Meteorological input data requirements to predict cross-pollination of GMO Maize with Lagrangian approaches

Kai LIPSIUS[1]*, **Ralf WILHELM**[2], **Otto RICHTER**[1], **Klaus Jürgen SCHMALSTIEG**[1] and **Joachim SCHIEMANN**[2]

[1] Institute for Geoecology, Environmental Systems Analysis Group, Technical University Braunschweig, Germany
[2] Federal Biological Research Center for Agriculture and Forestry (BBA), Institute for Plant Virology, Microbiology and Biosafety, Germany

Modeling pollen dispersal to predict cross-pollination is of great importance for the ongoing discussion of adventitious presence of genetically modified material in food and feed. Two different modeling approaches for pollen dispersal were used to simulate two years of data for the rate of cross-pollination of non-GM maize (*Zea mays* (L.)) fields by pollen from a central 1 ha transgenic field. The models combine the processes of wind pollen dispersal (transport) and pollen competition. Both models used for the simulation of pollen dispersal were Lagrangian approaches: a stochastic particle Lagrange model and a Lagrangian transfer function model. Both modeling approaches proved to be appropriate for the simulation of the cross-pollination rates. However, model performance differed significantly between years. We considered different complexity in meteorological input data. Predictions compare well with experimental results for all simplification steps, except that systematic deviations occurred when only main wind direction was used. Concluding, it can be pointed out that both models might be adapted to other pollen dispersal experiments of different crops and plot sizes, when wind direction statistics are available. However, calibration of certain model parameters is necessary.

Keywords: GMO / pollen dispersal / cross-pollination / particle Lagrange model / gene flow

INTRODUCTION

Adoption of genetically modified (GM) crops has led to increasing interest in evaluating and controlling pollen-mediated gene flow from crop to crop. Only with reliable estimation of cross fertilization it is possible to establish management strategies to limit gene flow and control the genetic purity of harvests, in order to fulfill market demands for genetically pure products, *e.g.* achieving levels below the labeling threshold in the EU (EU, 2003).

Maize (*Zea mays* L.) is the second most cultivated crop plant in the world (Eastham and Sweet, 2002; Jarosz et al., 2005). Maize is predominantly wind-pollinated, and cross-pollination is promoted by the separation of male and female flowers on the plant and protandry (*i.e.* male flowering begins before female flowering) (Treu and Emberlin, 2000). Pollen dispersal from maize fields has been examined in a number of studies (*e.g.* Ma et al., 2004; Raynor et al., 1972).

Weather conditions will determine the pollen dispersal, with higher wind velocities and unstable conditions promoting dispersal. Maize has large pollen grains (around 80 to 100 µm) with high settling velocity, on the order of 20 cm.s^{-1} (Aylor et al., 2003). Hence, maize

pollen is predominately deposited inside the source field, with deposition decreasing rapidly with distance from the source field (Jarosz et al., 2003; Luna et al., 2001; Raynor et al., 1972). As discussed in Devos et al. (2005), several experiments have shown that the dispersal kernel (*i.e.* distribution of pollen dispersal with distance) is more leptokurtic than Gaussian, since the tail of the distribution is longer than that of normal distribution. Besides the meteorological conditions, major factors explaining cross-pollination rates are spatial configuration of the emitting and receiving fields (Klein et al., 2006; Lavigne et al., 1998), and synchronization of flowering.

A number of field trials with maize have been performed to study some of these effects (*e.g.* Henry et al., 2003; Jarosz et al., 2005, Ma et al., 2004, Raynor et al., 1972; Weber et al., 2005). Such studies can give direct estimates of pollen dispersal, but cannot be generalized to other geometries, crops and atmospheric conditions, and are therefore of limited use as predictive tools (Devos et al., 2005). Computer simulations circumvent these shortcomings because they allow a dramatically increased number of contexts studied.

