processes, taking into account the aim for which such a model would be developed; that is, for the use of the COGEM to estimate outcrossing probabilities of GM plants in relation to separation distances. In the end, a description of the model we will propose is given.

Introgression is a very complicated process on which little information is available; therefore we will not consider modelling of this part.

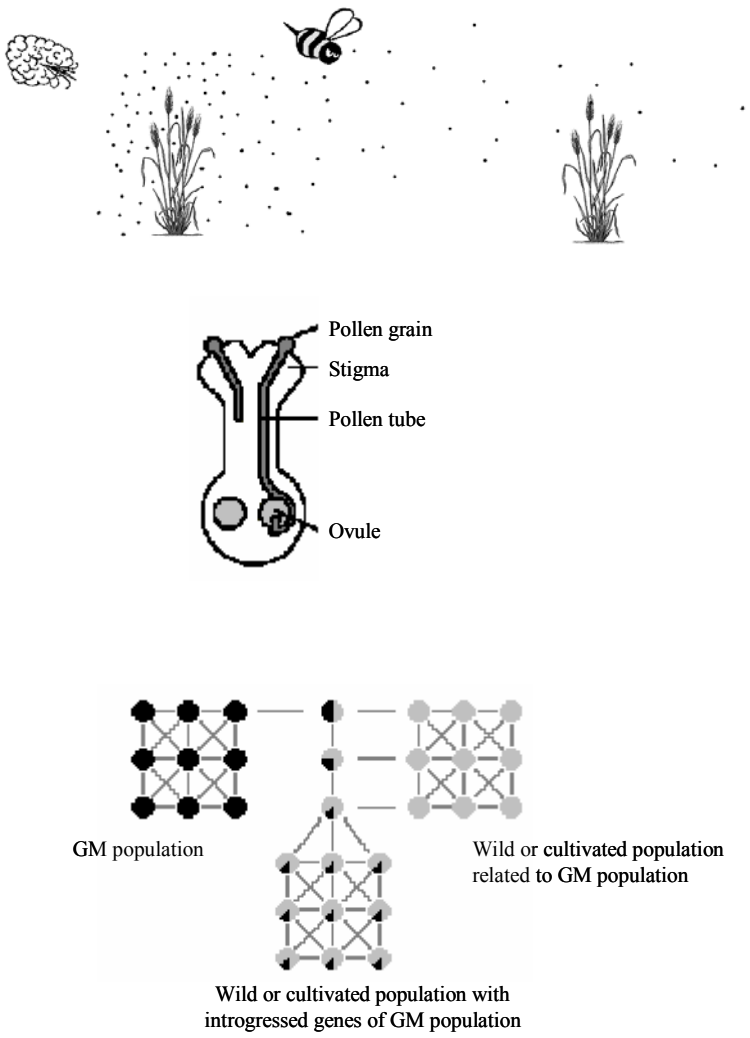


**Figure 1.** Schematic depiction of the outcrossing process, which can be divided into three distinct parts (A, B, C). A. Viable pollen grains must disperse over the distance between the GM population and related wild or cultivated populations. B. After reaching such population, the pollen must fertilise an ovule, resulting in hybrid seed. C. The modified DNA present in the hybrid must establish itself and persist (i.e. introgress) in the genome of the receiving population.

# CHAPTER 2. POLLEN DISPERSAL

## 2.1 INTRODUCTION

Before a pollen grain has any chance of fertilising an ovule of a compatible species, it must land on the stigma of this species (fig. 1A). In this chapter, we will consider processes and mechanisms that affect pollen dispersal, thereby affecting the chance of pollen reaching recipient populations growing at different distances. Self-fertilising species can be expected to lack specific (long-distance) dispersal mechanisms. These species will be compared with typically outcrossing species. The two most important pollination vectors, insects and wind, will be considered. During the time between release of the pollen from the GM plant and deposition on a wild or cultivated relative, the pollen may have lost viability, making the pollen grain unable to fertilise. Gene-flow barriers are intended to decrease pollen flow out of GM crops. We will briefly review different types of gene-flow barriers.

## 2.2 SELF-FERTILISING VERSUS OUTCROSSING

Plant species that are entirely self-fertilising are expected to lack mechanisms that promote pollen dispersal. Most self-fertilising species, however, show appreciable levels of outcrossing, especially animal-pollinated species. Vogler & Kalisz (2001) found most wind-pollinated species to be either highly selfing or highly outcrossing, with intermediate outcrossing rates rare but present. In contrast, animal-pollinated species exhibited a bimodal, but more continuous, distribution of estimates of outcrossing rates (fig. 2). A factor contributing to this difference could be that the presence of wind is relatively constant in natural populations, whereas animal abundance and visitation rates are highly variable.

Although most so-called selfers do show low levels of outcrossing, gene flow between different populations is low. Wagner & Allard (1991) estimated gene flow by pollen in barley (*Hordeum vulgare*), a predominantly selfing species with an outcrossing rate of about one percent, by counting the number of 'hybrids' formed (i.e. seeds of crosses between two homozygous parental plants having alternative genotypes). They found two hybrids, one from parents 7 m apart and one from parents 60 m apart. Certainly, more pollen travelled over these distances than only that which resulted in hybrids, but the indication is that pollen flow is low. Golenberg (1987) found that gene flow by pollen in wild emmer wheat (*Triticum dicoccoides*, outcrossing rates about 0.5 per cent) to be limited to distances up to 15 m. Other indications of low levels of gene flow can be found looking at the genetic differentiation between subpopulations (Box 2). Berge *et al.* (1998) found high levels of genetic differentiation between, and high levels of inbreeding within, subpopulations of *Arabis thaliana*, a highly selfing species, suggesting low levels of gene flow. This gene flow was accomplished by pollen and seed exchange. Therefore, pollen is expected to have contributed only partly to this already low differentiation.

Although highly selfing species cannot be excluded from our model, differences between outcrossers and selfers should be taken into account, as mainly self-fertilising and mixed-mating species are expected to have lower chances of outcrossing than typically outcrossing species, since pollen of the former species, being at low concentration in target populations, has to compete with much selfing pollen.

Unlike selfers, obligate outcrossing species are not able to fertilise themselves. Especially many animal-pollinated species are obligate outcrossers (fig. 2). Pollen of these species landing on an own stigma are not competing with outcrossing pollen landing on the same stigma for achieving fertilisation. The main mechanism of obligate outcrossing species to prevent fertilisation is self-incompatibility, which will be discussed in Chapter 3.2.

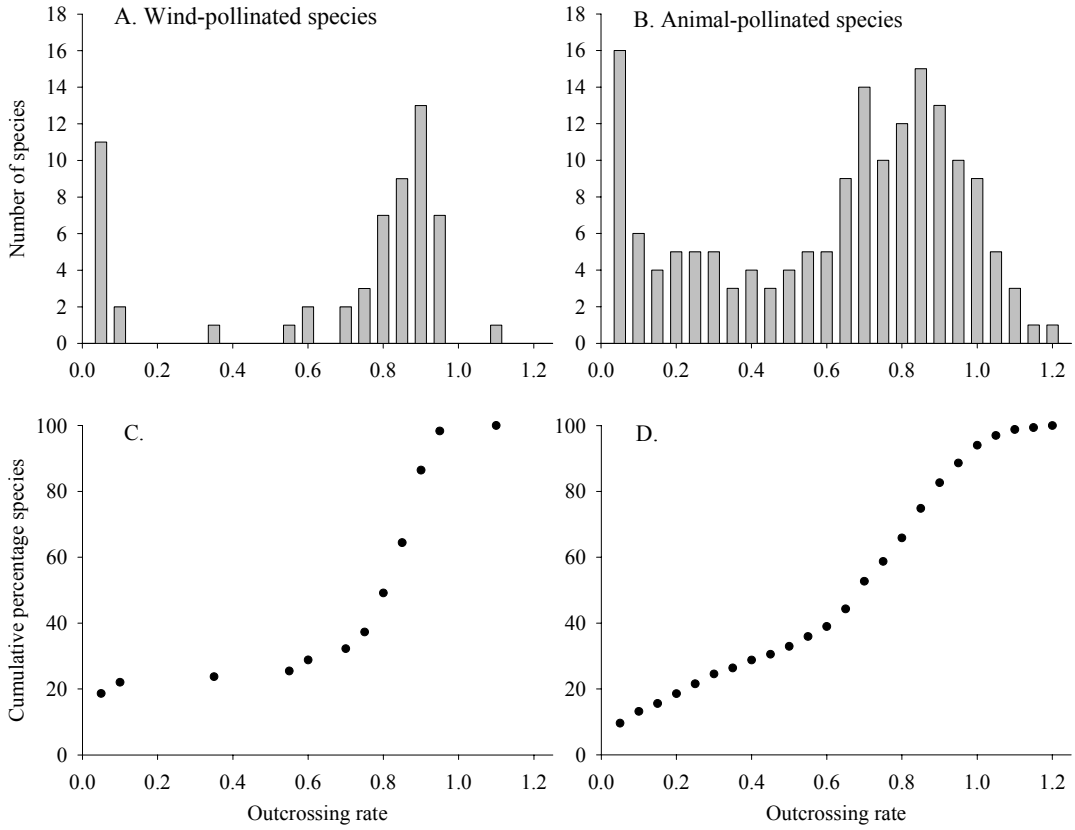


**Figure 2.** Distribution of outcrossing rate estimates for wind- (A, C, n=59) and animal- (B, D, n=169) pollinated species expressed as number of species (A, B) and cumulative percentage of species (C, D). Outcrossing rates are defined as the proportion of seed produced through outcrossing. Some of the estimated outcrossing rates are higher than one. This is a consequence of the estimation procedure. After Volger & Kalisz (2001).

BOX 2. GENETIC DIFFERENTIATION

F-statistics are commonly used to measure genetic differentiation, in which $F_{ST}$ is the measure of differentiation between populations. Specifically, $F_{ST}$ is the correlation between random gametes within each subpopulation relative to the gametes of all subpopulations together (Sork *et al.* 1999). It is calculated as follows:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where $H_S$ is the expected heterozygosity of an individual per subpopulation, averaged over all subpopulations, and $H_T$ is the expected heterozygosity of an individual in the total population (all subpopulations together). When $F_{ST} = 0$, there is no genetic differentiation, the allele frequencies in all subpopulations are equal, indicating high gene flow among populations. When $F_{ST} = 1$, all subpopulations are fixed for different alleles.

## 2.3 DISPERSAL VECTORS OF POLLEN

Pollen grains are not able to disperse themselves actively, but need dispersal vectors. These can be biotic, like insects, birds and mammals, or abiotic, like wind and water (Meeuse 1961, Holm 1978). Probably over 90 per cent of the angiosperms is pollinated by animals, and by far the greater part of this by insects (Wilcock & Neiland 2002), while 30 out of 300 plant families contain species that show adaptations for pollen dispersal by wind, for example grasses, sedges and rushes (Knox 1979). In our biogeographic region, insects and wind seem to be the predominant dispersal vectors. In this report we will therefore concentrate on pollen dispersal by insects and wind.

### *2.3.1 Dispersal by insects*

Pollen dispersal by insects is strongly dependent on ecological factors influencing the behaviour and occurrence of insects (Richards *et al.* 1999, Velterop 2000). Characteristics of the pollinating species determine pollen flow patterns and pollination efficiency. Hymenoptera, for example, are characterised by short flights, large pollen loads and high efficiency, while Lepidoptera are characterised by long flights, small loads and medium to low efficiency (Herrera 1987). Insects visiting flowers of the same species without visiting other species are more efficient than those that switch frequently between plant species, since the former deposit higher amounts of conspecific pollen (Velterop 2000). Some insect species visit exclusively certain plant groups (specialists), but most species use a broad range of different plant species (generalists). Even a generalist species may be able to deposit a large amount of conspecific pollen on plant stigmas, since different individuals may still specialise on only a few or a single species (flower constancy, Chittka *et al.* 1999). Flower constancy is known for several bee species, butterflies and hoverflies (e.g. Waser 1986, Goulson *et al.* 1997, Goulson & Wright 1998).

An important characteristic of insects is their flight distance, which gives an indication of the distance over which pollen flow occurs (Velterop 2000). Butterflies, for example, fly on average larger distances between subsequent flower visits than bumblebees, which fly between neighbouring flowers because of their high energy demands (Handel 1983, Herrera 1987). However, although most pollinators generally keep the flight distance low, many are capable of flying much longer distances. For honeybees was found that they forage till about 10 km from the hive, although especially in agricultural areas, a foraging radius of only a few hundred metres was found (Beekman & Ratnieks 2000). Furthermore, honeybees leaving the hive sometimes carry viable pollen that has remained on the body from an earlier flight trip, so a honeybee may be able to cross-pollinate plants more widely separated than it could visit in a single foraging trip. Another possibility is that pollen may be transferred from one honeybee to another in the hive. There seems to be no reason why some of this pollen should not also be viable (Free & Williams 1972).

