GENE FLOW FROM TRANSGENIC PLANT POPULATIONS: MODELS AND APPLICATIONS FOR RISK ASSESSMENT

Anna Kuparinen
Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

Academic dissertation
To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium B123, Exactum (Gustaf Hällströmin katu 2b), on January 4th, 2007, at 12 o'clock noon.

Helsinki 2006
ABSTRACT

The future use of genetically modified (GM) plants in food, feed and biomass production requires a careful consideration of possible risks related to the unintended spread of transgenes into new habitats. This may occur via introgression of the transgene to conventional genotypes, due to cross-pollination, and via the invasion of GM plants to new habitats. Assessment of possible environmental impacts of GM plants requires estimation of the level of gene flow from a GM population. Furthermore, management measures for reducing gene flow from GM populations are needed in order to prevent possible unwanted effects of transgenes on ecosystems.

This work develops modeling tools for estimating gene flow from GM plant populations in boreal environments and for investigating the mechanisms of the gene flow process. To describe spatial dimensions of the gene flow, dispersal models are developed for the local and regional scale spread of pollen grains and seeds, with special emphasis on wind dispersal. This study provides tools for describing cross-pollination between GM and conventional populations and for estimating the levels of transgenic contamination of the conventional crops. For perennial populations, a modeling framework describing the dynamics of plants and genotypes is developed, in order to estimate the gene flow process over a sequence of years.

The dispersal of airborne pollen and seeds cannot be easily controlled, and small amounts of these particles are likely to disperse over long distances. Wind dispersal processes are highly stochastic due to variation in atmospheric conditions, so that there may be considerable variation between individual dispersal patterns. This, in turn, is reflected to the large amount of variation in annual levels of cross-pollination between GM and conventional populations. Even though land-use practices have effects on the average levels of cross-pollination between GM and conventional fields, the level of transgenic contamination of a conventional crop remains highly stochastic.

The demographic effects of a transgene have impacts on the establishment of transgenic plants amongst conventional genotypes of the same species. If the transgene gives a plant a considerable fitness advantage in comparison to conventional genotypes, the spread of transgenes to conventional population can be strongly increased. In such cases, dominance of the transgene considerably increases gene flow from GM to conventional populations, due to the enhanced fitness of heterozygous hybrids. The fitness of GM plants in conventional populations can be reduced by linking the selectively favoured primary transgene to a disfavoured mitigation transgene. Recombination between these transgenes is a major risk related to this technique, especially because it tends to take place amongst the conventional genotypes and thus promotes the establishment of invasive transgenic plants in conventional populations.

KEYWORDS

Co-existence, Cross-pollination, Dispersal modelling, Gene flow, Pollen dispersal, Risk assessment of GM plants, Transgene spread, Transgenic crop, Transgenic forest

LIST OF ARTICLES

This thesis consists of a summary part and six research articles. The articles that are published or accepted for publication have been reproduced with the kind permission of the journals concerned.


II Kuparinen A., Markkanen T., Riikonen H., Vesala T. Modeling air-mediated dispersal of spores, pollen and seeds in forested areas. (submitted)


IV Kuparinen A., Schurr F., Tackenberg O., O'Hara R.B. Air-mediated pollen flow from genetically modified to conventional crops. Ecological Applications (in press)


VI Kuparinen A., Schurr F.M. Assessing the risk of gene flow from genetically modified trees carrying mitigation transgenes. (submitted)

AUTHOR'S CONTRIBUTION

I am fully responsible for the summary part of this thesis and the paper I. In paper II, I participated the model development, took major responsibility in planning and carrying out the simulation study and did major part of the writing. In paper III, I produced the pollen dispersal data, participated in developing code for model fitting procedures, carried out the model fitting and did major part of the writing. In papers IV-VI, I did major parts of the model development, had major responsibility in planning the simulation design and carrying out the simulations, and did most of the writing.
CONTENTS OF SUMMARY

1 INTRODUCTION 7
   1.1 Cultivation of genetically modified plants in the European Union 7
   1.2 Gene flow from genetically modified plant populations 7
   1.3 Models as tools for quantifying the gene flow process 8
   1.4 Aims of the research project 8

2 MODELS FOR POLLEN AND SEED DISPERSAL 9
   2.1 Mechanistic wind-dispersal models 9
      2.1.1 Modelling airflows in the atmospheric boundary layer 11
      2.1.2 Sensitivity of the wind-dispersal process 12
   2.2 Statistical modelling of empirically observed dispersal patterns 12

3 GENE FLOW FROM GENETICALLY MODIFIED CROPS 14
   3.1 Contamination of conventional crops 14
   3.2 Modelling gene flow from genetically modified to conventional wind-pollinated crops 15
   3.3 Stochasticity in pollen mediated gene flow 16
   3.4 Management measures for preventing air-mediated pollen dispersal 16

4 GENE FLOW FROM GENETICALLY MODIFIED FORESTS 17
   4.1 Simulating spatial population-genotype dynamics of perennial plants 18
   4.2 Sensitivity of gene flow in tree populations 20
   4.3 Transgene mitigation 21

5 CONCLUSIONS 22
   5.1 Models for gene flow from GM populations 22
   5.2 Features of the gene flow process to be carefully considered 23
   5.3 Measures for preventing transgene spread 23
   5.4 Challenges for future risk assessment and management 24

ACKNOWLEDGMENTS 25

REFERENCES 26
1 INTRODUCTION

1.1 Cultivation of genetically modified plants in the European Union

The rapidly increasing development and use of genetically modified (GM) crop plants and trees in food, feed and biomass production (e.g. Chapman and Burke 2006) has raised the prospect of their cultivation in the European Union (EU). In spite of the economical and environmental benefits expected to be gained from the use of GM plants, their large scale cultivation has also raised concerns about possible societal and environmental impacts (e.g. Schiemann et al. 2003, Snow et al. 2005). This has led to the EU setting regulations and guidelines for the use of GM plants within its area (directive 2001/18/EC). The underlying principles in the introduction of the GM plants to European agriculture and forestry are the co-existence of GM, conventional and organic productions (commission recommendation 2003/556/EC) and the control of environmental impacts of GM plants (commission decision 2002/623/EC and council decision 2002/811/EC).