Various modeling approaches have been used to predict pollen dispersal (Angevin et al., 2001; Arritt et al., 2003; Aylor et al., 2003; Bock et al., 2002; Jarosz et al., 2004; Klein et al., 2003; Klein et al., 2006;

* Corresponding author: K.Lipsius@tu-bs.de

Loos et al., 2003; Tolstrup et al., 2003; Yamamura, 2004). In spatially explicit models, dispersal is often modeled through individual dispersal kernel functions (IDF) (Klein et al., 2003, Lavigne et al., 1998). For pollen, these kernels are probability functions that describe the probability that a pollen grain is deposited at a given distance from its source (Lavigne et al., 1998). Dispersal kernels of single plants supposedly do not depend on the shapes, sizes and locations of the whole source and recipient plots (Devaux et al., 2005). The composition of the pollen cloud at a given site in a non-GMO field is determined by the pollen dispersal curves for all the plants in the neighborhood, whether close or farther away. Empirical models try to predict cross-pollination rates with different IDF. Devaux et al. (2005) and Klein et al. (2006) found that fatter-tailed functions predicting more long-distance dispersal, and therefore higher cross-pollination rates than the exponential function, were more consistent with the data. Angevin et al. (2001) also predicted adventitious presence due to cross-pollination with Klein's equation (2003). The main parameters of these empirical IDFs are main direction and mean speed of the wind during the entire course of flowering (Klein et al., 2003).

Apart from IDF approaches, several quasi-mechanistic or semi-empirical models have been developed (Arritt et al., 2003; Aylor et al., 2003; Jarosz et al., 2004; Loos et al., 2003). These models treat pollen grains as particles that move along simple diffusive-convective paths, either in normally-distributed Gaussian plumes (Ashton et al., 2001; Loos et al., 2003), or by a particle Lagrange (PL) or Lagrangian stochastic approach, in which independent particles follow statistically random paths (Arritt et al., 2003). The Gaussian approach was found to estimate both the quantity and direction of maize pollen movement with good accuracy, but underestimates cross-pollination near the source field (Loos et al., 2003). Jarosz et al. (2004) developed a quasi-mechanistic model. Most likely due to incorrect parameterization of turbulence, it also failed to correctly estimate pollen deposition immediately adjacent to the source. Arritt et al. (2003) showed that the PL model captured the sharp near-field deposition gradient evident in the observations, while under-predicting deposition at greater distances from the field. The PL approach has also been used in related dispersion applications, such as dispersion of particles (VDI, 2000), spores (Aylor and Flesch, 2001), and seeds (Nathan et al., 2003).

The probability of cross-pollination requires that foreign pollen out-compete local pollen and self-fertilization of the recipient plant (Aylor et al., 2003). So competition between locally shed pollen and pollen delivered from elsewhere on the wind needs to be considered.

We have compared two models for pollen dispersal and pollination by combining pollen deposition predictions with a pollen competition sub-model to predict cross-pollination rates. The first model is a PL model (VDI, 2000). The model simulates the path of individual pollen grains, using a single point meteorological observation to determine the mean pollen dispersal flow, and a random-walk approach to account for the effect of turbulence.

The second model uses the IDF concept (Klein et al., 2006). A transfer function (TF) models the probability for dispersion of an ensemble of pollen particles from source to sink. Transfer probabilities are determined using the Gaussian slender plume equation (Seinfeld, 1986). In contrast to other IDF models (Angevin et al., 2001; Klein et al., 2003), we did not use the average wind components for the entire flowering period to predict pollen dispersal. To investigate the required complexity of wind input data, we compared the deterioration in prediction accuracy when using aggregate wind statistics instead of the original measured time series of wind direction and velocity.

The objectives of this study were

- to compare the two Lagrangian approaches;
- to incorporate meteorological information in cross-pollination models;
- to assess the dependence of model performance on the quality of meteorological data;
- to test the models for their predictive power.