Plant population parameters such as size, density and isolation may affect pollen dispersal by insects, because it affects pollinator behaviour. Pollinators forage in such a way that the nectar gain per flower is independent of plant size and the rate of nectar production per flower (Dreisig 1995). Therefore, population size and distance between populations interact with each other, resulting in different patterns of gene flow among populations that are adjacent to each other or far apart (Richards *et al.*

1999). For wild radish (*Raphanus savitus* L.), it has been shown that large populations at larger distances contributed more to pollen import than small populations nearby (Ellstrand *et al.* 1989). In populations with low density, migrant pollen generally constitutes a higher relative fraction of the total pollen amount than in populations with high density (Handel 1983, Richards *et al.* 1999). Other ecological parameters, such as population shape, presence or absence of alternative hosts for the pollinators, plant biomass and rates of flower production, will be of influence as well (Ellstrand *et al.* 1989).

To summarise, pollen flow by insects can be highly variable among species, among plant populations and over time, but the dispersal distances are generally not very high, ranging from 100 to 200 m for short- to medium-distance dispersal, to approximately 10 km for long-distance dispersal. This infrequent long-distance dispersal is of importance, however. For many crops, it's not known how far its pollen is able to spread. Table 1 gives large dispersal distances of two important insect-pollinated crop species (oilseed rape and potato).

### *2.3.2 Dispersal by wind*

Pollen dispersal by wind is dependent on falling velocity and releasing height of the pollen and on wind characteristics. Not only horizontal speed and direction of the wind are of importance, but also turbulence. Two kinds of turbulence can be distinguished (Tackenberg 2003). *(i)* Mechanistic turbulence with high horizontal wind speed. This dominates in stormy weather and is associated with a mean downdraft (i.e. downward air current), although updrafts (i.e. upward air currents) are present. *(ii)* Thermal turbulence. This is caused by an increase in air temperature and is therefore associated with sunny weather. In this case, updraft dominates downdraft. Updrafts are particularly important for seed dispersal, because they lift seeds high in the air, which can explain long-distance dispersal. Therefore, contrary to popular belief, it is not stormy weather but sunny weather that causes seeds to travel over large distances. It is highly likely that the same is true for pollen dispersal.

Wind-dispersed pollen is usually small and light, resulting in low falling velocity and thus promoting pollen dispersal. In most alders, hazels and junipers, average pollen size is about 30 $\mu$m. Pollen from wind-pollinated plants, especially heavier ones, often has additional characteristics that support pollen dispersal, such as air sacs or a modified shape (Meeuse 1961).

Pollen grains have very low chances of landing on a compatible stigma, because they land in an arbitrary place. A large amount of pollen is therefore needed to ensure reasonable pollination success. Although some species, like wild oat and brome, release relatively few pollen grains (less than a thousand per spike, Knox 1979), most wind-pollinated species produce vast numbers. For example, the amount of pollen of one catkin can be higher than two million (Meeuse 1961), and some grasses, like ryegrass, cocksfoot, and canary grass, release between two and five million pollen grains from one spike (Knox 1979). To keep wastage of pollen to a minimum, most wind-pollinated species only release their pollen when conditions are favourable, for example in dry weather, and they do not release all of it at once (Meeuse 1961).

Pollination rate by wind depends not only on the characteristics of the pollen and the wind, but also on population size and density (Handel 1983). Raynor *et al.* (1971, 1972) found for ragweed, timothy and maize that relative pollen concentration in the air from a small source decreased more rapidly with distance than from larger sources.

For pollen dispersal by wind, topography is of importance too. For example, when a population is growing on a slope, a gust can carry it away from the hillside and thereby increase the releasing height.

To summarise, pollen dispersal by wind can be highly variable, not only depending on the weather, but also on population characteristics. Pollen dispersed by wind is able to travel over large distances (up to over 100 km). Watrud *et al.* (2004) found creeping bentgrass (*Agrostis stolonifera*) to outcross over a maximum distance of 21 km, although most gene flow by pollen occurred within 2 km. Tyldesley (1973) even found pollen of different tree species (e.g. birch, pine, juniper and larch) to travel for at least 250 km over sea. Distances travelled by pollen of wind-pollinated species are generally larger than the distances travelled by insect-dispersed pollen, but the number travelling over such large distances is very small. Table 1 gives large dispersal distances of two important wind-pollinated crop species (sugar beet and maize). For many crop species, however, it's not known how far its pollen is able to spread.

**Table 1.** Overview of the the most important pollination vectors and the largest measured pollen dispersal distance of four important crop species. Notice that for all four species, no measurements at further distances are made than the distances mentioned in the table. Therefore, it is very likely that pollen have travelled further than these distance. Reviewed by Treu & Emberlin (2000).

| Species | Main pollination vector | Largest measured distance |
|---|---|---|
| Oilseed rape<br>*Brassica napus* | Insects | 4000 m by insects<br>3000 m by wind |
| Potato<br>*Solanum tuberosum* | Insects | 1000 m |
| Sugar beet<br>*Beta vulgaris* | Wind | 800 m |
| Maize<br>*Zea mays* | Wind | 800 m |

**Table 2.** Comparison of wind- and insect-mediated pollen dispersal.

| Wind | Insects |
|---|---|
| Moves pollen in large masses and mainly downwind. | Move pollen independently from each other and more or less randomly in all directions, although dependent on the location of nest or hive. |
| No regard for species. All pollen are taken and deposited at an arbitrary place. Probability of arrival at a compatible stigma is low. Therefore, pollen production is high compared to insect-pollinated species. | Often distribution is systematically within plant species, due to specialisation or flower constancy. Only pollen of visited species is taken. Probability of arrival at a compatible stigma is high. Therefore, pollen production is low compared to wind-pollinated species. |
| Unlimited load of pollen. | Limited load of pollen. Only after deposition can new pollen be loaded. |

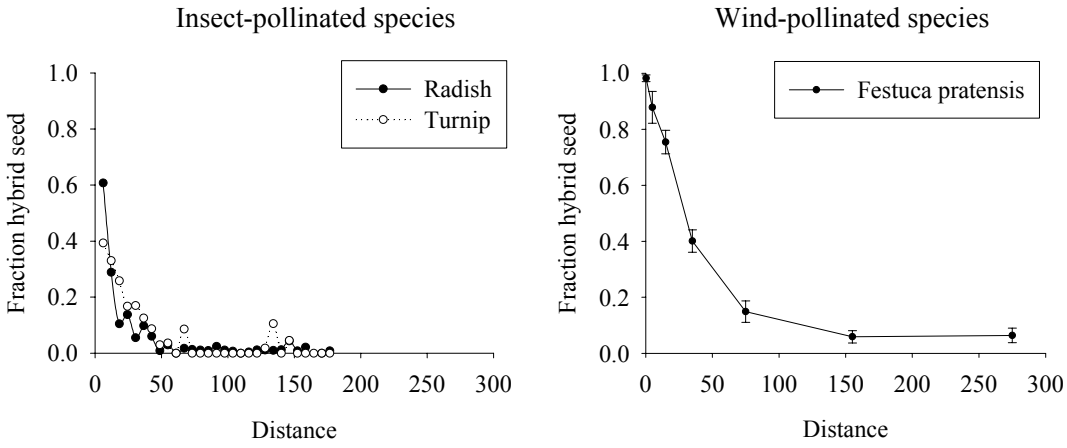


**Figure 3.** Examples of pollen dispersal curves for two insect-pollinated species (radish and turnip) and one wind-pollinated species (*Festuca pratensis*). In all cases, two different morphological types were used, one as a source, the other as a target. Pollen dispersal was measured as the fraction of hybrid seed in the target. More pollen grains would certainly have travelled over these distances than those that resulted in hybrids. Data from Bateman 1947ab and Rognli *et al.* 2000.

### *2.3.3 Comparing wind and insects as dispersal vectors*

Dispersal mechanisms by wind and insects differ considerably (table 2). Despite all these differences, the shape of wind and insect pollination curves relating pollination probability (or pollen frequency) to dispersal distance is generally very much the same, although the distances involved may differ (fig. 3, Bateman 1947c). A large fraction of the pollen lands close to the plant and only a small fraction disperses further, some of which may travel large distances (e.g. Bateman 1947ab, Paterniani & Stort 1974, Klinger *et al.* 1992, Arias & Rieseberg 1994, Lavigne *et al.* 1996, Cresswell *et al.* 2002, Richards *et al.* 1999).

From a modelling perspective, the similarity between the dispersal curves for wind-and insect-dispersed pollen has the great advantage that the same mathematical approach can be used for both.

## 2.4 POLLEN VIABILITY

During the dispersal process, a certain percentage of the dispersing pollen will have lost viability (i.e. died or lost the capability to germinate and achieve fertilisation). Pollen viability is influenced by three main types of factors: *(i)* internal factors, such as pollen metabolism, *(ii)* morphological factors, such as protected anthers or open flowers, and *(iii)* environmental factors, such as humidity, temperature and UV light (Dowding 1987, Dafni & Firmage 2000).

Dafni & Firmage (2000) reviewed pollen viability for 34 species. For wind-pollinated species, an average longevity of 21.5 ± 27.2 hours was found, whereas for insect-pollinated species it was 8.5 ± 10.4 days, but species differed greatly (respective ranges 0.05–72 hours and 1–40 days). The problem with these data is the great diversity in the methods used to measure pollen viability. The exposure conditions for the pollen were variable (field conditions, exposed to open air, greenhouses, growth chambers), as were the tests used to measure viability (seed set,

*in vivo* germination, FCR (fluorescein diacetate reaction) and other tests). These large differences in the methods used can probably partly explain the large ranges. For example, pollen longevity for the species *Oryza sativa* measured in a standard greenhouse was about 20 minutes, while when measured in an open greenhouse it was about four minutes. However, the difference could also be due to the different methods chosen to estimate pollen longevity. It is questionable, whether, on the basis of current knowledge, pollen viability rates of crops growing in the field can be predicted.

With so much information lacking, how can we incorporate pollen viability in a model? One study on pollen storage suggests that pollen longevity follows a normal distribution, with average longevity for most of the pollen grains and decreasing numbers of longer- and shorter-living grains (Hong *et al.* 1999). The same distribution is found for the longevity of seed and fungal spores. As no other information is currently available, this distribution seems the most sensible one to incorporate in the model.

## 2.5 GENE-FLOW BARRIERS

When undertaking cultivation of GM populations in the environment, it is worthwhile considering reducing pollen flow by means of gene-flow barriers. Different gene-flow barriers can be distinguished. One type is a vegetation barrier. Insects can be discouraged from moving between fields by planting a vegetation barrier of a heterospecific species not pollinated by insects around an insect-pollinated crop, thereby limiting pollen flow (Morris *et al.* 1994).

Another kind of gene flow barrier is a trap crop, i.e. a border of plants of the same crop, but not genetically modified. Such traps could "absorb" pollen that disperses out of the GM crop (Morris *et al.* 1994). Indeed, in comparison with bare land, GM pollen flow dispersed by insects outside the trial patch is decreased when a trap crop is present (Morris *et al.* 1994, Reboud 2003). Paterniani & Stort (1974) suggest from an experiment with maize, a typical wind-pollinated species, that the number of plants a pollen grain has to cross is more important than the actual distance. The effectiveness of a trap improves, when the trap-crop area is increased relative to the area of the GM population (Hokanson *et al.* 1997). To limit pollen flow out of agronomic-scale plantings would be extremely difficult, however, because borders are only effective in reducing long-distance dispersal if they are substantially larger than the crop field (Hokanson *et al.* 1997).

The above-mentioned barriers reduce pollen flow between populations, but generally do not prevent it. Genetic isolation mechanisms, such as male sterility, are able to prevent gene flow and would therefore be a more effective barrier (Van de Wiel *et al.* 2003). Nevertheless, when planting small fields for research trial, it is customary to use trap crops or vegetation barriers as an extra precaution measure. In the model, we will not consider the effect of trap crops or vegetation barriers.

## 2.6 CONCLUSIONS

Plant species differ considerably in their pollen-dispersal mechanisms. Some species are mainly self-fertilising and lack mechanisms for pollen dispersal over large distances. These species are expected to have smaller chances of outcrossing than

species that are mainly outcrossing and that have mechanisms for long-distance pollen dispersal. The most important pollen-dispersal mechanisms are dispersal by wind and insects. Generally, pollen dispersed by wind travels further than pollen dispersed by insects. The shape of the dispersal curves, however, is similar, such that, from a modelling perspective, the same mathematical approach can be used for both.

Not all pollen grains will be viable when reaching the stigma of a compatible plant, especially not all of the immigrant grains, which include those from GM populations. In our model, we will assume that most pollen has an average lifespan, with the rest spread either side of the mean.

# CHAPTER 3. FERTILISATION

## 3.1 INTRODUCTION

In the previous chapter, we discussed the pollen-dispersal process, resulting in a certain amount of viable pollen that lands on the stigma of a compatible species (fig. 1A). This dispersal is only effective if the pollen that has reached the stigma is able to achieve fertilisation (fig. 1B). Each pollen grain has to compete with other grains present on the same stigma.

This chapter considers the most important aspects affecting the chances that pollen from GM populations will fertilise plants in wild or cultivated populations.

## 3.2 POLLEN COMPETITION

Often, more pollen grains are present on a stigma than are needed to fertilise all the ovules (Walsh & Charlesworth 1992). These grains have to compete with each other to achieve fertilisation. The number of competing pollen grains produced by the receiving population can have large effect on the success of incoming pollen (Ingram 2000).