Co-existence of GM, conventional and organic production is meant to ensure the ability to choose between GM, conventional and organic products for both consumers and farmers, and thus also several economical and societal aspects are related to it (commission recommendation 2003/556/EC). In addition to traditionally cultivated plants, GM plants may also have impacts on wild plant and animal populations, ecosystem functioning and soils organisms. Such environmental impacts may be direct or indirect, and immediate, delayed or cumulative, so that detecting the effects of GM plants may be difficult (e.g. Kjellson and Strandberg 2001, commission decision 2002/623/EC, Snow et al. 2005).

In order to detect and manage the possible impacts of GM plants on other forms of cultivation and on the surrounding environment, systematic, case-by-case based risk assessment procedures will be applied to each GM plant variety accepted to cultivation in EU (commission decision 2002/623/EC). Preliminary risk assessment will be complemented with follow-up studies after the first release of a GM plant variety (council decision 2002/811/EC). The aim of such procedures is to identify and prevent possible environmental hazards and set required measures to manage the risks of unwanted impacts of GM plants. Practical risk management procedures will be embedded to national legislation of each member state of EU, according to the general guidelines for the usage of GM plants set by EU (directive 2001/18/EC).

1.2 Gene flow from genetically modified plant populations

One of the major concerns related to the cultivation of GM plants is the possible spread of transgenes outside the GM plant populations (e.g. Snow et al. 2005). The main processes through which gene flow from GM populations may occur are transgene introgression into conventional plants due to hybridization and the invasion of GM plants to unintended areas (Eastham and Sweet 2002, Messequer 2003, Snow et al. 2005, Chapman and Burke 2006). Consequently, the key underlying mechanisms for gene flow are the dispersal of pollen and seeds from GM plant populations (Haygood et al. 2004, Hails and Morley 2005, Snow et al. 2005).

The risks related to gene flow from GM populations depend on the impacts of the transgenes on recipient organisms and the impacts of the GM plants on the ecosystem in the
habitat into which they may invade (e.g. Snow et al. 2005, Chapman and Burke 2006). On the other hand, the extent to which the impacts of transgenes may take place depends on the spatial abundance of the transgenes, and it is thus directly related to gene flow from GM populations. The level and spatial dimensions of gene flow define the areas in which the cultivation of GM plants may have direct effects, and, thus, on which possible management measures and follow-up studies should be focused. Consequently, for assessing and managing risks related to GM plants, it is important to quantify the level of gene flow from GM populations (e.g. Easthem and Sweet 2002, Messequer 2003). Furthermore, a precautionary measure against the possible impacts of transgenes on unintended ecosystems is to find means for preventing or reducing the gene flow process from GM populations (Messequer 2003). Some practical measures suggested to control gene flow are isolation distances, local and areal restrictions and barriers (e.g. Easthem and Sweet 2002, commission recommendation 2003/556/EC ).

1.3 Models as tools for quantifying the gene flow process

The level of gene flow from GM populations cannot easily be estimated empirically. Pollen and seeds can easily disperse over so long distances that their whole dispersal patterns cannot be directly observed (e.g. Katul et al. 2005). In addition, the release of GM plants to the environment is strictly regulated (directive 2001/18/EC), so that it would not be possible in practice to conduct field trials at a spatial scale over which information would be needed for decision making (James et al. 1998, Schiemann 2003).

Mathematical modelling of the gene flow process and its key mechanisms provides an alternative approach to quantify the spatial spread of transgenes. Models for gene flow are especially convenient for systematic risks assessment and when a gene flow process should be investigated over a large variety of environmental conditions and farming or forestry scenarios (Boch et al. 2002, Messequer 2003, van Frankenhuyzen and Beardmore 2004, Williams and Davis 2005, Hails and Morley 2005). Models can also be used to explore the gene flow process prior to setting up field trials, so that empirical work can be focused on the most relevant factors and spatial scales (James et al. 1998). Furthermore, models are useful for scaling up the results of empirical trials to the spatio-temporal scales over which the decisions are to be made (Schiemann 2003, van Frankenhuyzen and Beardmore 2004).

Gene flow models developed for systematic risk assessment and management have to be both flexible and realistic. The models must be general enough to be applied to several plant species and they should be able to operate over large spatial and temporal scales. In particular, a model must be able to describe the uncertainty in the gene flow process caused by the stochasticity in underlying demographic processes and environmental conditions.

1.4 Aims of the research project

The main aim of this study is to develop mathematical modeling tools for systematically estimating the level and spatial range of gene flow from GM plant populations in boreal environment. This include models for pollen and seed dispersal, hybridization and spread and establishment of invading plants (I-V). By describing and investigating the mechanisms affecting gene flow from GM populations, this study aims to identify the species-related, environmental and genetic factors to which the gene flow process is most sensitive (II,
IV-VI). Furthermore, the efficiencies of several proposed management measures for preventing the gene flow are investigated using the developed modeling techniques (IV-VI). Therefore, this study provides both tools for quantifying the level of gene flow from GM populations and knowledge about the factors and processes affecting it.

2 MODELS FOR POLLEN AND SEED DISPERSAL

The dispersal of pollen and seeds are key mechanisms leading to gene flow from GM populations. Quantifying the patterns of pollen and seeds dispersing from GM populations is essential for estimating and predicting over which distances and at which rates transgenes may spread in space. In particular, estimating the frequencies in which particles disperse over unusually long distances is important, since gene flow caused by such long-distance dispersal (LDD) events is difficult to control with local management measures.

Wind is the main vector for pollen for most boreal crop species (e.g. wheat, rye, corn) and trees (e.g. birch, spruce, pine, poplar), although insect pollination also occurs (e.g. oilseed-rape). Similarly, seeds of several tree species are wind dispersed (e.g. birch, pine, poplar), and crop seeds may also be spread by wind during harvesting (e.g. oilseed rape). For wind-dispersed pollen grains and seeds, LDD may easily occur due to vertical updrafts caused by atmospheric instability (e.g. Nathan et al. 2002, Tackenberg 2003). Therefore, modeling wind dispersal of pollen and seeds is crucial for assessing the level of gene flow from populations of several GM plant varieties.