RESULTS AND DISCUSSION

Comparison of models

For better comparison, the figures in the following section are all structured in the same way. In A, measured and predicted cross-pollination rates are shown in grey-scaled squares at their respective coordinates in the field. Transects with cross-pollination rates against distance from the central GM field are plotted logarithmically in B. The dashed line represents the minimal measurement result of one cross-pollination event in 3000 kernels. Measured cross-pollination rates of 0% are shown in the figures at half of this value. Transects were aligned orthogonally, north, south, east and west of the GM field (the winds were predominantly from the west). In C, a simple scatter plot of predicted *versus* measured values is shown, together with the correlation coefficient r. To investigate whether residuals appear to be systematic or are distributed randomly around the predictions, the logarithmic deviation from the logarithm of the measured value is plotted against distance from the GM in D.
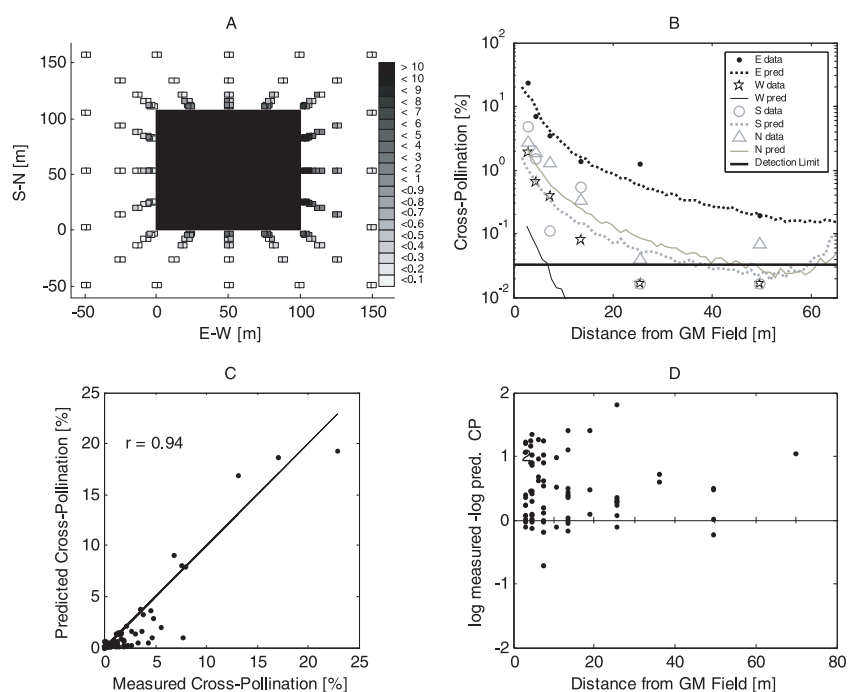
Figure 1. Cross-pollination rates for particle Lagrange model with data from 2000. Meteorological data is time series. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.

Table 1. Parameter estimations of GM cross-pollination rate prediction with particle Lagrange (PL) and transfer function (TF) models for 2000, with pollen competition sub-model (3 parameters, η_e, η_f, α).

Model	Wind	Maximum likelihood estimator					Least sum of square estimator				
		η_e	η_f	α	$-L_n$	ME	η_e	η_f	α	S_n	ME
PL	$u(\phi)$	3.01	119	0.051	999	0.821	1.75	103	0.060	0.019	0.830
TF	$u(\phi)$	2.50	131	0.036	1009	0.790	1.60	108	0.045	0.023	0.800
TF	u, ϕ	2.64	126	0.042	963	0.806	1.68	105	0.050	0.021	0.816
TF	$\overline{u}, \phi$	2.66	124	0.046	851	0.861	1.53	101	0.058	0.016	0.866
TF	$\overline{u}, \overline{\phi}$	3.61	94	0.073	978	0.915	2.24	99	0.085	0.009	0.919

$u(\phi)$ designates wind data as time series, u, ϕ wind data in classes, $\overline{u}, \overline{\phi}$ mean values over entire flowering period. L_n is the log likelihood value and S_n the sum of squared errors.