In Chapter 2, we saw that some pollen may have lost viability by the time it reaches the target population. In fact, viability is not lost from one moment to the next, but declines gradually over time. For example, germination time and time necessary for the pollen tube to reach the ovule both increase with age (Shivanna *et al.* 1991). This reduces the competitive ability of GM pollen compared to resident pollen, since GM pollen has travelled larger distances and is likely to have aged. However, without information on the time it takes to travel certain distances, no inferences can be made about the magnitude of the age differences.

The position of a pollen grain on the stigma affects its speed of germination. Pollen that lands in the fluid filled cleft between two rows of stigmatic papillae germinates relatively quickly, whereas pollen landing on the papillae themselves germinates more slowly. Pollen that is travelled by insects is often sticking together. Pollen being part of such clump have different fertilisation probabilities, depending on the position in the clump (Thomson 1988).

Another factor affecting competitive ability is the relative timing of arrival. A pollen grain that arrives early, when none or relatively few other grains are present on the stigma, is expected to have higher fertilisation success. Indeed, more seed is sired by pollen that arrives first than pollen that arrives later. Nevertheless, part of the seed can be sired by pollen that arrives two hours after the first pollen (Spira *et al.* 1996). The moment of pollen arrival may thus determine which pollen tubes are more likely to be successful (Walsh & Charlesworth 1992).

A further aspect of pollen competition is tube growth rate, which varies between pollen grains. Differences are found between individuals within a population, between selfing and outcrossing pollen, between pollen from different donors and between pollen from different species (e.g. Snow & Spira 1991a, Walsh & Charlesworth 1992, Skogsmyr & Lankinen 1999). Pollen tubes that grow relatively quickly have an advantage, as they have high chance of reaching the ovule before other pollen tubes, and so have a higher chance of fertilising the ovule (Snow & Spira 1991ab). Direct evidence for the existence of genetic variation in pollen-tube growth rate is lacking

(Chasan 1992), but there are indications that it may be heritable (Skogsmyr & Lankinen 1999). Although we know that pollen grains do differ in growth rate, no generalisations can be made, such as a slower growth rate for selfing than for outcrossing species (Snow & Spira 1991a). There is no reason to suspect that pollen containing GM genes has a different growth rate to pollen without such genes, so we will not include pollen-tube growth rate in our model.

BOX 3. SELF-INCOMPATIBILITY

Self-incompatbility (SI) can be divided in two types, namely heteromorphic and homomorphic. Heteromorphic SI species produce morphologically distinct flowers with regard to their relative style length and anther level, resulting in efficient transfer of intermorph pollen by insects. This mechanistic self-fertilisation barrier augments the biochemical incompatibility that also exists (Ebert *et al.* 1989). Both morphological and biochemical barriers are governed by the same cluster of genes designated *S* and *s*, which consist of some genes, coding for, among other things, style length, anther length and style-pollen incompatibility (De Nettancourt 1977). The genotype of the parent producing the pollen determines the compatibility of two flowers (i.e. compatibility is sporophytically determined). When the parent is short-styled (*Ss)*, its pollen is compatible only with long-styled plants (either when the pollen has genotype *S* or *s*), and vice versa. The genetic control of this cluster ensures that both plant types are present in the same proportion in the population.

Homomorphic SI can be either sporophytic or gametophytic SI. As for heteromorphic SI, rejection of self-pollen in the sporophytic SI is controlled by the diploid genotype of the sporophytic generation (Ebert *et al.* 1989). The control is in the so-called *S*-locus, which is actually a cluster of three tightly-linked loci. Because the plants cannot fertilise themselves, they tend to be heterozygous, carrying two different *S* loci. Pollen will not germinate on the stigma of a flower that contains either of the two alleles in the sporophytic parent that produced the pollen (fig. B3.1A). This holds true, even though each pollen grain – being haploid – contains only one of the alleles (Ebert *et al.* 1989). This is because the active protein on the exine of the pollen is a product of the internal disomic cell layers in the anthers of the parental plant, which contains the products of both alleles.

The gametophytic SI is controlled by the single *S* allele in the haploid pollen grain. A pollen grain will grow in any pistil that does not contain the same allele (fig. B3.1B, Ebert *et al.* 1989). In the gametophytic SI system, as well as in the sporophytic SI system, the *S*-locus is highly polymorphic, containing dozens of different *S*-alleles.

.

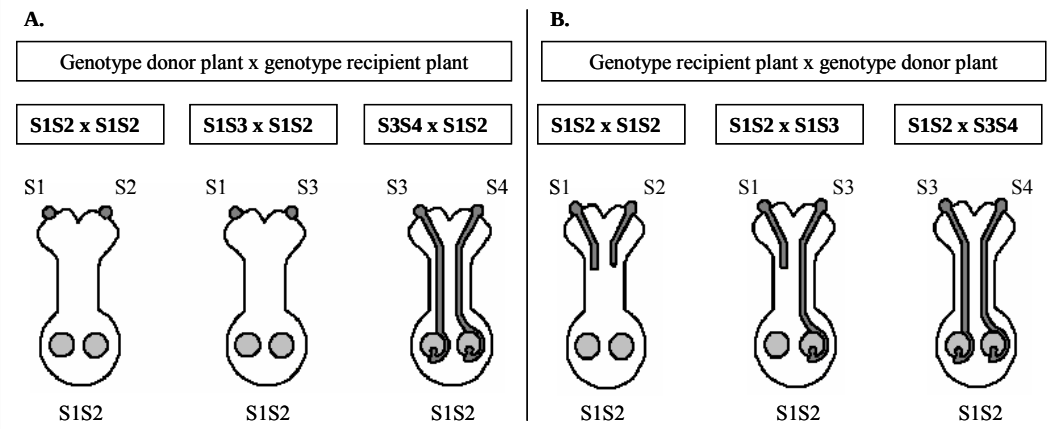


**Figure B3.1.** Schematic depiction of homomorphic self-incompatibility (SI) systems. A. Schematic depiction of a sporophytic SI system. The genotype of the pollen-producing sporophyte (donor plant) determines compatibility of the pollen with the recipient plant. If the recipient plant has at least one allele in common with the donor plant (regardless of whether the pollen grain carries this allele or the other one), fertilisation is prevented. Only when both alleles of the donor plant are different from those of the recipient plant can fertilisation occur. B. Schematic depiction of a gametophytic SI system. If the pollen grain's allele is the same as one of the alleles of the recipint plant, the pollen tube grows for only a few millimetres and fertilisation is unsuccessful. Pollen grains with a different allele, even if the pollen-producing sporophyte

A final important aspect of pollen competition is the impossibility of many plant species to self-fertilise. Most flowering plants have hermaphroditic (bisexual) flowers, which greatly increase the efficiency of insect pollination, because deposition of foreign pollen on the stigma and removal of self-pollen from the anthers are accomplished in a single insect visit. However, bisexual flowers have a disadvantage in the increased risk of self-pollination and self-fertilisation, which can result in inbreeding depression. Many flowering plant species, therefore, have evolved mechanisms to prevent self-fertilisation. Some angiosperms have dispersal and reception of pollen separated in time. Other species have unisexual flowers, having either male or female reproductive organs. Dispersal and reception of pollen can also be separated in space within a flower. All these structural barriers prevent selfing to a greater or lesser degree (Fægri & Van der Pijl 1979).

All these barriers do not preclude fertilisation between pollen and ovule of the same plant, but they make it less likely by reducing the chances that dispersing pollen will land on the plant's own stigma. Other plant species have a system in which pollen grains do land on the plant's own stigma (i.e. self-pollination is present), but fertilisation is prevented, because pollen and stigma are incompatible. Such self-incompatibility (Box 3) is a genetically determined pre-zygotic barrier to fertilisation by self or self-related pollen that eliminates any risk of inbreeding and therefore optimises the potential for outbreeding (Hiscock & McInnes 2003). SI systems prevent self-fertilisation and hence decrease the effective number of competing pollen on a stigma. Particularly in populations with low polymorphism, the effective number of competing pollen on a stigma is expected to be reduced, since plants are expected to share genes. In such a situation, pollen from other populations is expected to be favoured, because it has alleles different from those present in the local population. This can have large effects on crop-to-crop fertilisation. Crops that are harvested before seed set can be highly homogeneous, meaning that pollen grains from a cultivated population are almost all incompatible with their own population. Pollen arriving from other (e.g. GM) crop populations, being of a different type to the target population, enjoys an increase in relative effective numbers. In fact, the same holds for crop pollen entering a wild related population: there will be a small chance of overlap of *S*-alleles.

## 3.3 CONCLUSIONS

Fertilisation is a complicated process. Important mechanisms that influence fertilisation success are reduced viability, time of arrival, the exact place where the pollen lands and the presence of self-incompatibility. Some of these effects are too complicated to include in a model. For example, the relative time of arrival not only depends on the distance travelled by the pollen, but also on the timing of its emission, which can differ between individual flowers.

In a model, the process of interest needs to be simplified. Walklate *et al.* (2004) simulated fertilisation probabilities by considering the effective deposition of pollen from the GM population and expressing this as a proportion of the total effective pollen deposition. We propose a similar kind of approach in our model, but envisage that it will be difficult to get reliable data bearing on this.

# CHAPTER 4. INTROGRESSION

## 4.1 INTRODUCTION

After the production of a hybrid [1] seed containing modified DNA, several scenarios are possible. One option is that the hybrid is unable to establish, or some plants are able to establish but are not able to backcross or persist. In this case, outcrossing has occurred, but it has no consequences for the wild or cultivated population. A second possibility is that the hybrid does establish and persists as a new species. This can occur in several ways. (*i*) The hybrid can spread vegetatively. In this case, only one successful hybrid needs to establish. (*ii*) Several hybrids develop and cross with each other, producing a new population. (*iii*) A sterile allopolyploid hybrid becomes fertile through chromosome doubling (Box 4). All three processes can affect a wild or cultivated population by competing with them.

In this chapter, we will not deal with these scenarios, but will instead consider a fourth possibility, namely introgression. The reason for restricting our focus in this way is that introgression is the only process that leads to the incorporation of modified DNA into the genome of wild or cultivated populations.

BOX 4. RECOVERY OF AN ALLOPOLYPLOID

Many crops are polyploid while their wild relatives are diploid. Crosses between species with different ploidy levels give rise to allopolyploid hybrids, i.e. hybrids that contain a number of chromosomes intermediate to those of the parental species (fig. B4.1). When these hybrids reproduce, problems with pairing of homologous chromosomes typically occur in meiosis, making them sterile. Sometimes, however, crosses between species with different ploidy levels can be successful, as in *Spartina* (Ellstrand 2003). The hybrid *S.* x *townsendii* is the result of a cross between *S. maritima*, which has 30 chromosome pairs (2n = 60), and *S. alterniflora*, which has 31 chromosome pairs (2n = 62). *S.* x *townsendii* had a chromosome number intermediate to the two parental species (2n = 61), and therefore was not able to reproduce sexually, but it was able to spread vegetatively. Out of this sterile hybrid the new species *S. anglica* evolved by chromosome doubling (2n = 122).

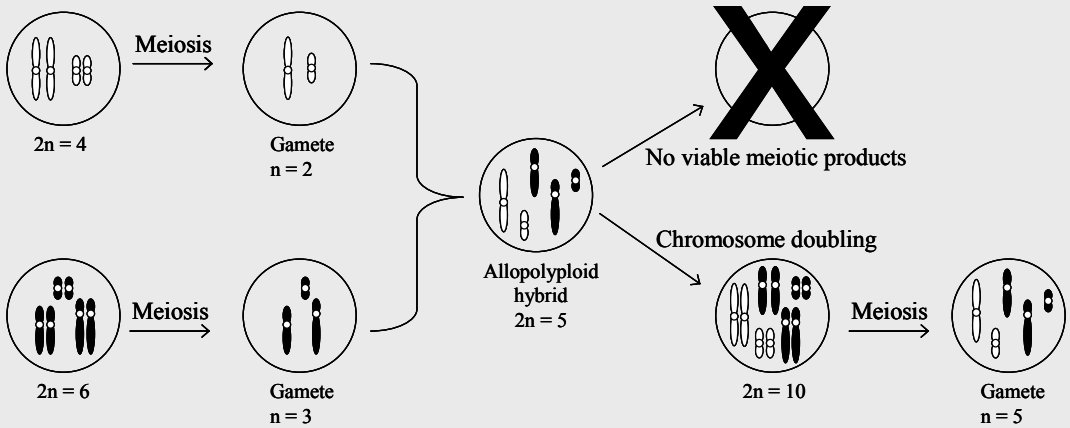


**Figure B4.1.** Schematic depiction of the recovery of a sterile hybrid from a cross between two species with different numbers of chromosomes. After spontaneous chromosome doubling, chromosomes are able to pair and create gametes.

[1] With hybrid is meant offspring produced by crossbreeding between a pollen originating from a GM source population and an ovule originating from a non-GM target population.