2.1 Mechanistic wind-dispersal models

The basic processes affecting the dispersal of airborne particles, such as pollen grains and seeds, are relatively well-known. This has led to a rapid development of models that try to describe and predict the dispersal of particles by combining information on underlying mechanisms of the dispersal process, such as turbulent airflows and gravity (I). These so-called mechanistic dispersal models are able to predict particle dispersal over both short and long distances based on information on atmospheric and environmental conditions and particle type. In contrast to more traditional empirical models they do not require measurements of the pattern of dispersed particles for model parameterization. A general lack of empirical data on dispersal patterns, especially over long dispersal distances (Cain et al. 2000, Higgins et al. 2003, Levin et al. 2003, Katul et al. 2005), has thus made mechanistic models particularly popular.

Describing airborne particle dispersal requires sufficiently detailed descriptions of atmospheric conditions, autocorrelated turbulent airflows and their interactions with the vegetation (I). Describing these features with analytical models has turned out to be impossible in practice. Simulation models that are more realistic in their description of airflows and that are flexible in incorporating several atmospheric, environmental and biological features are therefore considered preferable. These simulation models typically describe the movement of an individual particle by temporary three-dimensional wind speeds and constant particle settling due to gravity. Particle trajectory can be described by the equations (e.g. Tackenberg 2003, Soons et al. 2004)
\[ x(t' + \Delta t) = x(t') + u(t') \Delta t \]  \hspace{1cm} (1.1) \\
\[ y(t' + \Delta t) = y(t') + v(t') \Delta t \]  \hspace{1cm} (1.2) \\
\[ z(t' + \Delta t) = z(t') + (w(t') - v_t) \Delta t \]  \hspace{1cm} (1.3)

where \( \{x, y, z\} \) are coordinates of a particle, \( \{u, v, w\} \) are temporary wind speeds in corresponding directions, and \( v_t \) is the particle's falling velocity. Temporary wind speeds can be described based on direct empirical measurements (Tackenberg 2003) or they can be modelled (e.g. Soons et al. 2004). Lagrangian stochastic (LS) trajectory models have turned out the most promising way to describe the impacts of temporary wind speeds on particle dispersal due to their realistic description of turbulent flow-fields (I). These meteorological trajectory models simulate temporary wind speeds along an air-parcel trajectory based on canopy features and micro-meteorological parameters describing current atmospheric and wind conditions (II).

**Figure 1** Schematic figure illustrating the structure of the atmospheric boundary layer (ABL). The ABL consists of the surface layer and the convective layer. The roughness sublayer of the surface layer contains the canopy and the areas just above it. Heights of the layers in the figure are approximate examples and are meant to give an idea of the relative sizes of each layer.
Recently several airborne particle dispersal models that are based on LS trajectory models have been developed and successfully validated (I). These LS particle dispersal models are typically designed for rather local scale particle transport, and thus do not include descriptions for turbulence conditions in upper parts of the atmospheric boundary layer (ABL) (Figure 1). However, those conditions may be relevant for the larger scale transport of particles. Within the canopy, turbulence conditions have also been found to be relevant in describing LDD of particles, due to their impact on particle release (Skarpaas et al. 2006) and the formation of turbulent updrafts that can lift particles well above the canopy (Nathan and Katul 2005). These reasons gave an incentive to develop a LS particle dispersal model that includes a description of turbulence conditions everywhere in the ABL and to use this model to further assess the sensitivity of the particle dispersal on several meteorological and biological features (II).

2.1.1 Modelling airflows in the atmospheric boundary layer

Turbulent wind speeds within and directly above the canopy depend on the density of the vegetation and the vertical distribution of the foliage (Massman 1997, Massman and Weil 1999). In upper parts of the ABL turbulence, conditions no longer directly interact with the vegetation, but turbulent structures are height and stability dependent (Massman 1997, Massman and Weil 1999). Within the ABL, in the surface layer (Figure 1) vertical, lateral and horizontal turbulent wind speeds can be described by Gaussian distributions with mean 0 (Thomson 1987), whereas in the convective layer (Figure 1) vertical turbulence is characterized by strong, temporary updrafts and weaker but more frequent downdrafts (Luhar and Britter 1989). In practice these vertical convective structures can be described by a mixture of two Gaussian distributions with non-zero means, so that the resulting distribution for vertical turbulence is positively skewed (Luhar and Britter 1989). The convective layer turbulence model suggested by Luhar and Britter (1989) is cross-wind integrated, and thus it can only be used for modelling atmospheric transport in the downwind direction. For modelling convective layer turbulence in all three dimensions, an alternative approach has been suggested by Arya (2001). In this model convective layer turbulence is described by similar Gaussian distributions as in the surface layer, but the allocation of energy in turbulent airflows between vertical, cross-wind and down-wind components is changed compared to the allocation in the surface layer.

The features and models for the ABL turbulence described above and models for the mean horizontal airflow within and above the canopy (Massman 1997, Massman and Weil 1999) jointly give a complete description of the airflows within the ABL. Using this flow-field description, airborne particle movement can be described throughout the ABL. Depending on the model used for the convective layer turbulence, particle dispersal can be recorded in the downwind direction or fully in three dimensions. The difference between these models is in practice only visible in unstable atmospheric conditions during which strong vertical updrafts mostly occur, whereas in near-neutral conditions the performances of the models in predicting particle dispersal pattern are similar (II).

Heavy particles (e.g. large seeds) very rarely get lifted up far above the canopy, so that the turbulence conditions in convective layer do not affect their dispersion. Those particles that get lifted up to the convective layer (e.g. pollen and spores) tend to disperse far, due to the
impact of heavy horizontal airflow in upper parts of the ABL. Therefore their near-field dispersal is not affected by the turbulence in the convective layer, whereas it strongly affects LDD. Consequently, the required level of complexity in convective layer turbulence description for predicting particle dispersion depends on the size of the dispersing particles, stability conditions and the spatial scale over which the particle dispersal is considered (II).

2.1.2 Sensitivity of the wind-dispersal process

When predicting the wind dispersal of particles, particle size strongly guides the complexity and accuracy required of the model. Because of gravity, dispersing particles fall from one airflow to another, which reduces the autocorrelation in the wind conditions faced by the particle along its trajectory (e.g. Wilson 2000). However, incorporating this feature into the turbulence description is only relevant when modelling the dispersal of heavy particles: for light particles the fall is so slow that its effects on airflow autocorrelation are negligible (II). Similarly, among light particles the differences in the falling velocities have little effect on the dispersal distances. This suggests that for light particles the estimation of the falling velocity can be done roughly, and the effects of particle clustering (Di-Giovanni 1995) are negligible (II).