Particle Lagrange model (PL), data from 2000 (wind data as time series, $u(\phi)$)

The prediction of the 2000 data was generally good (Fig. 1A), which is also shown by the modeling efficiency of 0.83 (Tab. 1). Logarithmic residuals decrease with distance (Fig. 1D). In 2000, the cross-pollination rates at small distances were underestimated on the majority of transects. The very high cross-pollination rates immediately adjacent to the field in an easterly direction were well predicted. In all other directions, the observed

cross-pollination rates were much lower, and the model tended to underestimate (Fig. 1A).

The transects in Figure 1B illustrate that prediction in the south and north directions, which were orthogonal to the main wind direction, was good and showed no systematic bias. Prediction in main wind direction (downwind; east) was even more precise; whereas the model systematically underestimated cross-pollination values to the west (upwind) of the GM plot (Fig. 1B). The dispersal of GM donor pollen upwind was slightly underestimated by the PL model. The oscillations in transects predicted
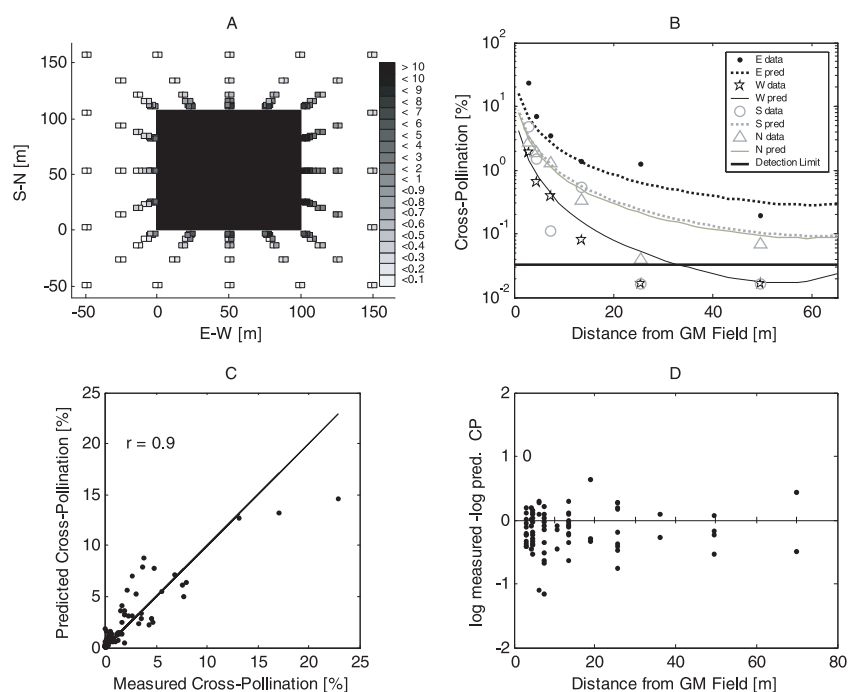
Figure 2. Cross-pollination rates for transfer function model with data from 2000. Meteorological data is time series. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.

with the PL model (4B) were due to the random nature of the PL model.

Transfer function model (TF), data from 2000 (wind data as time series, $u(\phi)$)

The TF model was also able to predict the measured cross-pollination rates for data from 2000. The modeling efficiency was 0.79 (Tab. 1). In contrast to the PL model, values near the GM field edge, *i.e.* at small distances in Figure 2D, were rather overestimated. However, the residuals were generally distributed randomly, with deviations decreasing with distance, which was also true for the logarithm. The transects in Figure 2C showed good fit of the data, without systematic bias. Compared to the PL model, the simulated curve shows a much smoother decrease with distance, because the model incorporates no random walk.