## 4.2 INTROGRESSION

Introgression can be defined as the permanent incorporation of one or more genes from the gene pool of one taxon into the gene pool of another taxon (mostly different species), through hybridisation and repeated backcrossing with one of the parental populations. In order to predict introgression chances, it is necessary to decide when you consider a gene to be introgressed. This could be, for example, when the gene is fixed in the population, but such an approach might take too many generations to be workable. A more workable option would be to estimate chances that the modified DNA persists in the population for a certain number of generations by means of backcrossing.

To become introgressed, first the modified DNA should become established[1], something that is highly dependent on stochastic processes. In the intial stages, while the modified DNA is present in only very low frequencies, there is only a small chance that it will become established in the recipient population. To have a reasonable chance to establish, hybridisation must occur regularly. The cumulative probability of individual hybrids becoming established determines the final chance of establishment, so the more hybrids that develop, the higher the overall establishment chances.

After successful establishment, the modified DNA must be able to persist. Here, deterministic factors and processes become important. One of these is the fitness of hybrids and backcrosses. Hybrids can have either a higher fitness (hybrid vigour, e.g. in crosses within *Raphanus sativus*, within *Oryza sativa* and between *Brassica napus* and *B. rapa*), lower fitness (outbreeding depression, e.g. in crosses between *Brassica napus* and *Hirschfeldia incana*, between *Raphanus sativus* and *R. raphanistrum* and within *Helianthus anuus*), or the same fitness as their parents (e.g. in crosses between *Cucurbia pepo* and *C. texana* and between *Sorghum bicolor* and *S. halepense*) (Ellstrand 2003). A complicating factor is that the effect of the modified DNA on the persistence of the plant is not always known beforehand, which makes it hard to estimate introgression probabilities. The fitness effect depends not only on phenotypic characteristics caused by the gene (such as herbicide resistance), but also on other features, such as dominance, association with deleterious crop alleles or traits and location on the chromosome (Stewart *et al* 2003). For example, certain characteristics of domesticated species may behave recessively in a cross with a wild species. These characteristics do not become expressed in first-generation hybrids. Therefore, first-generation hybrids having a high fitness does not always imply high fitness for later-generation hybrids (Groot *et al*. 2004).

Besides the effect of the modified DNA on the persistence of hybrids and backcrosses, the environmental conditions are important. If, for example, fields with hybrids carrying an herbicide-resistance gene were sprayed with herbicide, this would create a strong selection pressure favouring the gene which would not have existed had the fields not been treated.

Introgression into cultivated populations is a different kind of problem to introgression into wild populations. Most crops are harvested each year. Farmers are then concerned about contamination levels of the seed. Crops cultivated to be used as food or intended for seed production are allowed to be contaminated to a certain level, which height is continuous under discussion. Introgression will only occur, when

[1] The modified DNA is considered established in a population, when the chances that it will persist in the populations are not dependent on merely stochastic processes, but mainly on the fitness of the individual plants carrying the modified DNA.

contaminated seed is repeatedly harvested and plants grown from this will be fertilised next year, so in fact when the farmer collects his own seed for next year's sowing.

## 4.3 MODELLING APPROACH

When hybridisation between a GM plant and its wild or cultivated relatives has occurred, the modified DNA will be present in very low frequencies in the population. Whether the modified DNA will establish is initially a mainly stochastic process. Therefore, a model should use initial establishment probabilities that depend on stochastic processes.

When the modified DNA is established, it should be able to persist. To estimate this part, different modelling approaches can be taken. One is a population genetic approach (e.g. Van Raamsdonk & Schouten 1997, Haygood *et al.* 2003). Such models are based on changes in allele frequencies from one generation to the next depending on the fitness of the different genotypes and on the number of alleles received from the GM population every generation.

Another approach is to divide species into categories with high or low chances of introgression. Stewart *et al.* (2003) based different categories on experimental knowledge of hybridisation and introgression. Species for which no molecular evidence of introgression has been found were considered very low-risk crops, while species that hybridise with wild relatives and for which there is good molecular evidence for introgression were considered to be high-risk crops. Hancock (2003) based the categories on fitness characteristics of the modified DNA combined with characteristics determining invasiveness. Decisions about invasiveness were based on the number of weediness traits carried by the GM crop and the recipient population (traits such as broad germination requirements, high seed longevity, rapid growth to flowering, seed production in variable environments and vigorous vegetative reproduction). The potential impact of the modified DNA could be ranked by its likely effect on reproductive success, ranging from advantageous to neutral to detrimental. Gressel & Rotteveel (2000) developed a detailed decision-tree-based risk-assessment categorisation methodology for GM herbicide-resistant crops.

## 4.4 CONCLUSIONS

Groot *et al.* (2004) recently reviewed the current knowledge of hybridisation and introgression between GM or conventional crops and their wild relatives. They conclude their report with a long list of knowledge gaps, showing that much experimental work needs to be done before introgression probabilities can be estimated reliably. Therefore, we did not include introgression in our model.

Nevertheless, we would recommend that future models incorporate introgression as soon as these knowledge gaps can be filled, because "the general conclusion with respect to the phenomenon of gene flow between crops and wild relatives is that although chances may vary, in many crop-wild relative complexes sooner or later gene flow will occur. Incorporation of crop genes into recipient taxa will occur through further introgression processes after initial hybrid formation" (Groot *et al.* 2004).

# CHAPTER 5. MODELLING APPROACHES

## 5.1 INTRODUCTION

Before starting with a modelling endeavour, one has to realise that all models have their inherent limitations. The user might want a model that is at the same time simple, robust, realistic, precise, reliable and discriminating. However, some of these desirable properties are inherently incompatible. For example, the outcome of a robust model is not much affected by a small change in parameters. A discriminating model, however, is expected to reflect precisely such differences. Accordingly, a model cannot be robust and discriminating at the same time. Depending on the purpose of a model, the developer has to decide which properties are more important and which less. In Box 5, a short overview is given of different types of models and their properties.

BOX 5. TYPES OF MODELS AND THEIR PROPERTIES

Roughly, models can be classified in three types: conceptual, mechanistic and statistical (e.g. Lavigne *et al.* 2004). A short description of these types will be given.

**Conceptual** models would probably strive to describe pollen dispersal in a relatively simple way, appealing to intuition and without putting too much emphasis on the details of the process. Such an approach would most probably be based on a negative exponential distribution of pollen away from its source, for two reasons. First, this approach is mathematically simple and elegant. Second, it has a simple statistical interpretation: if pollen is moving at a given speed and in a given horizontal direction, an exponential distribution is generated if the probability of landing is constant and independent of the distance from the source.

A **Mechanistic** model of pollen dispersal by wind would most probably be based on the physical principles of transportation by air. Such a model would probably take into account factors such as wind direction, horizontal wind speed, thermal turbulence and several weather and landscape parameters having influence on the process. Given information on all these processes, a mechanistic model could then derive a pollen dispersal curve. It is by no means sure that this curve would be an exponential one. A mechanistic model of pollen dispersal by insects would mostly be based on insect behaviour, taking into account such factors as flight distance and direction of the insect involved, pollen load and pollen carry-over. For examples of mechanistic models of seed dispersal by wind see Tackenberg (2001, 2003) and Tackenberg *et al.* (2003) and for those of dispersal by insects see Morris (1993), Cresswell *et al.* (1995, 2002) and Cresswell (2003).

With a **statistical** description of the process, a given data set is used, e.g. an experimentally determined distribution of pollen around a source. The statistical approach then fits some curves with simple and well-known statistical properties through the data and chooses that statistical model that provides the best balance between goodness-of-fit and number of parameters that have to be estimated from that data (Myung *et al*. 2000).

This rough distinction does not mean that any given model fits perfectly to any of the three categories described above. For example, no model can take all mechanisms into account (otherwise the model would be as complex as reality). Hence, a mechanistic model is only mechanistic to a certain degree. Similarly, a conceptual model may incorporate elements that are based on statistical analysis. For example, it would be easy to replace a negative exponential pollen distribution by an inverse power law if it turns out that power functions give a better description of the process.

Figure B5.1 illustrates how the type of model relates to the properties it typically has. Conceptual models are mainly intended for giving qualitative insights guiding intuition, and are therefore less suitable for making quantitative predictions and providing guidance to management decisions. Statistical models can easily lead to a description which fits well to the data, but for unknown reasons. These models therefore contribute little to a better understanding of the processes involved. This may have important implications for management decision, since these decisions often involve considering situations for which reliable data are not yet available.

CONTINUATION BOX 5. TYPES OF MODELS AND THEIR PROPERTIES

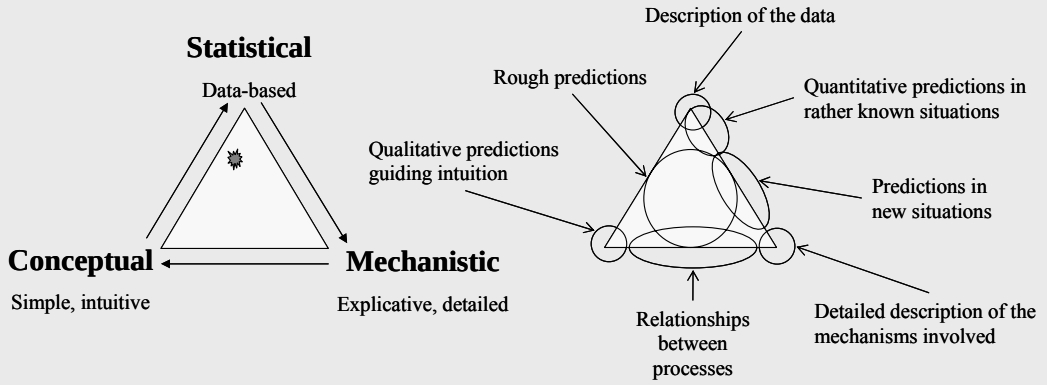


**Figure B5.1.** *Left triangle:* Distinction of three types of models which have different purposes and properties. The rough distinction into conceptual, mechanistic and statistical models does not mean that a given model fits exactly into one of these categories. Most models fall in between, represented by the symbol in the triangle. *Right triangle:* Illustration of some important aspects related to a proper choice of model. After Lavigne *et al.* (2004).

In such situations one has to extrapolate from the given data to unknown situations and such an extrapolation can be risky if it is based on a (statistical) model whose mechanistic foundation is not known. In the context of a mechanistic model, extrapolation from a known situation to unknown ones is much less risky, at least in those situations where the underlying mechanisms are well understood. The problem with a mechanistic approach is that typically a large number of mechanisms are responsible for a given process and these processes may interact in a rather complicated way. Fully mechanistic models therefore tend to be highly complicated. Moreover, these models are often dependent on many parameters on which information is not readily available in a given situation.

Concerning the question of genetic exchange, the model we have developed is mainly statistical, since otherwise reliable quantitative predictions are hard to obtain. However, as indicated above, one has to be aware of the extrapolation problem. It might therefore be useful to develop in parallel a suite of "mechanistic models of intermediate complexity" (models incorporating a few mechanisms) in order to judge the reliability of the conclusions derived from the statistical approach.

## 5.2 MODELLING APPROACH

Figure 4 gives an overview of the mathematical model we propose, and have partly developed, for estimating the probabilities that GM pollen will land in populations of compatible species and achieve fertilisation in such a population. The model consists of three modules. The first module addresses the question: how does pollen, originating from a GM source population, disperse over the landscape? In this module, pollen dispersal of a source population will be simulated. The second module addresses the question: what is the expected frequency of seeds in a target population that is fertilised by pollen originating from a given GM source population? In this module, the percentage of seeds that would originate from a cross between pollen from the source population and ovules from the target populations is calculated, thus giving an estimation of the contamination level of the target population with DNA from the GM source population. The third module addresses the question: which percentage of seed of a collection of target populations is fertilised by pollen originating from a given GM source population? In this module, many target populations are situated in a landscape with one GM source population. Depending on population size, proximity to other target populations and to the GM source

population, different target populations are expected to be contaminated to a greater or lesser extent.

In the following paragraphs, we will go into the different modules of the model in more detail.