Particle dispersal distances are strongly affected by the release conditions. Particles released higher in the vegetation tend to disperse further, due to both a longer falling time and higher horizontal wind speeds (II). If particle release is regulated by a threshold release velocity (e.g. Schippers and Jongejans 2005), then the typical negative correlations between down-wind and vertical turbulent gusts (Finnigan 2000) may considerably reduce the LDD of the particles (II). However, several uncertainties are still related to the effects of the vegetation, turbulence and stability conditions on particle abscission (Schippers and Jongejans 2005, Skarpaas et al. 2006). For example, within-canopy turbulence may also promote particle release by inducing a waving motion in the vegetation (Skarpaas et al. 2006). This suggests that when modeling airborne particle dispersion, particular effort should be invested in accounting for possible direct and indirect mechanisms affecting particle release, and their impacts on the particle dispersal distances should be investigated.

2.2 Statistical modelling of empirically observed dispersal patterns

In contrast to mechanistic dispersal models that are able to predict particle dispersal patterns based on measurements and models on mechanismisms causing the dispersal, dispersal modelling can also be based on observed particle dispersal distances. If the only information available on the dispersal process consists of empirical measurements of a dispersal pattern, the dispersal process can be modelled using empirical dispersal models. These models are especially useful in situations where underlying mechanisms affecting the particle dispersal process are not known or there may not be any empirical measurements on the mechanisms available. Empirical dispersal models describe the expected density of the dispersal pattern with an analytical dispersal kernel, the parameters of which are estimated by fitting the kernel to observed data using statistical procedures. Selection between alternative models is done by comparing the fits of the models to the observed data.

The major challenge of empirical modelling is that one model should also be able to describe the process in situations from which there are no direct observations available. In the
context of empirical dispersal modelling this means that a model should be able to predict the
density of the dispersal pattern over longer distances than at which the dispersal pattern was
directly measured (III). Moreover, a dispersal model should be able to predict other outcomes
of the dispersal process (i.e. other dispersal patterns of similar particles). In other words, the
validity of an empirical model in describing the dispersal process depends on the possibilities
to extrapolate it both in time and in space.

Wind dispersal of airborne particles is strongly affected by varying atmospheric
conditions, so that the patterns of dispersing particles obtained under varying wind conditions
can greatly differ (Loos et al. 2003, Tackenberg 2003, III, IV). This natural variation in the
dispersal process makes the description of particle dispersal using empirical models difficult.
Different dispersal patterns can support the validity of different dispersal models, so that it
may not be possible to find one model that could be considered as a good description of the
whole underlying dispersal process (Fitt et al. 1987). In addition, the dispersal models that
have been commonly suggested are not flexible enough to be able to account for variation in
the dispersal process as variation in model parameters (III). In practice this means that the
confidence intervals of model parameters estimated from one dispersal pattern could be very
different from those estimated from another pattern.

The range over which the dispersal pattern has been observed can have a large
influence on the extrapolation of modelling results (III). Selection between alternative models
may depend on the distances over which the dispersal data has been collected; data collected
only over a small range may support a different model than data collected over longer range.
The choice of the model in turn affects the predicted particle dispersal over the observation
range and other predicted patterns. Moreover, the predicted amount of LDD in one dispersal
pattern may considerably vary because of the differences between the model parameters
estimated using dispersal data collected over shorter or longer spatial ranges.

The stochastic nature of the dispersal process (III, IV) and practical difficulties in
collecting dispersal data over long distances from the particle source (e.g. Cain et al. 2000,
Levin et al. 2003) can create uncertainty in the selection and estimation of dispersal models
(III). These uncertainties in turn reduce the predictive power of empirical dispersal models. In
practice this means that empirical experiments on pollen and seed dispersal may not provide
much information about the underlying dispersal processes, because the results of those
experiments can hardly be generalized. Therefore, assessment of the risks related to pollen
and gene dispersal should not be solely based on few field experiments, or, at the very least,
considerable effort should be invested in accounting for the uncertainty in the predictions that
may arise due to uncertainty in model selection and model parameterization.
3 GENE FLOW FROM GENETICALLY MODIFIED CROPS

3.1 Contamination of conventional crops

Farming of GM crops has lead to concerns about their possible impacts on conventional fields located in the areas surrounding transgenic fields. In particular, pollen dispersal from a GM crop to a conventional crop of the same species may lead to cross-pollination between GM and conventional plants (Eastham and Sweet 2002). As a result of this cross-pollination, a crop that is considered GM-free may still contain transgenic seeds (Figure 2). The risk of this spread of transgenes to conventional crops is highly species specific. For mostly self-pollinating species (e.g. wheat) pollen-mediated gene flow only occurs at very low rates, whereas for cross-pollinating crops, such as maize, sugar-beet and rye, the risk of cross-pollination between GM and conventional varieties is considerable (Eastham and Sweet 2002, commission recommendation 2003/556/EC).

The adventitious presence of transgenes in conventional crops is regulated by the EU, with a threshold contamination of 0.9% having been set for conventional crops (regulation 1829/2003). If the level of contamination of a crop exceeds this threshold, the crop has to be labelled as containing transgenes. Similar, but more strict, thresholds will also be set for seeds in order to ensure seed purity and avoid unintended admixture of GM and conventional varieties (commission recommendation 2003/556/EC).

Gene flow from GM to conventional crops is problematic in many ways. According to the principles of co-existence of GM and conventional farming, consumers and farmers should be able to choose between GM and non-GM production (commission recommendation 2003/556/EC). If the level of transgene contamination of conventional crops exceeds the threshold, the crop can no-longer be considered GM-free. This may cause considerable financial losses for conventional farmers due to lower market prices and possible management and monitoring regulations that should be applied to the contaminated field in years following the exceeding of the threshold. Additionally, frequent exceeding of the contamination thresholds may also reduce the general supply of non-GM products, which can affect consumer choices (commission recommendation 2003/556/EC). Transgene escape from GM to conventional fields may also promote the spread of transgenes outside cultivated areas and make crop management and monitoring generally more difficult, because they may also have to be applied in areas where GM crops have not been sown.