Particle Lagrange model, data from 2001 (wind data as time series, $u(\phi)$)

The goodness of fit of both modeling approaches was much worse in 2001 compared to 2000. For the PL model, the model efficiency decreased to 0.421 (Tab. 2). The

plot of logarithmic residuals *versus* distance (Fig. 3D) showed a general tendency towards overestimation of measured cross-pollination rates, especially with increasing distance from the GM field. Prediction was good upwind and downwind (Fig. 3B) although cross-pollination downwind was systematically overestimated. In contrast, the measured high level of cross-pollination for all distances north of the GM plot was not mirrored in the modeling results. Especially the high cross-pollination rate at 2.5 m east-northeast of the GM plot was not simulated (Fig. 3A). The same holds true for the TF model (Fig. 4A).

Transfer function model, data from 2001 (wind data as time series, $u(\phi)$)

In 2001, prediction of the TF model was much less precise than in 2000, with a modeling efficiency of 0.523 (Tab. 2). The high cross-pollination rates in the northerly direction, orthogonally to the main wind direction, were predicted much better than with the PL model. The decrease of the cross-pollination rates with increasing distance from the GM plot in the southerly direction was however underestimated (Fig. 4B). For both directions parallel to the main wind direction, the decrease of cross-pollination rates with distance from the GM field edge

Table 2. Parameter estimations of GM cross-pollination rate prediction with particle Lagrange (PL) and transfer function (TF) model for 2001, with pollen competition sub model (3 parameters, η_e, η_f, α).

Model	Wind	Maximum likelihood estimator					Least sum of square estimator				
		η_e	η_f	α	$-L_n$	ME	η_e	η_f	α	S_n	ME
PL	$u(\phi)$	1.01	95	0.200	2912	0.421	1.00	98	0.165	0.0273	0.423
TF	$u(\phi)$	1.04	68	0.202	1740	0.523	1.02	75	0.168	0.0264	0.524
TF	u, ϕ	1.08	71	0.191	1712	0.522	1.04	77	0.165	0.0264	0.524
TF	$\overline{u}, \phi$	1.00	68	0.201	1754	0.543	1.00	81	0.154	0.0252	0.545
TF	$\overline{u}, \overline{\phi}$	1.00	119	0.172	2242	0.504	1.00	129	0.116	0.0269	0.515

$u(\phi)$ designates wind data as time series, u, ϕ wind data in classes, $\overline{u}, \overline{\phi}$ mean values over entire flowering period. L_n is the log likelihood value and S_n the sum of squared errors.
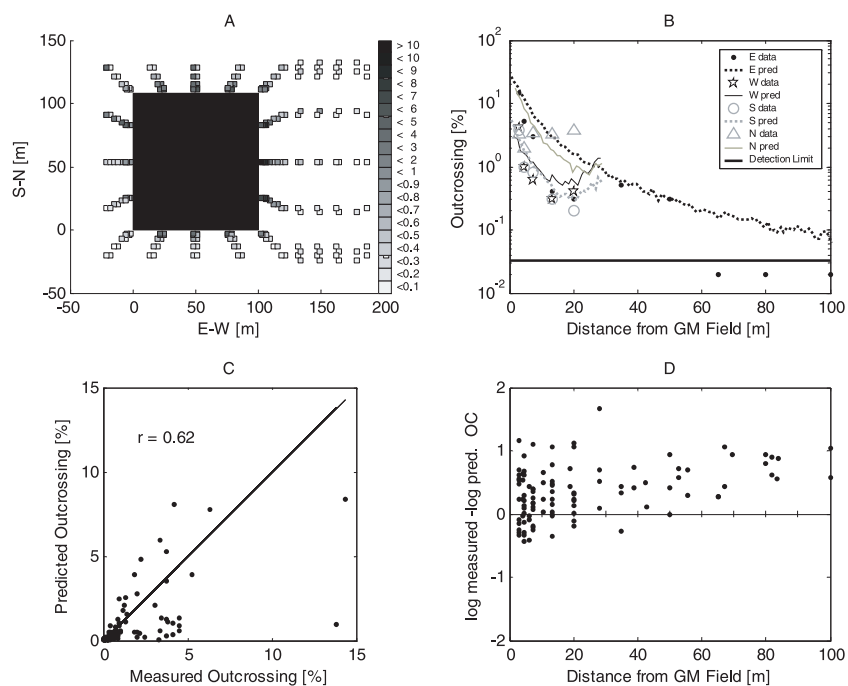

Figure 3. Cross-pollination rates for particle Lagrange model with data from 2001. Meteorological data is time series. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.

predicted by the TF model was lower than the measured values (Fig. 4B). The plot of logarithmic residuals *versus* distance from GM plot also confirmed a general bias towards overestimation of cross-pollination at larger distances (Fig. 4D).