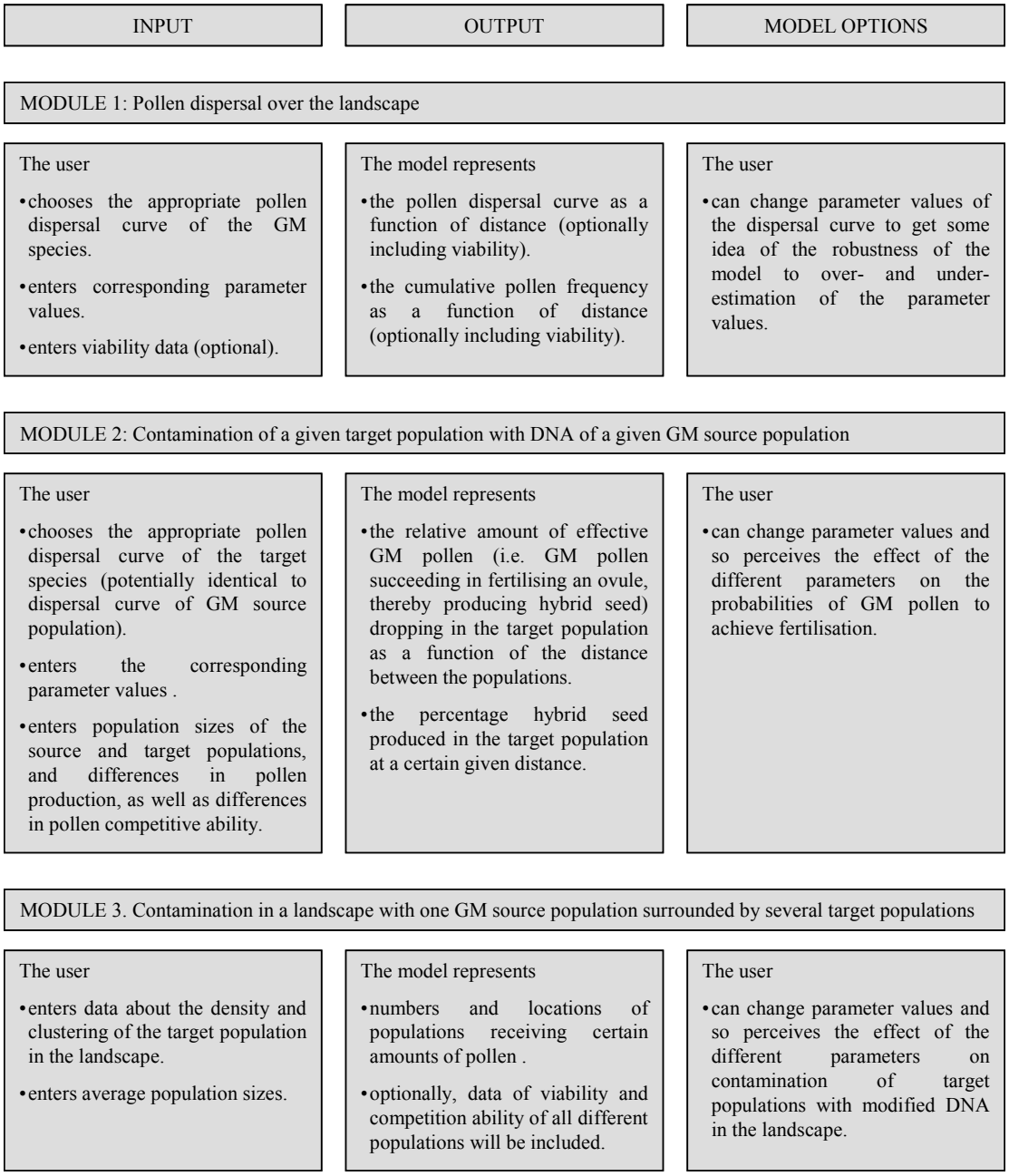


**Figure 4.** Schematic depiction of the mathematical model that is proposed. In the first column (user), the parameter values that have to be entered by the user are given. In the second column (output), the model calculations are represented. In the third column (model options), the possibility of the user to change parameters is given.

BOX 6. EQUATIONS USED TO DESCRIBE POLLEN DISPERSAL CURVES

Many different equations can be used to describe the dispersal pattern of pollen through the landscape. Here, different equations are discussed. The equations are defined as a probability density. The mathematical definition of a continuous probability density function, *f(D)*, is a function that satisfies the following properties:

1. The probability $P$ that the distance $D$ is between two points $D_1$ and $D_2$ is

$$P(D_1 \le D \le D_2) = \int_{D_1}^{D_2} f(D)\, dD$$

2. The integral of the probability function is one, that is

$$\int_{-\infty}^{+\infty} f(D)\, dD = 1$$

What does this actually mean? Since continuous probability functions are defined for an infinite number of points over a continuous interval, the probability at a single point is always zero. Probabilities are measured over intervals, not single points. That is, the area under the curve between two distinct points defines the probability for that interval.

Below, we consider some of the different types of equations used for pollen dispersal curves.

1. One-parameter families of dispersal curves

**Negative exponential distribution**

The standard negative exponential distribution is described by

$$f(D) = a \exp(-\lambda D)$$

where $\lambda$ is a shape parameter and $a$ is a scale parameter. If it is used to describe the two-dimensional distribution of pollen around a point source, the parameter $a$ is determined by $\lambda$ and given by

$a = \frac{\lambda^2}{2\pi}$ (see appendix A1). Hence, $f(D) = \frac{\lambda^2}{2\pi} \exp(-\lambda D)$

The parameter $\lambda$, characterising the shape of the dispersal curve, has a simple relationship to the mean dispersal distance $\bar{D}$ and the variance in dispersal distance:

Mean dispersal: $\bar{D} = \frac{2}{\lambda}$

Variance: $\text{var}(D) = \frac{6}{\lambda^2} - \bar{D}^2$

**Inverse power law**

The standard inverse power law is given by

$$f(D) = a D^{-\lambda}$$

where $\lambda$ is a shape parameter and $a$ is a scale parameter. Since the inverse power law becomes unreliable towards $D = 0$, $D + 1$ is used instead of $D$. If the inverse power law is used to describe the two-dimensional distribution of pollen around a point source, the parameter $a$ is determined by $\lambda$ and given by

$a = \frac{(1-\lambda)(2-\lambda)}{2\pi}$ (see appendix A2). Hence,

CONTINUATION BOX 6

$$f(D) = \frac{(\lambda-2)(\lambda-1)}{2\pi}(D+1)^{-\lambda} \quad \text{for } \lambda > 2$$

The parameter $\lambda$, characterising the shape of the dispersal curve, has a simple relationship to the mean dispersal distance $\overline{D}$ and the variance in dispersal distance:

Mean dispersal: $\overline{D} = \frac{2}{\lambda-3}$ for $\lambda > 3$

Variance: $\text{var}(D) = \frac{6}{(\lambda-3)(\lambda-4)} - \overline{D}^2$ for $\lambda > 4$

**Uniform distribution**

The standard uniform distribution is given by

$$f(D) = \frac{a}{D_{\max}}$$

where $a$ is a scale parameter and $D_{max}$ the maximum dispersal distance. If this equation is used to describe the two-dimensional distribution of pollen around a point source, the parameter $a$ is given by

$a = \frac{1}{\pi D_{\max}}$ (see appendix A3). Hence, $f(D) = \frac{1}{\pi {D_{\max}}^2}$

The mean dispersal distance $\overline{D}$ and the variance in dispersal distance are given by

Mean dispersal: $\overline{D} = \frac{2}{3} D_{\max}$

Variance: $\text{var}(D) = \frac{1}{2} D_{\max} - \overline{D}^2$

2. Two-parameter families of dispersal curves

**Equation from the exponential power family**

The standard equation from the exponential power family is given by

$$f(D) = a \exp\left(-(\lambda D)^b\right)$$

If it is used to describe the two-dimensional distribution of pollen around a point source, the parameter $a$ is determined by $\lambda$ and given by

$a = \frac{\lambda^2 b}{2\pi\,\Gamma(2/b)}$. Hence, $f(D) = \frac{\lambda^2 b}{2\pi\,\Gamma(2/b)} \exp\left(-(\lambda D)^b\right)$

The parameter $\lambda$ characterises the shape of the curve. $\Gamma(n)$ is the gamma function, in which $\Gamma(n+1) = n!$ (for $n = 1, 2 \ldots \infty$). By adding the extra parameter $b$, the equation has adjustable kurtosis (fig. 7). This parameter has following properties:

CONTINUATION BOX 6.

When $b = 1$, the curve follows a negative exponential distribution,
When $b < 1$, the tail of the curve is fat-tailed compared to a negative exponential distribution,
When $b > 1$, the tail of the curve is thin-tailed compared to a negative exponential distribution

The parameters $\lambda$ and $b$ have following relationship to the mean dispersal distance $\overline{D}$, the variance in dispersal distance var($D$) and the kurtosis $\kappa$:

Mean dispersal: $$\overline{D} = \frac{\Gamma(3/b)}{\lambda\,\Gamma(2/b)}$$

Variance: $$\text{var}\,(D) = \frac{\Gamma(4/b)}{\lambda^2\,\Gamma(2/b)}$$

Kurtosis: $$\kappa = \frac{\Gamma(6/b)\Gamma(2/b)}{\Gamma^2(4/b)}$$

Many other dispersal curves defined by two parameters are used to estimate pollen dispersal patterns. Below, three of them are listed. See Austerlitz *et al.* (2004) for detailed information about these equations.

**Equation from the Weibull family** $$f(D) = \frac{b\,(\lambda\,D^{(b-2)})}{2\pi\,\lambda^{(-b)}}\,e^{-(\lambda D)^b}$$

**Equation from the bivariate Student's t family** $$f(D) = \left(\frac{\lambda^2\,(b-1)}{\pi}\right)\left(1+\lambda^2 D^2\right)^{-b}$$

**Equation from the geometric family** $$f(D) = \frac{\lambda\,(b-2)(b-1)}{2\pi}\left(1+\lambda\,D\right)^{-b}$$

3. Dispersal curves defined by more than two parameters

**Summing two functions**

Sometimes, experimentally determined dispersal curves can best be viewed as a weighted average of two dispersal curves $f_1$ and $f_2$ (e.g. the exponential and power curve):

$$f(D) = \alpha\,f_1(D) + (1-\alpha)\,f_2(D)$$

This equation is described by three parameters: the weight factor $\alpha$, the average of equation 1, $\overline{D}_1$, and the average of equation 2, $\overline{D}_2$. The mean dispersal distance and variance can be calculated as follows:

Mean dispersal: $$\overline{D} = \alpha\,\overline{D}_1 + (1-\alpha)\,\overline{D}_2$$

Variance: $$\text{var}\,(D) = \alpha\,\text{var}\,(D_1) + (1-\alpha)\,\text{var}\,(D_2) + \alpha(1-\alpha)\,(\overline{D}_1 - \overline{D}_2)^2$$

### *5.2.1 Module 1: Pollen dispersal over the landscape*

In Chapter 2, we saw that the shape of wind and insect pollination curves relating pollination probability (or pollen frequency) to dispersal distance is very much the same. A large fraction of the pollen lands close to the source plant and only a small fraction disperses further, of which some travels over large distances (fig. 3).

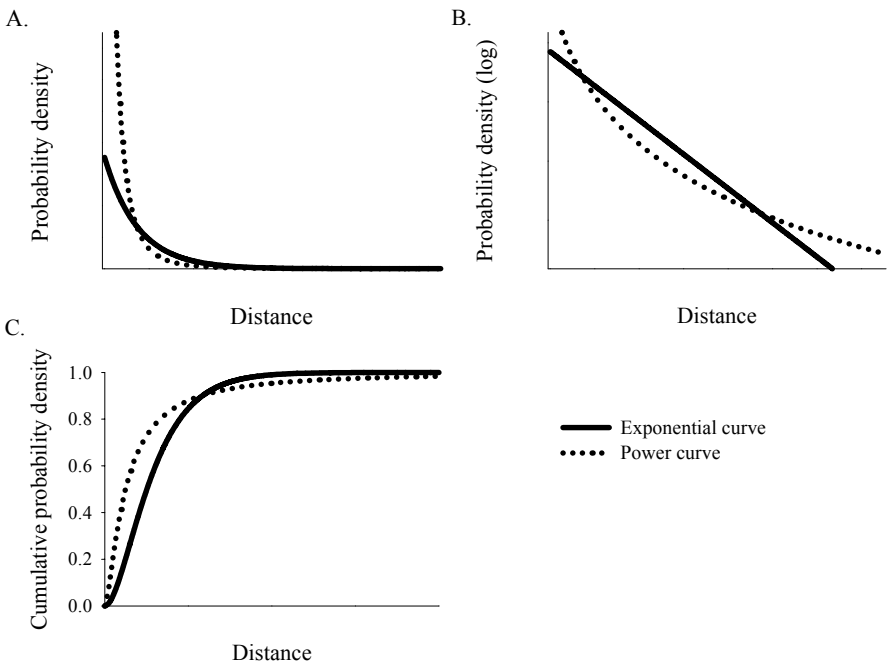


**Figure 5.** Graphic representation of the negative exponential curve and inverse power law with the same average dispersal distance on A. a linear scale, B. a semi-log scale and C. represented as the cumulative probability density. The power curve is more leptokurtic than the exponential curve, predicting higher pollen densities close to the donor plant (very short dispersal distances) and at large dispersal distances and lower pollen densities at intermediate dispersal distances.

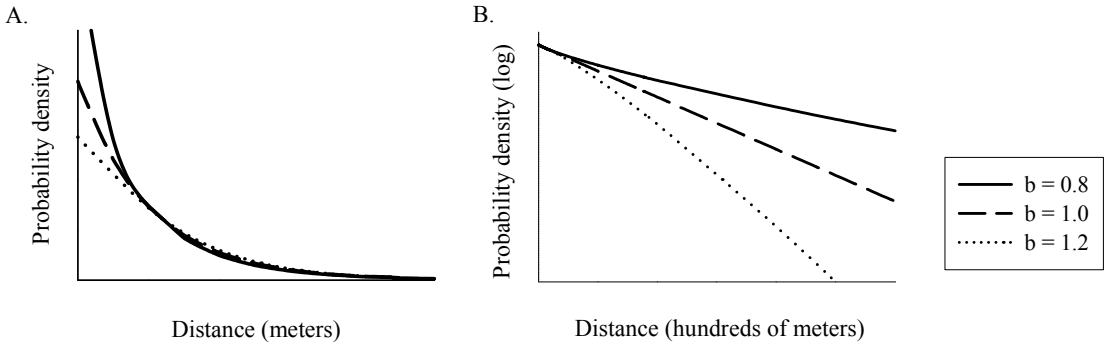


**Figure 6.** Different exponential power curves with the same average dispersal distances on A. a linear scales and B. a semi-log scale. When b > 1 the curve is thin-tailed, and when b < 1 it is fat-tailed compared to the standard exponential curve (b = 1).