In order to ensure equal possibilities for both GM and non-GM farming, members of the EU are supposed to develop national co-existence strategies. Such strategies will include species specific local and regional regulations and restrictions regarding farming practices and land-use (commission recommendation 2003/556/EC).
**Figure 2** Transgenes may escape from GM to conventional fields via pollen dispersal and cross-pollination. As a result, the conventional crop will include transgenic seeds, the proportion of which is the level of contamination of the conventional crop.

### 3.2 Modelling gene flow from genetically modified to conventional wind-pollinated crops

Pollen dispersal and the resulting gene flow between adjacent, wind-pollinated crops has been studied in several field trials (e.g. Bateman 1947, Raynor et al. 1972, Klein et al. 2003). However, the results of those are highly dependent on local wind conditions and the trial set-up, so it is difficult to generalize them to other situations (Loos et al. 2003). Models that can be applied for different systems of fields and that are able to describe the variation in wind conditions within and between pollination periods are therefore a more flexible way to investigate the cross-pollination between fields. Such models can be used for obtaining a general understanding of the crop-to-crop gene flow process, and for predicting case-specific levels of transgene contamination in conventional fields. In particular, models can be used for rapidly exploring the effects of possible management measures on cross-pollination, in order to find cost-efficient ways to reduce the gene flow.
Quantifying gene flow from GM to conventional crops requires a model for the dispersal of airborne pollen (I-III) and a model describing the fertilization of the plants in the conventional crop (IV) (Figure 2). Accounting for the stochasticity in cross-pollination requires consideration of a range of possible wind conditions and pollen dispersal patterns. Therefore the pollen dispersal model must be able to account for the natural variation in wind conditions (III, IV). Furthermore, pollen dispersal has to be realistically predicted over both short and long distances. For these reasons, mechanistic simulation models for pollen dispersal are preferable (I-IV).

The final level of cross-pollination in conventional fields depends on local pollen competition determined by the amounts of GM and conventional pollen available (Walklate et al. 2004, Loos et al. 2003, IV). In addition to the relative amounts of different pollen types, the final level of outcrossing also depends on the fitness of GM and conventional genotypes (Walklate et al. 2004), the viability of pollen grains (e.g. Aylor 2004) and flowering dynamics of fields (commission recommendation 2003/556/EC). All these factors affect the success of a transgenic pollen grain in fertilizing an ovule of the conventional genotype (IV).

3.3 Stochasticity in pollen mediated gene flow

On average, the level of gene flow between adjacent fields is relatively low. However, because of the large variation in environmental conditions affecting pollen dispersal, cross-pollination can occasionally occur at much higher rates. Therefore, even though the typical levels of contamination of conventional crops are low, the contamination threshold can be exceeded if pollination occurs under favourable wind conditions (IV).

The amount of temporary updrafts in airflows turns out to be the most important mechanism causing large amounts of pollen dispersal from a GM to a conventional field (IV). Such updrafts are caused by instability in atmospheric conditions (Tackenberg 2003). Large amounts of pollen may thus occasionally disperse over long distances, for example due to strong updrafts before storm fronts (Emberlin et al. 1999).

This stochasticity in the pollen dispersal and in the resulting gene flow suggests that it is almost impossible to completely prevent the contamination threshold being exceeded, and field-specific levels of transgenic contamination are largely unpredictable. Routinely monitoring of conventional crops may thus be needed in order to detect occasional high levels of transgenic contamination.

On the other hand, investigating the levels of pollen mediated gene flow under a large variety of possible wind conditions estimates the effects of the natural variation in wind conditions on the variation in the levels of contamination (IV). Rather than directly predicting the level of contamination of conventional crops, their probability of exceeding the contamination threshold can be estimated. This probability, in turn, can be used in estimating the expected financial losses over a large number of fields.

3.4 Management measures for preventing air-mediated pollen dispersal

The traditional way to reduce pollen mediated gene flow from a plant population is to restrict the populations of compatible relatives in the surroundings (Chapman and Burke 2006). Increasing the distance between GM and conventional fields is an effective way to reduce wind-dispersal of pollen from GM to conventional fields (commission
recommendation 2003/556/EC, II-IV). Similarly, the relative supply of GM and conventional pollen can be affected by the allocation of the farming area between GM and conventional crops (commission recommendation 2003/556/EC). Both of these measures affect the amounts of different pollen types available in the conventional field, and, consequently, have an impact on the average level of cross-pollination (IV).

Considerable stochasticity of the air-mediated pollen dispersal process (III, IV) limits the applicability of isolation distance and relative field sizes and shapes of the fields in preventing gene flow. These management measures affect the average level of cross-pollination, which means that their effects are visible over a number of fields or years. Consequently, the effects of the isolation distance of fields and their relative sizes can be seen in the estimated probabilities of exceeding the contamination threshold (IV). When considering a system with a large number of fields these management measures have clear impacts on the level gene flow, but for an individual field the level of cross-pollination may still remain stochastic (IV).

Long-distance pollen dispersal cannot easily be affected by on-field management measures. Those airborne particles that disperse far from the source are typically lifted well above the canopy by vertical turbulent updrafts (Nathan and Katul 2005, II). Therefore pollen traps and barriers placed along field edges (commission recommendation 2003/556/EC) will only influence the movement of pollen grains that fly low and that are thus likely to be deposited at shorter distances, whereas the pollen grains that are lifted up can easily cross such elements (IV). This further suggests that local turbulence conditions in GM fields are in a key position for the determination of the potential levels of gene flow to neighbouring fields.

4 GENE FLOW FROM GENETICALLY MODIFIED FORESTS

Over the past few years, a number of transgenic tree varieties have been developed and grown in trial plantations (van Frankenhuyzen and Beardmore 2004). GM trees are not yet established in commercial forestry, but in future they are expected to be widely used, for example in wood production and environmental management (e.g. Gartland et al. 2003, van Frankenhuyzen and Beardmore 2004, Walter 2004).

Despite the anticipated economic and environmental benefits, there are concerns related to the possible gene flow from GM to conventional forests. Due to the long life-cycles of trees, many tree species reproduce for a long time before being harvested. Compatible conventional relatives are also expected to be found in close proximity to GM tree populations, so that hybridization between GM and conventional trees may occur easily (James et al. 1998, van Frankenhuyzen and Beardmore 2004, Williams et al. 2005).