Summary of model comparison (simulation with wind data as time series, $u(\phi)$)

In summary, the PL and TF models both achieved good prediction of cross-pollination rates for the data from 2000. In 2001, both models achieved lower goodness of fit, but still yielded acceptable predictions throughout the field. In 2001, the predicted cross-pollination values differed considerably between the two models. However, the same observation points account for the highest residuals in both models. The deterioration of model performance in 2001 can be explained by high infield heterogeneity in plant development, which is supported by observations of individual flowering times. However observations of flowering times were too infrequent to be incorporated in the model. When male flowering of the GM maize and female flowering in
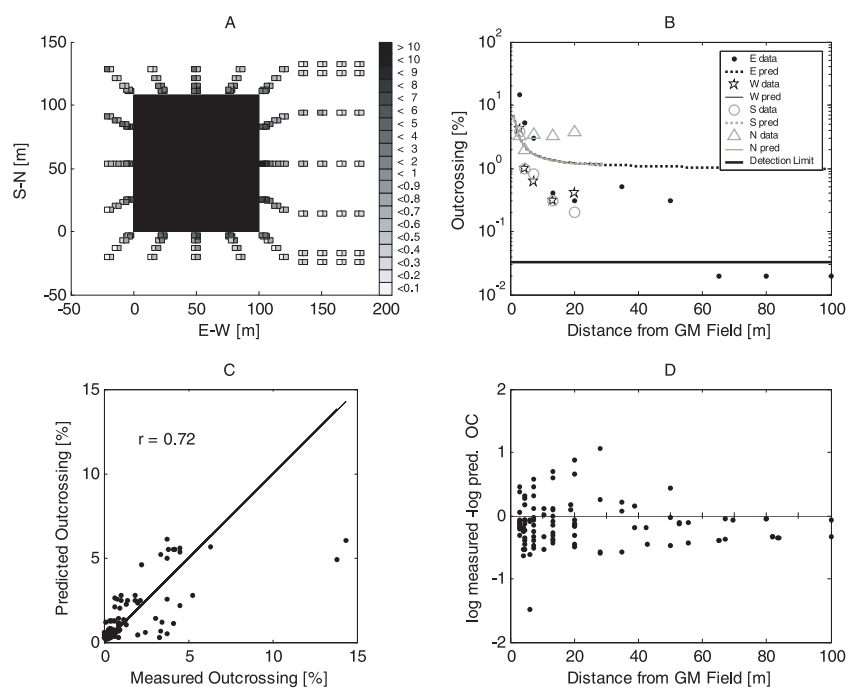
Figure 4. Cross-pollination rates for transfer function model with data from 2001. Meteorological data is time series. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.

the receptor field deviate largely from the assumption of perfect synchrony, cross-pollination were overestimated. Single plants with high cross-pollination rates that are under-predicted can be explained by asynchrony of development of the plant itself and surrounding receptor maize plants, which would lead to lower pollen competition from receptor pollen and higher cross-pollination. Another possible explanation for the difference between the years is the prevalent wind conditions during the flowering periods. In 2000 the variability in wind was much lower, and mean wind velocity was higher (Fig. 14). The models might not be suitable to predict pollen dispersal for low wind speeds. If the models overestimate pollen dispersal at low wind speeds, cross-pollination is overestimated at low and changing winds as measured in 2001.