Many different equations are used to describe pollen dispersal curves. The most promising ones are the negative exponential distribution (NED) and the inverse power law (IPL, Box 6). Having the same average dispersal distance, the NED predicts higher pollen frequencies close to the donor plant and lower frequencies at larger distances compared with the IPL (fig. 5AB). Although commonly used, both curves seem to underestimate pollen frequencies at large distances; pollen dispersal curves are generally more leptokurtic (i.e. more fat-tailed) than predicted by the NED (Nurminiemi *et al.* 1998, Austerlitz *et al.* 2004).

The NED and IPL are both determined by one parameter and therefore restricted. More realistic dispersal curves are defined by two parameters (Box 6). One of these comes from the exponential power family. In this equation, the 'fatness' of the tail of the dispersal distribution is determined by the kurtosis parameter *b*. When $b < 1$, the tail of the curve is fat-tailed, when $b > 1$, the tail of the curve is thin-tailed

compared to the exponential distribution (fig. 6). Austerlitz *et al.* (2004) used this exponential power curve to estimate pollen dispersal curves using genetic markers. They were able to estimate correctly the general trend of the curve, i.e. fat-tailed or thin-tailed. The same equation was used by Clark (1998) and Clark *et al.* (1998) to estimate seed dispersal curves. Other functions that are defined by two parameters are functions of the Weibull family and the geometric and 2Dt families (Box 6).

A different approach is not to estimate the whole dispersal curve at once, but to cut the curve into two parts and estimate each part separately. Lavigne *et al.* (1998) applied this method. They fitted dispersal curves to experimental data from oilseed rape (*Brassica napus*) and found that 55 per cent of the pollen dropped within a few metres, the other 45 per cent landing at larger distances. This latter part of this distribution could best be described by a negative exponential function. Other possibilities are to describe the latter part by a power function or a uniform distribution (fig. 7). The mathematically correct way to use this method is to take the weighted average of the two curves involved (e.g. the exponential and power curve, Bullock & Clarke 2000). Three parameters are needed to do so, namely the average dispersal distances for both equations and a weighting factor for the curves (box 6). The weighting factor corresponds to the proportion of pollen governed by the curve in question. This method can only be used when it is clear that two different equations are involved, as was the case in the experimental study by Lavigne *et al.* (1998).

How should one choose among these competing models of the same phenomenon? Here we enter the realm of model selection. The model that fits observed data sufficiently well (i.e. is descriptively adequate) in the least complex way (i.e. using fewest parameters) should be preferred (Myung *et al.* 2000). A complex model with many parameters and highly flexible form can often fit data better than a simple model with few parameters; however, beyond a certain point, the improved fit from including extra parameters does not outweigh the increased complexity of the model. Most of the distributions mentioned above are estimated by a few parameters, which makes them relatively simple, but it is questionable that they will describe the data sufficiently well. For more information about model selection see Burnham & Anderson (2002), Pitt & Myung (2002) and Johnson & Omland (2004).

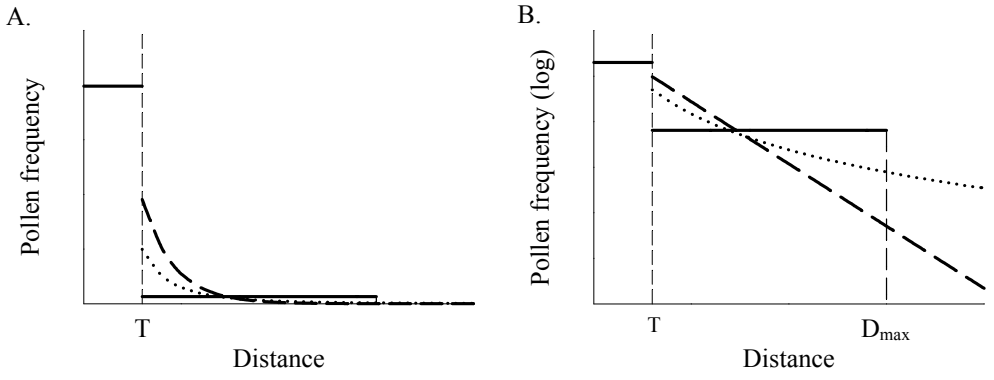


**Figure 7.** Schematic representation of how the pollen distribution pattern is estimated with three different equations. Here, the first part of the pollen dispersal curve is estimated with a uniform distribution (solid line), the second part of the curve (the tail) is estimated by a negative exponential curve (dashed line) an inverse power law (dotted line) and a uniform distribution (solid line). To be able to estimate the pollen distribution pattern in this way, there should be a clear 'cutting point', i.e. it should be clear where to end the first equation and to start the second one (T in this figure). Furthermore, it should be known what fraction of the pollen is described by the first equation and what fraction by the second equation.

For many species the pollen distribution is not known and therefore must be estimated. In the model, the uniform distribution is included to make a kind of worst-case estimation. This distribution assumes that a constant frequency of pollen lands at every distance. The uniform distribution can only be used as a truncated distribution, since the amount of pollen is limited; at a certain distance all pollen will have landed. Therefore, a maximum dispersal distance should be set. The pollen will be distributed evenly over the area below this distance. Choosing a large maximum dispersal distance implies little pollen per unit area (since the total area is large); choosing a small maximum dispersal distance implies a larger amount of pollen per unit area. Using the uniform distribution, an overestimation of the pollen frequency is made in the tail of the curve and an underestimation at the short dispersal distances. The uniform distribution can also be used to describe only the tail of the curve. Some authors (e.g. Paterniani & Stort 1974) suggest that the latter scenario is most realistic.

How does this work in the model? The negative exponential distribution (NED), the inverse power law (IPL) and the uniform distribution are included in the model as possible alternatives. The user can also choose to enter an equation that is described by one parameter. In some cases, it may be better not to describe the whole curve, but to estimate only the tail of the curve (fig. 7). For all curves, the parameter values have to be entered. This means the lambda for the NED and the IPL (Box 6) and the maximum dispersal distance for the uniform distribution. If the user chooses to estimate only the tail of the curve, the weight factor should be entered (i.e. the proportion of pollen that lands in this part of the curve). When the appropriate parameters have been entered, the dispersal pattern can be shown graphically, e.g. as a 'standard' pollen dispersal curve or as a cumulative pollen dispersal curve, both as a function of distance. The standard pollen dispersal curve plots the pollen frequency as a function of the distance. The cumulative pollen dispersal curve plots the frequency of all pollen that has dropped up to and including that distance (fig. 5C).

### *5.2.2 Module 1: Viability of the pollen*

It takes some time for pollen to travel from the releasing plant to a recipient population. During this time, part of the pollen is expected to have lost viability. Does a common distribution exist that describes the loss of viable pollen in time? Hong *et al.* (1999) found that, over time, the changing fraction of surviving stored pollen of *Typha latifolia* followed a negative cumulative normal distribution (Box 7). To our knowledge, this is the only study so far that has tried to find a distribution describing the loss of viable pollen. Other studies analysing pollen *in vivo* (e.g. James & Knox 1993, Fernando & Cass 1997, Aylor *et al.* 2003) seem consistent with the results of Hong and colleagues.

In the model developed, we used this negative cumulative normal distribution to describe the frequency of viable pollen. This curve is defined by two parameters, mean viability and the variation around the mean (Box 7). As knowledge is lacking about the time it takes a pollen grain to arrive at a compatible stigma, we define this equation as a function of the distance.

Besides the negative cumulative normal distribution, it is possible for the user to enter another function that is described by one parameter. The viability data are taken into account in the first module of the program (fig. 4).

BOX 7. NEGATIVE CUMULATIVE NORMAL DISTRIBUTION

The normal (or Gaussian) distribution is an extremely important probability distribution in many fields. It is actually a family of distributions of the same general form, differing only in their location and scale parameters: the mean ($\mu$) and standard deviation ($\sigma^2$). The distribution is symmetric. The probability density function *P(D)*, with *D* being distance, is:

$$P(D) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(D-\mu)^2}{2\sigma^2}\right)$$

Considering the mortality of dispersing pollen grains, this distribution gives the frequency of individual deaths per distance.

In the model we developed, we used the negative cumulative normal distribution as a function of distance to calculate the probability of pollen surviving over the corresponding distance. The cumulative distribution is a special way to represent the normal distribution. Generally speaking, cumulative distribution functions give the probability that the variable takes a value less than or equal to *x*. In our situation, the variable is the probability of death (or survival) and *x* is the distance. The cumulative distribution function of the normal distribution does not exist in a simple closed formula. It is computed numerically.

The cumulative **positive** normal distribution gives the probability of a pollen grain to have died before or at the corresponding distance. This curve is called the **mortality** curve. The negative cumulative normal distribution is the opposite of the mortality curve (1 minus the mortality curve) and is called the **survival** curve, as it gives the probability of surviving till the corresponding distance (fig B7.1A). The location and scale parameter define the exact shape of the survival curve (fig B7.1B).

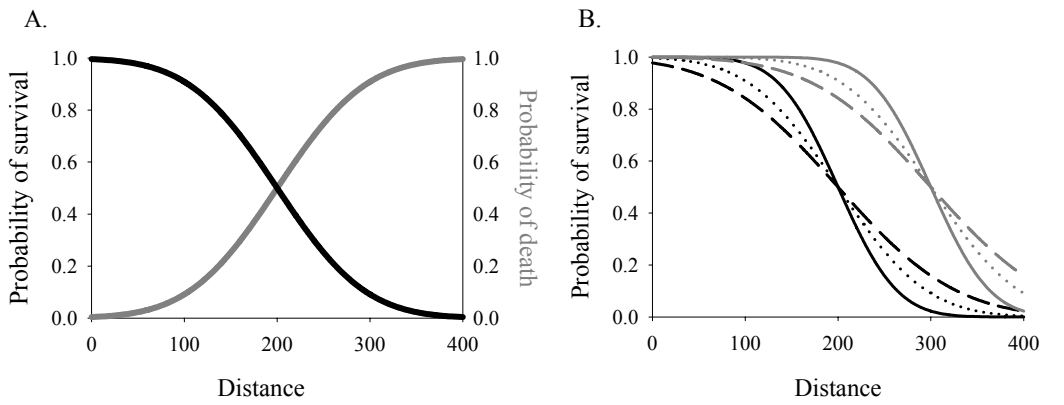


**Figure B7.1.** A. The mortality curve (grey) gives the probability that a pollen grain will be dead by the time it reaches the given distance. The survival curve (black) gives the probability that the pollen grain will still be alive by the time it reaches the given distance. Since the vertical axis is a probability, it must fall between zero and one. The horizontal axis is the allowable domain for the given probability function. B. The location and scale parameters define the exact shape of the survival curve. An increase in mean viability, i.e. an increase in the distance at which half of the pollen is viable and half of the pollen dead, results in the same frequency of pollen surviving to higher distances (compare the black lines, with a mean of 200, with the grey lines, with a mean of 300). The standard deviation (sd) around the mean determines the rate of loss of viability. A higher standard deviation results in a lower rate (compare the solid line (sd = 50), dotted line (sd = 75) and dashed line (sd = 100) within each colour).

### *5.2.3 Module 2: Fertilisation*

In the second module, the model estimates contamination levels of a given target population with DNA from the GM source population. In a target population, two types of pollen can land: pollen from the target species itself and pollen from the GM source population. The higher the relative number of GM pollen grains landing on a stigma compared to the pollen grains of the resident population, the higher the

fertilisation chances of these grains. The amount of pollen containing modified DNA that reaches the target population depends on the distance between source and target population, as well as on population sizes and the number of pollen grains produced by the different populations. With this information, we can estimate the relative amount of pollen containing modified DNA that lands on a stigma in the target population.

Perhaps resident pollen will have a higher competitive ability than GM pollen; for example, because GM pollen, although closely related to the target species, is heterospecific, lowering its compatibility, or because the GM pollen will have aged more than resident pollen by the time it reaches the stigma. In other situations, resident pollen might have a lower compatibility than GM pollen, for example due to self-incompatibility systems active within the target population. Resident pollen can be subdivided in several types: self-pollen, pollen from other flowers within the target population and pollen from populations of the same species as the target population growing nearby. Table 3 lists the effects of different mechanisms on the relative compatibility of different types of resident pollen.