Crop plants have been under selective breeding for centuries and, thus, they have lost many traits that would be needed for survival outside cultivated areas. In contrast, tree genotypes selected for genetic modification are expected to come from native stands, so that GM trees may spread easily to wild habitats where they could replace the native trees (James et al. 1998, Walter 2004). The invasiveness of GM trees depends on their fitness in wild habitats and populations, which may be increased due to the demographic effects of the transgene (James et al. 1998, Walter 2004, Williams and Davis 2005).
Due to the longevity of trees, it is not possible to use field trials to study the level of gene flow from GM tree population, and its long-term impacts on ecosystems, in a time-frame relevant for commercial decision making (James et al. 1998, Williams and Davis 2005). In practice, large scale trials may only be possible by incorporating them to the early stage of commercialization of GM trees (James et al. 1998). Therefore, models for population dynamics are needed to estimate and investigate the gene flow process from GM tree populations (James et al. 1998, van Frankenhuyzen and Beardmore 2004, Williams and Davis 2005). These models should be able to incorporate spatial structure in genetic variation and its impacts on local population-genetic processes (Epperson 1995) and the effects of LDD dispersal events on gene flow (Nicholson and Hewitt 1994). Spatially explicit simulation models are most flexible in describing these features (e.g. Durrett and Levin 1994a,b).

4.1 Simulating spatial population-genotype dynamics of perennial plants

Modelling gene flow in perennial plant populations requires descriptions of the key demographic processes affecting the dynamics of the population (Figure 3). For long-living plants, such as trees, the demographic processes may strongly depend on the size of individuals, for example so that the reproduction starts only when a plant has reached a certain size. Describing the differences between the plants of different sizes can be done by structuring the population into size classes that differ in their demographic characteristics (e.g. Hara 1988, Bohrer et al. 2005). Demographic processes can then be modelled separately within each size class (V).

The dynamics of a population may affect its genetic structure, which, in turn, can feedback to the population dynamics due to possible demographic differences between the alleles (e.g. Davis and Shaw 2001, V). This means that rates determining demographic processes, such as growth or mortality, may depend on the current genetic structure of the population. To model this, demographic rates of a population has to be modelled as being genotype specific. In practice demographic rates for the conventional genotype can be considered as reference levels that are then increased or decreased according to the effect of the transgenic trait on the demographic process the rate is describing (V).

At a large spatial scale, individual plants will not affect all others, but rather plant interactions are restricted to more local scales. In forests, such interactions are typically one-sided, so that smaller trees compete for resources with larger individuals, but larger trees are not affected by smaller ones (Kohyama 1992). Competition between neighbouring plants can be described by dividing demographic processes of the whole population into location-specific subprocesses. To account for the one-sided competition, location-specific demographic rates can be modelled further as depending on the density of larger individuals and the carrying-capacity of the habitat in the neighbourhood considered (V).

Stochasticity has a large influence on both the dynamics of populations and genotypes. Local population dynamics are affected by demographic variability and rare LDD events (Clark et al. 2001, Higgins and Cain 2002). Demographic stochasticity also affects the fixation or extinction of rare alleles, through genetic drift (e.g. Gale 1980). Similarly, spatial and temporal variation in environmental conditions has an impact on demographic processes and is thus required for quantifying the natural variation in gene flow processes. In simulations for population dynamics, environmental variation in demographic processes can be described by
introducing variation between years and locations in the size and genotype specific demographic rates (V). Stochasticity in the outcome of demographic processes can be accounted for by considering the outcome of the process as a random variable with expected value indicated by the demographic rate and the number of individuals to which the rate should be applied (V).

A simulation model, AMELIE, that incorporates the features described above provides a description of population-genotype dynamics of perennial plants and the impacts of environmental conditions on those (V). The model is spatially explicit, and it characterizes individual plants according to their size, genotype and location in which they are growing. It describes hybridization and the spread of transgenic plants in a discretized study domain over a sequence of years. It is therefore able to describe the spread of transgenes from a GM to a conventional population both in time and in space.

**Figure 3** Demographic processes affecting the dynamics of a perennial plant population. Each of these processes has to be described when modelling gene flow in forests over a sequence of years.
4.2 Sensitivity of gene flow in tree populations

The level of gene flow from a GM to a conventional tree population is initially limited by the amount of cross-pollination between populations and the number of seeds carrying the transgene that migrate to the conventional population. However, the final level of transgene spread is determined by the establishment of transgenic plants among the conventional trees (e.g. James et al. 1998). This depends on the relative fitness of GM plants in the conventional population, the amount of local competition, species-specific characteristics and the prevailing environmental conditions (V).

To be invasive, a transgenic genotype must exhibit increased fitness compared to conventional genotypes. This is not necessarily always the case: for example, altered lignin concentration is a trait that is typically selectively disfavoured in conventional populations (James et al. 1998). On the other hand, traits that improve the viability of a plant may strongly support the establishment of GM genotypes. Therefore, when quantifying the risks of rapid gene flow from GM forest, the main interest is focused on traits that may have advantageous effects on the plants' fitness (e.g. Williams and Davis 2005).

The fitness of GM plants in a conventional population may also depend on the availability of resources in the conventional population. Such interactions between the genetically modified traits and local populations affect the final level of gene flow. For example, in a dense population the supply of pollen and seeds is typically not a factor that would limit population growth. Consequently, changes in reproductive output due to genetic modification may only have rather small effects on the level of gene flow (V). On the other hand, increased growth speed and decreased density-dependent mortality are traits that may considerably enhance the fitness of GM plants, especially in the presence of intensive, one-sided competition. Therefore these traits can increase the rate of establishment of GM plants and thus increase the final level of transgene spread (V).

Interactions between demographic and genetic characteristics of GM plants have clear effects on the level of gene flow. When comparing the spread of dominant and recessive transgenes from a homozygous GM plant population, the level of gene flow depends on both the expression of the transgene and the impact of the transgenic trait on a plant's fitness. If the genetically modified trait increases the fitness of GM plants in the conventional population, dominance of the transgene strongly increases the level of gene flow (V). On the other hand, if the transgene does not have a large impact on fitness, the difference between the levels of gene flow from GM populations with dominant or recessive transgenes is small. This can be explained through the fitness of heterozygous hybrids in the conventional population, which affects the local competition between them and conventional plants. If the heterozygous hybrid genotype expresses the transgene (i.e. if the transgene is dominant), and the transgene gives the hybrid genotype an advantage in competition with the conventional genotype (i.e. increases hybrid's fitness), the hybrid plants tend to replace the conventional plants.

GM plants can initially be either homozygous or heterozygous (V). If the transgene is dominant, the level of gene flow is markedly lower from initially heterozygous GM plant populations than from homozygous populations. This is because a heterozygous GM plant produces only half of the amount of transgenic gametes produced by a homozygous GM plant. However, the comparison between heterozygous GM plants with a dominant transgene and homozygous GM plants with a recessive transgene is not as clear. Even though the level of
gene flow from a heterozygous GM population is lower, a strong positive effect of transgene dominance may in time support the establishment of transgenic plants so much that the final level of gene flow is eventually higher from the heterozygous populations (V). This suggests that even though the gene flow process can be regulated by the initial genetic composition of GM plants and transgene recessiveness, these features should always be considered together with the possible fitness effects of the transgene.

4.3 Transgene mitigation

The fitness of GM genotypes in conventional populations can be reduced by genetic modification. A fitter transgene can be linked to another transgene that gives a large fitness disadvantage in conventional populations (Chapman and Burke 2006). This technique is called transgene mitigation. The transgene causing the originally intended trait is called as primary transgene, while the one introduced in order to reduce the plant's fitness is the mitigation transgene. The major risk related to transgene mitigation is the possible break-up of primary and mitigation transgenes (i.e. through recombination). This may give rise to hybrid genotypes that only express the primary transgene, and that are thus fitter than both the original GM genotype and the conventional genotype (Chapman and Burke 2006, VI).

The formation of hybrids expressing only the primary transgenic trait depends on the recombination rate between the genetically modified loci (Ewens 2004). If the modified loci are on different chromosomes this rate equals 0.5, whereas it approaches 0 when the distance between loci on the same chromosome diminishes (Ewens 2004).

In future, transgene mitigation may be a particularly relevant management measure for gene flow from GM tree populations. It has been suggested that gene flow from GM forests could be reduced by reducing the reproductive output of GM trees with a mitigation transgene (James et al. 1998, Walter 2004). Genetic modification techniques, e.g. for preventing flower formation of trees (Lemmetyinen et al. 2001), are already being developed. However, if the cross-pollination between genotypes cannot be completely prevented, break-up of the linkage of transgenes will still occur and lead to the spread of invasive hybrids.

For homozygous GM plants with two genetically modified loci, the formation of break-up hybrids expressing only one of the transgenes happens in two stages. Cross-pollination between plants of the conventional and transgenic genotypes leads to the formation of heterozygous plants that still express both transgenic traits. In the formation of gametes of these hybrids, recombination may lead to the formation of pollen grains and ovules that only contain one of the transgenes. Combinations of these gametes or their hybridization with gametes produced by plants of the conventional genotype may thus form 'break-up hybrids' that only express the primary or the mitigation transgene.

The spatial dependence of these two fertilization processes shifts the distribution of the break-up hybrids into the conventional population (VI). Consequently, their formation is a process that tends to occur at highest rates in the conventional populations. For practical risk management of GM plants, the consequences may be dramatic: if the break-up of primary and mitigation transgenes occur, it is likely to take place in populations where the GM plants are not intended to be grown.

The extent and the spatio-temporal scales over which break-up hybrids spread depend on the amount of pollen exchange between GM and conventional populations and the growth
speed of the plants. The initial proportions of the break-up hybrids depend on the linkage between the transgenes, but their establishment is further affected by their relative fitnesses in the conventional populations. Consequently, the spread of the hybrids expressing only the primary transgene is affected by both the genetic features, local dynamics and the dispersal potential of the plants (VI).

The break-up of the primary and mitigation transgenes causes a considerable risk for the formation and spread of invasive trangenic genotypes into the surrounding conventional populations. On the other hand, the formation of such genotypes can be prevented by increasing the linkage between the transgenes. Therefore, transgene mitigation can be considered as a promising tool for reducing the fitness of GM plants to conventional populations, but the risks related to this technique depend on how the tight linkage between the transgenes can be ensured (Chapman and Burke 2006, VI). Techniques increasing the linkage, such as creating a 'tandem construct' in which the primary transgene is placed between two copies of the mitigation transgene, have recently been developed (e.g. Al-Ahmad 2006).

5 CONCLUSIONS

5.1 Models for gene flow from GM populations

In the this thesis, I have presented mathematical models that can be used for estimating and predicting the level of gene flow from GM populations. Identifying areas where a GM population may have an impact is directly related to the estimation of the spatial patterns of dispersal of pollen and seeds from the GM population. Methods for modelling such dispersal patterns based on the underlying mechanisms of the dispersal process or on empirically observed patterns are presented in papers I-IV. Based on the pollen dispersal pattern, the level of cross-pollination between GM and conventional genotypes can be estimated and the modeling approach for this is presented in papers IV and V. Seeds resulting from such hybridizations, or those directly dispersing from a GM to a conventional populations may establish themselves as transgenic plants within a population of conventional genotypes. This process is limited by the initial amount of transgenic seed available but it also depends on local demographic processes and interactions between genotypes. To describe this, a modeling framework with submodels for individual demographic processes is presented in paper V.

The developed models can be used for estimating gene flow in both annual and perennial plant populations. Cross-pollination between annual crops is simulated in paper IV, and the resulting level of transgene contamination of conventional crop is estimated. In papers V and VI, gene flow from a GM to a conventional tree population is modelled over a sequence of years. The flexibility of the gene flow models in describing the effects of environmental, landscape and species specific features has been demonstrated by simulating gene flow under several environments, and demographic and genetic scenarios (IV-VI).
5.2 Features of the gene flow process to be carefully considered

Stochasticity plays a considerable role in the dispersal of pollen and seeds. For wind-dispersed particles, the dispersal distances can highly vary both within and between dispersal patterns due to the variation in wind conditions, particle release conditions and varying particle sizes (II-IV). Consequently, the dispersal patterns of pollen and seeds have to be modelled on a case-by-case basis. This requires measurements of species-specific canopy characteristics and features related to the particle and its release mechanisms (II). In addition, variation in wind conditions during typical particle dispersal periods as well as possible impacts of varying regions and habitats on wind conditions should be taken into account (II-IV). Understanding the role of these features in particle dispersion is crucial for capturing natural variation in the dispersal process and for understanding the effects of varying environments on particle dispersal patterns.