**Table 3.** The effects of different mechanisms on the competitive ability of GM pollen ($D_{gm}$), self-pollen ($D_{self}$), pollen from other flowers within the target population ($D_{in}$) and pollen from populations of the same species as the target populations growing nearby ($D_{out}$). The competitive ability of the GM pollen is set to one, with that of other pollen given relative to this.

| Pollen types | $D_{gm}$ | $D_{self}$ | $D_{in}$ | $D_{out}$ |
|---|---|---|---|---|
| Effect on compatibility factor $\eta$ | $\eta$ set to 1 | $\eta$ relative to GM pollen | $\eta$ relative to GM pollen | $\eta$ relative to GM pollen |
| **Mechanism** | | | | |
| Reduced compatibility due to being different species | 1 | > 1 | > 1 | > 1 |
| Rejection of pollen that differ too much from own genotype | 1 | > 1 | > 1 | (>) 1 (possibly slightly larger than one, due to overlap in genotype as a result of regular gene flow with target population) |
| Self-sterile or containing barriers preventing self-fertilisation | 1 | 0 | 1 | 1 |
| Partly self-sterile | 1 | <1 | 1 | 1 |
| Heteromorphic incompatibility system | 1 | 0 (self-sterile) | 1 (compatible with half of the plants (distyly), but so is GM pollen) | 1 (compatible with half of the plants (distyly), but so is GM pollen) |
| Homomorphic incompatibility system (sporophytic as well as gametophytic) | 1 | 0 (self-sterile) | < 1 (dependent on number of different alleles in population; if variation is low, then << 1) | 1 (but when exchange with target population is high, then probably <1) |

Now, consider a plant with a certain amount $D_{gm}$ of pollen derived from GM plants and an amount $D_{res}$ derived from resident plants. If the compatibility of the GM pollen is set to one, the effective pollen number of the resident pollen is reduced or increased with a factor $\eta_{res}$. The proportion of GM pollen ($P_{gm}$) is then given by

$$P_{gm} = \frac{D_{gm}}{D_{gm} + \eta_{res} D_{res}}$$

The user of the model should estimate $\eta_{res}$, which is the relative fertilisation probability of the resident pollen compared to the GM pollen. This parameter can be split up in a factor estimating the relative fertilisation probability of self-pollen $\eta_{self}$ and the relative fertilisation of pollen originating from other plants in the target population $\eta_{in}$:

$$\eta_{res} D_{res} = \eta_{self} D_{self} + \eta_{in} D_{in}$$

in which $D_{self}$ is the amount of pollen that lands on the own stigma and $D_{in}$ is the amount of pollen originating from other plants in the population. $\eta_{self} D_{self}$ corresponds to the selfing rate of a species. $\eta_{in} D_{in}$ corresponds to successful outcross fertilisations within the target population. Of many species, the selfing rate (or the range in which the selfing rate is) is known, but the amount of pollen that lands on the own stigma or on another stigma within a population is mostly unknown. Therefore, $\eta_{in} D_{in}$ is not easy to estimate. $\eta_{in}$ covers many different processes that together lead to a certain fertilisation probability relative to the fertilisation probability of outcrossing pollen originating from the GM source population.

With these data, the model can estimate the fertilisation chances of pollen originating from a given GM source population in a given target population for a given distance.

### *5.2.4 Module 3: Contamination in a landscape*

The third module, which will, due to time constraints, not be programmed at present, concerns more complex situations. In this module, a landscape is simulated that contains one or more GM source populations surrounded by several target populations. The contamination levels of these target populations, resulting in seed containing modified DNA, will be estimated. Since pollen originating from every individual population disperses over the entire landscape, every target population has certain chance of being reached by pollen originating from the GM source populations and from the other target populations. This makes the situation much more complex. Pollen originating from every population now has certain probability of reaching a given target population, depending on the distance between the populations and on population characteristics like size and pollen production.

Three types of pollen can be present in a given target population: pollen originating from a GM source population, pollen originating from the target population itself and pollen originating from one of the other target populations. This makes that the relative fertilisation chances have to be estimated not only for the GM source and the target population, but as well for pollen of the other target populations

arriving at the focal target population (table 3). This estimation should be performed for all target populations.

To simulate a landscape in the model, the user should enter, in addition to the earlier-mentioned dispersal and fertilisation characteristics, the density and clustering of the populations in the landscape, as well as the average population sizes (fig. 4). The model then calculates the number of populations that will have a higher contamination level than the threshold specified by the user or it gives an overview of classes with different levels of contamination.

# CHAPTER 6. CONCLUSIONS

The COGEM uses environmental risk analysis (ERA) to evaluate proposals for the cultivation of GM plants. Estimating the probability of outcrossing is only one step in the ERA, the other being estimation of the consequences of such outcrossing. In this report, we were concerned solely with the first part: our aim was to evaluate the main processes that should be included in a model for outcrossing probabilities. For outcrossing to occur, a pollen grain originating from the GM source population must reach a given target population, fertilise a plant in that target population and then the resulting hybrid seed must establish. Together, these processes determine the probability of modified DNA introgressing in the DNA of the target population.

The process of pollen dispersal is highly complicated, not only differing between species, but also within a species, depending on characteristics such as insect abundance, weather and population characteristics. However, the main dispersal pattern seems to be ubiquitous. Most pollen lands close to the dispersing plant, but the small fraction that travels further may cover large distances. In the model we develop, we include several different equations for pollen dispersal, so that the most appropriate curve can be chosen for every species. However, every user of the model should keep in mind that even within a species the dispersal curve can differ considerably.

At present, pollen viability is poorly understood. There are some indications that pollen survival follows a negative cumulative normal distribution, but the available information is too preliminary to depend on. Nonetheless, we include loss of viability as a component of our model and describe it using just such a distribution, for two main reasons. First, the same distribution is known to fit well for the survival of seeds and spores. Second, the negative normal distribution is intuitively the obvious choice: most pollen will survive for a certain time period, but a few grains will be able to survive for much longer. In the model, we use loss of viability as a function of distance instead of time, since it is unknown if and how the two are correlated.

After reaching the target population, the next step is fertilisation. In the model, fertilisation chances are based not only on pollen numbers present on a stigma, but also on compatibility data between source and target species. Even when a great deal of information is available regarding the compatibility differences between pollen from the GM source population and pollen from the target population, it may still be difficult to give an exact estimate of this compatibility.

Introgression will not be considered in this first version of our model. Too much information is lacking to know how to simulate this process realistically. Our model can therefore be used to estimate probabilities of outcrossing up to the stage of hybrid seed formation, but no further.

The outcome of the model will largely depend on the parameter values entered by the user. The user should be aware of two types of uncertainty associated with this. One type of uncertainty is whether the parameter values used have been estimated correctly. The other uncertainty is caused by variation in parameter values due to stochastic processes, such as the effect of weather. For the evaluation procedure, it is important to give an exact estimation of the contamination level of a given target population with modified DNA. Including confidence intervals in the program would give an idea about possible deviation from the contamination levels found. One possible way of calculating confidence intervals to account for incorrectly estimated parameter values would be to take a number of random samples around the estimated value. This

feature is not included in the model at present, but we recommend that it is added to future versions.

The model will be helpful for estimating the separation distances required to reduce contamination levels with modified DNA to acceptably low levels. With additional time to develop the model further, we could consider several more complex situations. The most obvious step for further development concerns the estimation of gene flow at the landscape level, with multiple target populations surrounding one or more GM source populations. We recommend that in the future the COGEM aims to extend and refine the present model, to continually improve our estimates of the outcrossing probabilities of GM populations with cultivated or wild relatives.

## REFERENCES


Arias DM & Rieseberg LH 1994. Gene flow between cultivated and wild sunflowers. *Theoretical and Applied Genetics* 89: 655-660.

Aylor DE, Schultes NP & Shields EJ 2003. An aerobiological framework for assessing cross-pollination in maize. *Agricultural and Forest Meteorology* 119: 111-129.

Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE & Sork VL 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13: 937-954.

Bateman AJ 1947a. Contamination of seed crops. I. Insect pollination. *Journal of Genetics* 48: 257-275.

Bateman AJ 1947b. Contamination of seed crops. II. Wind pollination. *Heredity* 1: 235-246.

Bateman AJ 1947c. Contamination in seed crops. III. Relation with isolation distance. *Heredity* 1: 303-336.

Beekman M & Ratnieks FLW 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. *Functional Ecology* 14: 490-496.

Berge G, Nordal I & Hestmark G 1998. The effect of breeding systems and pollination vectors on the genetic variation of small plant populations within an agricultural landscape. *Oikos* 81: 17-29.

Bullock JM & Clarke RT 2000. Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia* 124: 506-521.

Burnham KP & Anderson DR 2002. *Model Selection and Multi-Model Inference. A Practical Information – Theoretic approach.* Springer-Verlag, Telos.

Chasan R 1992. Racing pollen tubes. *The Plant Cell* 4: 747-749.

Chittka L, Thomson JD & Waser NM 1999. Flower constancy, insect psychology and plant evolution. *Naturwissenschaften* 86: 361-377.

Clark JS, Macklin E & Wood L 1998. Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs* 68: 213-235.

Clark JS 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist* 152: 204-224.

COGEM 1999. Richtlijnen van de COGEM bij de beoordeling van veldproefaanvragen: criteria waaraan de beschrijving van de genetische modificatie moet voldoen, in relatie tot de maximale omvang van veldexperimenten. COGEM kenmerk CGM/990518-41.

Cresswell JE, Bassom AP, Bell SA, Collins SJ & Kelly TB 1995. Predicted pollen dispersal by honey-bees and three species of bumble-bees foraging on oilseed rape: a comparison of three models. *Functional Ecology* 9: 829-841.

Cresswell JE, Osborne JL & Bell SA 2002. A model of pollinator-mediated gene flow between plant populations with numerical solutions for bumblebees pollinating oilseed rape. *Oikos* 98: 375-384.

Cresswell JE 2003. Towards the theory of pollinator-mediated gene flow. *Philosophical Transactions of the Royal Society of London B* 358: 1005-1008.

Dafni A & Firmage D 2000. Pollen viability and longevity: practical, ecological and evolutionary implications. *Plant Systematics and Evolution* 222: 113-132.

De Nettancourt D 1977. *Incompatibility in Angiosperms.* Springer-Verlag, Berlin.

Dowding P 1987. Wind pollination mechanisms and aerobiology. *International Review of Cytology* 107: 421-437.

Dreisig H 1995. Ideal free distributions of nectar foraging bumblebees. *Oikos* 72: 161-172.

Ebert PR, Anderson MA, Bernatzky R, Altschuler M & Clarke AE 1989. Genetic polymorphism of self-incompatibility in flowering plants. *Cell* 56: 255-262.

Ellstrand NC, Devlin B & Marshall DL 1989. Gene flow by pollen into small populations: data from experimental and natural stands of wild radish. *Proceedings of the National Academy of Science of the United States of America* 86: 9044-9047.

Ellstrand NC, Prentice HC & Hancock JF 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30: 539-563.

Ellstrand NC 2003. *Dangerous Liaisons? When cultivated plants mate with their wild relatives.* Johns Hopkins University Press, Baltimore.

Fægri K & Van der Pijl L 1979. *The principles of pollination ecology.* Pergamon Press, Oxford.

Fernando DD & Cass DD 1997. Developmental assessment of sexual reproduction in *Butomus umbellatus* (Butomaceae): Male reproductive component. *Annals of Botany* 80: 449-456.

Free JB, Williams IH 1972. The transport of pollen on the body hairs of honeybees (*Apis mellifera* L.) and bumblebees (*Bombus* spp. L.). *Journal of Applied Ecology* 9: 609-615.

Golenberg EM 1987. Estimation of gene flow and genetic neighborhood size by indirect methods in a selfing annual, *Triticum dicoccoides*. *Evolution* 41: 1326-1334.

Goulson D, Ollerton J & Sluman C 1997. Foraging strategies in the small skipper butterfly, Thymelicus flavus: when to switch? *Animal Behaviour* 53: 1009-1016.

Goulson D & Wright NP 1998. Flower constancy in the hoverflies Episyrphus balteatus (Degeer) and Syrphus ribesii (L.) (Syrphidae). *Behavioral Ecology* 9: 213-219.

Gressel J & Rotteveel AW 2000. Genetic and ecological risks from biotechnologically-derived herbicide-resistant crops: decision trees for risk management. *Plant Breeding Review* 18: 251-303.

Groot MHM, Van de Wiel CCM, Van Tienderen PH & Den Nijs JCM 2004. *Hybridization and introgression between crops and wild relatives.* COGEM Report, Report Nr. 2003-02.

Hancock JF 2003. A framework for assessing the risk of trangenic crops. *BioScience* 53: 512-519.

Handel SN 1983. Pollination ecology, plant population structure and gene flow. In: Real, L., ed. *Pollination biology*, pp 163-211. Academic Press, Orlando, FL.

Haygood R, Ives AR & Andow DA 2003. Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society of London Series B – Biological Sciences.* 270: 1879-1886.

Herrera CM 1987. Components of pollinator "quality": comparative analysis of a diverse insect assemblage. *Oikos* 50: 79-90.

Hiscock SJ & McInnes SM 2003. The diversity of self-incompatibility systems in flowering plants. *Plant Biology* 5: 23-32.

Hokanson SC, Grumet R & Hancock JF 1997. Effect of border rows and trap/donor ratios on pollen-mediated gene movement. *Ecological Applications* 7: 1075-1081.