Variation in pollen dispersal patterns is reflected in variation in the levels of cross-pollination between GM and conventional populations (IV). Due to favourable wind conditions, occasionally large amounts of pollen may disperse over long distances and cause high levels of transgene contamination of conventional crops. Therefore, when assessing the level of gene flow from a GM crop, the variation in the level of gene flow should be estimated in addition to the average levels of gene flow. Modelling stochasticity in gene flow from GM populations is particularly important for exploring the frequency of rare extreme cases when the level of gene flow is unusually high (IV). Such events are highly relevant for risk assessment, since they are unexpected and typically not easily controllable, and may thus considerably increase the escape of transgenes to unintended areas.

The spread of transgenic plants to conventional populations depends on the traits introduced by the genetic modification and on the fitness advantage those may give to the genotypes carrying the transgene (V,VI). Careful estimation of the relative fitness of transgenic plants in conventional populations is thus needed, in order to assess the risk of transgene spread. This requires information about basic demographic parameters of GM and conventional plants. In addition, consideration of which demographic traits may be key to the regulation of plant establishment is needed for estimating the impacts of the changes in these demographic traits on the fitness of transgenic plants.

The effect of transgene dominance on gene flow depends on the size of the fitness advantage the transgene gives to a plant. In cross-pollinating populations, the more a transgene enhances the fitness of a plant the more dominance of a transgene promotes transgene spread due to the fitness advantage the transgene gives also to heterozygous hybrids (V). Therefore, genetic features of a GM plant variety should always be considered together with the demographic effects of the transgene. In particular, if the transgene gives a plant high fitness advantage, risk related to the dominance of the transgene should be acknowledged.

5.3 Measures for preventing transgene spread

Pollen and seed dispersal over long distances cannot easily be prevented by local management measures. Particles dispersing far from their release place are typically lifted well above the vegetation (II, IV), so that they cannot easily be stopped by physical barriers located around the particle source (IV). Therefore, small amounts of pollen and seeds are always likely to disperse over long distances.
The annual average levels of cross-pollination between GM and conventional populations can be affected by land-use practices, but because of the variation in pollen dispersal, the level of transgene contamination of a conventional population remains stochastic (IV). Due to this stochasticity, it is virtually impossible to prevent levels of contamination from ever exceeding the threshold of 0.9% for conventional crops. In practice this means that routine monitoring of conventional crops may be needed, in order to detect the cases when cross-pollination between GM and conventional crops has occurred at high levels.

For perennial plants, the establishment of transgenic plants in a conventional population can be reduced by the recessiveness of the transgene, or, in the case of a dominant transgene having the initial GM genotype be heterozygous. These management measures have large impacts on transgene spread (V), especially if the transgene considerably enhances the fitness of transgenic plants in conventional populations. This suggests that management of gene flow should be not only focused on agricultural and forestry practices, but the impacts of the genetic characteristics of a plant on gene flow should be considered already in the early stages of the development of GM plant varieties.

Reducing the fitness of invasive transgenic plants in a conventional population by linking it to a selectively disfavoured mitigation transgene may lead to the break-up of primary and mitigative transgenic traits. Invasive hybrids that express only the primary transgenic trait are likely to get established amongst conventional genotypes, which implies that spread of transgenes to conventional populations surrounding such a GM population may be a significant risk (VI). Therefore, the effectiveness of transgene mitigation as a management measure for preventing gene flow strongly depends on the possibilities of tightly link the transgenes together, so that recombination is very unlikely to occur.

5.4 Challenges for future risk assessment and management

The level of gene flow from GM populations strongly depends on the interactions between environmental, demographic and genetic characteristics of the plant and its populations. In the assessment and management of risks posed by GM plants, it is therefore important to jointly consider the impacts of these three components. In practice, this means that in many cases risks have to be investigated on a case-by-case basis, which may further lead to setting management measures and regulations that are specific to the species and habitat. Such risk assessment and management procedures require frameworks designed for rapid but effective decision making.

As in numerous ecological processes, the outcome of the process of gene flow may be highly stochastic. This is particularly true for annual species, for which averaging over years does not occur. The uncertainty in the gene flow process arising from environmental and demographic stochasticity has to be carefully accounted for when planning the cultivation of GM plants. Despite of any management measures, stochasticity in gene flow may occasionally lead to the spread of transgenes to unexpected areas. This has to be accounted for, both in the development of strategies for the co-existence of GM and conventional plants and in developing monitoring schemes.
ACKNOWLEDGMENTS

This study was conducted as a part of the ARGUE (Assessment and Regulation of the Ecological Effects of GMOs in Boreal Environment) research project that belongs to ESGEMO (Environmental, Societal and Health Effects of Genetically Modified Organisms) research programme. My work was funded by the Academy of Finland (project number 207423). During these two and a half years I have been working with my thesis, I visited the Department of Ecological Modelling of the UFZ research center in Leipzig for three months and the Chair of Plant Ecology and Nature Conservation of the University of Potsdam for six months. My visit in Leipzig was funded by a grant from UFZ, and thanks for arranging that grant for me belongs to Dr. Karin Frank. For financial support for my visit in Potsdam I want to thank Deutsche Akademische Austauschdienst (DAAD).

During my research work I have received a lot of help and support from my supervisors Doc. Robert B. O'Hara and Prof. Elja Arjas, for which I am highly grateful. They have given me a lot of freedom in planning and carrying out my research and given me opportunities to travel around the world visiting research institutes and participating in conferences and workshops. I have been able to develop my skills and obtain a good amount of international research experience. The freedom I have been given has been of great benefit for me on the way to becoming an independent researcher.

I want to address especial thanks to Dr. Frank Schurr, for our productive co-operation during my visits in Leipzig and in Potsdam. During the preparation of our three joint manuscripts he has shared with me his knowledge of ecological modelling, programming and scientific writing. He has patiently tolerated my enthusiastic way of working and sacrificed several nights for commenting my manuscripts. Thank you!

I am grateful to all the other co-authors of my manuscripts and ARGUE project researches for our co-operation. In particular, I would like to thank Prof. Timo Vesala for introducing me with atmospheric dispersal modelling and for helping me establishing contacts with other researchers on the same field. Also, I want to address especial thanks to Drs. Marja Ruohonen-Lehto and Kirsi Törmäkangas for assisting me with the political side of the GM risk assessment.
REFERENCES


Cited EU documents