Holm E 1978. *Bloembiologie.* W. J. Thieme & Cie, Zutphen.
Hong TD, Ellis RH, Buitink J, Walters C, Hoekstra FA & Cranes J 1999. A model of the effect of temperature and moisture on pollen longevity in air-dry storage environments. *Annals of Botany* 83: 167-173.
Ingram J 2000. The separation distances required to ensure cross-pollination is below specified limits in non-seed crops of sugar beet, maize and oilseed rape. *Plant Varieties and Seeds* 13: 181-199.
James EA & Knox RB 1993. Reproductive biology of the Australian species of the genus *Pandorea* (Bignoniaceae). *Australian Journal of Botany* 41: 611-626.
Johnson JB & Omland, K. S. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19: 101-108
Klinger T, Arriola PE & Ellstrand NC 1992. Crop-weed hybridization in radish (Raphanus sativus): effects of distance and population size. *American Journal of Botany* 79: 1431-1435.
Knox RB 1979. *Pollen and allergy.* Edward Arnold, London.
Lavigne C, Godelle B, Reboud X & Gouyon PH 1996. A method to determine the mean pollen dispersal of individual plants growing within a large pollen source. *Theoretical and Applied Genetics* 93: 1319-1326.
Lavigne C, Klein EK, Vallée P, Pierre J, Godelle B & Renard M 1998. A pollen-dispersal experiment with transgenic oilseed rape. Estimation of the average pollen dispersal of an individual plant within the field. *Theoretical and Applied Genetics* 98: 886-896.
Lavigne C, Devaux C, Deville A, Garnier A, Klein EK, Lecomte J, Pivard S & Gouyon PH 2004. Petential and limits of modelling to predict the impact of transgenic crops in wild species. In: Den Nijs, H. C. M., Bartsch, D. & Sweet, J., eds. *Introgression from genetically modified plants into wild relatives.* CABI publishing, Oxfordshire.
Meeuse BJD 1961. *The story of pollination.* Ronald Press Company, New York.
Morris WF 1993. Predicting the consequences of plant spacing and biased movement for pollen dispersal by honey bees. *Ecology* 74: 493-500.
Morris WF, Kareiva PM & Raymer PL 1994. Do barren zones and pollen traps reduce gene escape from transgenic crops? *Ecological Applications* 41: 157-165.
Myung IJ, Forster MR & Browne MW 2000. Special issue on model selection. *Journal of mathematical psychology* 44: 1-2.
Nurminiemi M, Tufto J, Nilsson N-O & Rognli OA 1998. Spatial models of pollen dispersal in the forage grass meadow fescue. *Evolutionary Ecology* 12: 487-502.
Paterniani E & Stort AC 1974. Effective maize pollen dispersal in the field. *Euphytica* 23: 129-134.
Pitt MA & Myung IJ 2002. When a good fit can be bad. *Trends in cognitive sciences* 6: 421-425.
Raynor GS, Ogden EC & Hayes JV 1971 . Dispersion and deposition of pollens as a function of source and particle size. *Bulletin American Meteorological Society* 52: 309.
Raynor GS, Ogden EC & Hayes JV 1972. Dispersion and deposition of timothy pollen from experimental sources. *Agricultural Meteorology* 9: 347-366.
Reboud X 2003. Effect of a gap on gene flow between otherwise adjacent transgenic Brassica napus crops. *Theoretical and Applied Genetics* 106: 1048-1058.
Richards CM, Church S & McCauley DE 1999. The influence of population size and isolation on gene flow by pollen in *Silene alba*. *Evolution* 53: 63-73.

Rognli OA, Nilsson N-O & Nurminiemi M 2000. Effects of distance and pollen competition on gene flow in the wind-pollinated grass *Festuca pratensis* Huds. *Heredity* 85: 550-560.

Shivanna KR, Linskens GF & Cresti M 1991. Pollen viability and pollen vigor. *Theoretical and Applied Genetics* 81: 38-42.

Skogsmyr I & Lankinen Å 1999. Selection on pollen competitive ability in relation to stochastic factors influencing pollen deposition. *Evolutionary Ecology Research* 1: 971-985.

Snow AA & Spira TP 1991a. Differential pollen-tube growth rates and nonrandom fertilization in Hibiscus moscheutos (Malvaceae). *American Journal of Botany* 78: 1419-1426.

Snow AA & Spira TP 1991b. Pollen vigour and the potential for sexual selection in plants. *Nature* 352: 796-797.

Sork VL, Nason J, Campbell DR & Fernandez JF 1999. Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution* 14: 219-224.

Spira TP, Snow AA & Puterbaugh MN 1996. The timing and effectiveness of sequential pollinations in *Hibiscus moscheutos*. *Oecologia* 105: 230-235.

Stewart Jr CN, Halfhill MD, Warwick SI 2003. Transgene introgression from genetically modified crops to wheir wild relatives. *Nature Reviews Genetics* 4: 806-817.

Tackenberg O 2001. *Methoden zur Bewertung gradueller Unterschiede des Ausbreitungspotentials von Pflanzenarten. Modellierung des Windausbreitungspotentials und regelbasierte Ableitung des Fernausbreitungspotentials.* Dissertationes Botanicae 374, Cramer, Berlin. [Online, URL: <http://archiv.ub.uni-marburg.de/diss/z2001/0107>.]

Tackenberg O 2003. Modeling long-distance dispersal of plant diaspores by wind. *Ecological Monographs* 73: 173-189.

Tackenberg O, Poschlod P & Bonn S 2003. Assessment of wind dispersal potential in plant species. *Ecological Monographs* 73: 191-205.

Thomson J D 1988. Germination schedules of pollen grains: implications for pollen selection. *Evolution* 43: 220-223.

Treu R & Emberlin J 2000. Pollen dispersal of the crops Maize (*Zea mays*), Oilseed rape (*Brassica napus* spp. *oleifera*), Potatoes (*Solanum tuberosum*), Sugar beet (*Beta vulgaris* spp. *vulgaris*) and Wheat (*Triticum aestivum*). Soil Association, Bristol.

Tyldesley JB 1973. Long-range transmission of tree pollen to Shetland. *New Phytologist* 72: 175-181.

Van de Wiel C, Groot M, & Den Nijs H 2003. Gene flow from crops to wild plants and its population-ecological consequences in the context of GM-crop biosafety, including some recent experiences from lettuce. In: Wesseler JHH (ed.) in press. Environmental costs and benefits of transgenic crops. Springer.

Van Raamsdonk LWD & Schouten HJ 1997. Gene flow and establishment of transgenes in natural plant populations. *Acta Botanica Neerlandica* 46: 69-84.

Velterop O 2000. *Effects of fragmentation on pollen and gene flow in insect-pollinated species.* Dissertation University of Groningen. [Online, URL: <http://www.ub.rug.nl/eldoc/dis/science/o.velterop/>.]

Vogler DW & Kalisz S 2001. Sex among the flowers: the distribution of plant mating systems. *Evolution* 55: 202-204.

Wagner DB & Allard RW 1991. Pollen migration is predominantly self-fertilizing plants: Barley. *Journal of Heredity* 82: 302-304.

Walklate PJ, Hunt FCR, Higson HL & Sweet JB 2004. A model of pollen-mediated gene flow for oilseed rape. *Proceedings of the Royal Society of London B.* 271: 441-449.

Walsh NE & Charlesworth D 1992. Evolutionary interpretations of differences in pollen tube growth rates. *The Quarterly Review of Biology* 67: 19-37.

Waser NM 1986. Flower constancy: definition, cause, and measurement. *The American Naturalist* 127: 593-603.

Watrud LS, Lee EH, Fairbrother A, Burdick C, Reichman JR, Bollman M, Storm M, King G & Van de Water PK 2004. Evidence for landscape-level, pollen-mediated gene flow from genetically modified creeping bentgrass with *CP4 EPSPS* as a marker. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14533-14538.

Wilcock C & Neiland R 2002. Pollination failure in plants: why it happens and when it matters. *Trends in Plant Science* 7: 270-277.

Woolf CM 1968. *Principles of Biometry*. D. Van Nostrand Company, Inc., New York.

# APPENDIX A. DERIVATION OF POLLEN DISPERSAL CURVES

## A.1 EXPONENTIAL DISTRIBUTION

A negative exponential function is given by

$$f(D) = a \exp(-\lambda D)$$

in which $\lambda$ is a shape parameter and $a$ is a scale parameter.

For a given $\lambda$, $a$ can be determined by the requirement that the integral of a probability density function over the whole space of possible events has to be equal to 1. In two dimensions, this consistency requirement corresponds to

$$\int_0^{\infty} \int_0^{2\pi} f(D)\, d\varphi\, dD = 1,$$

where $D$ is the distance to a posit source of pollen and $\varphi$ is the angular direction.
For a given distance $D$,

$$\int_0^{2\pi} d\varphi = 2\pi D,$$

implying

$$1 = \int_0^{\infty} 2\pi D\, a \exp(-\lambda D)\, dD = \frac{2\pi a}{\lambda^2}.$$

As a consequence

$$a = \frac{\lambda^2}{2\pi}$$

For a continuous distribution function, the arithmetic mean $\overline{D}$ is given by

$$\overline{D} = \int D\, f(D)\, dD$$

The average dispersal distance $\overline{D}$ of the negative exponential function is inversely proportionate to $\lambda$ and given by

$$\overline{D} = \int_0^{\infty} \int_0^{2\pi} D\, f(D)\, d\varphi\, dD = \lambda^2 \int_0^{\infty} D^2 \exp(-\lambda D) = \frac{2}{\lambda}.$$

For a continuous distribution function, the variance var(D) is given by

$$\mathrm{var}(D) = \int (D - \overline{D}^2)\, f(D)\, dD = \int D^2\, f(D)\, dD - D^2$$

The variance of the negative exponential function is given by

$$\mathrm{var}(D) = \int_0^{\infty} \int_0^{2\pi} D^2\, f(D)\, d\varphi\, dD - \overline{D}^2 = \lambda^2 \int_0^{\infty} D^3 \exp\left(-\lambda D\right) dD - \overline{D}^2 = \frac{6}{\lambda^2} - \overline{D}^2 .$$

## A.2 INVERSE POWER LAW

An inverse power law is given by

$$f(D) = a\, D^{-\lambda}$$

where $\lambda$ is a shape parameter and $a$ is a scale parameter. A power function has the undesirable property that $f(D)$ tends to infinity for $D$ approaching zero. We therfore use the modified version

$$f(D) = a\, (D+1)^{-\lambda}$$

For a given $\lambda$, $a$ can be determined by the requirement that the integral of a probability density function over the whole space of possible events has to be equal to 1. In two dimensions, this consistency requirement corresponds to

$$\int_0^{\infty}\int_0^{2\pi} f(D)\, d\varphi\, dD = 1,$$

implying

$$1 = \int_0^{\infty} 2\pi D\, a\, (D+1)^{-\lambda}\, dD = \frac{2\pi a}{(\lambda-2)(\lambda-1)}.$$

As a consequence

$$a = \frac{(\lambda-2)(\lambda-1)}{2\pi} \qquad \text{for } \lambda > 2.$$

The average dispersal distance $\overline{D}$ is inversely proportionate to $\lambda$:

$$\overline{D} = \int_0^{\infty}\int_0^{2\pi} D\, f(D)\, d\varphi\, dD = (\lambda-2)(\lambda-1)\int_0^{\infty} D^2 (D+1)^{-\lambda} = \frac{2}{\lambda-3} \qquad \text{for } \lambda > 3.$$

The variance is given by

$$\int_0^{\infty}\int_0^{2\pi} D^2\, f(D)\, d\varphi\, dD - \overline{D}^2 = (1-\lambda)(2-\lambda)\int_0^{\infty} D^3\, (D+1)^{-\lambda} dD - \overline{D}^2$$

$$= \frac{6}{(\lambda-3)(\lambda-4)} - \overline{D}^2 \qquad \text{for } \lambda > 4.$$

## A.3 UNIFORM DISTRIBUTION

A uniform distribution is given by

$$f(D)=\frac{a}{D_{\max}},$$

in which $a$ is a scale parameter.

The parameter $a$ can be determined by the requirement that the integral of a probability density function over the whole space of possible events has to be equal to 1. In two dimensions, this consistency requirement corresponds to

$$\int_0^{D_{\max}} \int_0^{2\pi} f(D)\, d\varphi\, dD = 1$$

implying

$$1 = \int_0^{D_{\max}} \frac{2\pi a}{D_{\max}} D\, dD = \pi\, a\, D_{\max}.$$

As a consequence

$$a = \frac{1}{\pi\, D_{\max}}.$$

The average dispersal distance $\overline{D}$ is given by

$$\overline{D} = \int_0^{D_{\max}} \int_0^{2\pi} D\, f(D)\, d\varphi\, dD = \frac{2}{(D_{\max})^2} \int_0^{D_{\max}} D^2 = \frac{2}{3} D_{\max}$$

The variance is given by

$$\int_0^{D_{\max}} \int_0^{2\pi} D^2\, f(D)\, d\varphi\, dD - \overline{D}^2 = \frac{2}{(D_{\max})^2} \int_0^{D_{\max}} D^3\, dD - \overline{D}^2 = \frac{1}{2} D_{\max} - \overline{D}^2.$$