The transects show that cross-pollination rates are predicted to rise slightly at the outer field edge of the receptor fields. An increase in cross-pollination rates at the outer field edge seems to be reasonable, since the number of neighbor plants and therefore receptor pollen sources (especially immediate neighbors with high competitive advantage, because they can pollinate without wind transportation) decreases. However, no measurements were taken close enough to the field edge to confirm the predicted rise.

Effect of the quality of meteorological input data on model performance

In this section are presented the results of the models using different complexities of meteorological input data. The results when the full information is used, *i.e.* the original time series of wind direction and wind velocity during the whole flowering period, were discussed above.

For the TF model, the information content of meteorological input data was decreased in three steps. In the first step, wind measurements were sorted into distinct classes for wind speed and wind direction. Wind velocity was grouped into nine classes, according to the classification rules of the WMO (Seinfeld, 1986). The wind direction was no longer considered as directly measured, but only frequency of measurements in 10° steps was used. In the next simplification step, mean wind velocity was assumed for all wind directions. The most simplified input data were mean wind speed and main wind direction.

For the PL model, cross-pollination rates were calculated with the competition sub-model after each hour during the flowering period in 2000, *i.e.* the first prediction was derived only from the pollen distribution after one hour, the second prediction from pollen distribution for hours one and two, and so on. The resulting model efficiencies are shown in Figure 5 and discussed below.
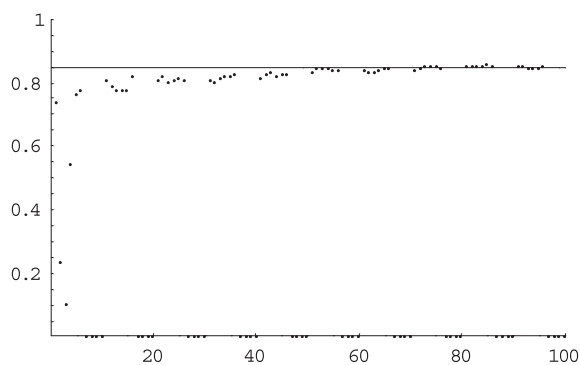
Figure 5. Modeling efficiency of cross-pollination rate prediction with particle Lagrange model during flowering period in 2000 assuming for each time step, that pollen shed stopped and no further pollen was transported.

Particle Lagrange, data from 2000 (time series, $u(\phi)$)

In Figure 5, it can be seen that the modeling efficiency varied strongly for model predictions using only the first hours, assuming values of down to less than 0.2. After the third day of the male flowering period, the goodness of fit hardly varied further. Modeling the pollen transport and subsequent cross-pollination in the later stages of the flowering period therefore hardly improved the model performance.

Transfer function, data from 2000 (wind direction in classes, u, ϕ and $\overline{u}, \phi$)

Predictions are shown in Figure 6 for wind direction and wind velocity in classes (u, ϕ), and in Figure 7 for mean wind velocity ($\overline{u}, \phi$). Model efficiency was high with 0.806 and 0.861, respectively, but systematic errors increased. For the transects parallel to the main wind direction, the predictions of cross-pollination values were systematically but only slightly underestimated (Figs. 6B, 7B). In contrast, cross-pollination rates at larger distances north of the GM plot were overestimated. However, for all directions together, no systematic bias with distance was observed (Figs. 6D, 7D).

Transfer function, data from 2000 (mean wind direction and wind velocity, $\overline{u}, \overline{\phi}$)

While yielding the highest modeling efficiency of all model runs with 0.915 (Tab. 1), predictions exhibited strong systematic deviations. The high modeling efficiency was due to the good estimation of the high cross-pollination rates near the field in the downwind

Table 3. Parameter estimation results for TF model shape parameters for longitudinal, *ax*, transversal, *ay*, and vertical, *az*, dispersion.

Year	ax	ay	az
2000	21	33	4
2001	76	61	5

direction. The simplest model version underestimated cross-pollination rates in all directions but downwind (Fig. 8B). A general bias towards underestimation of cross-pollination rates can be observed for all distances from the GM field edge, as can be seen in Figure 8D. Overall, the prediction quality with only mean wind direction and velocity was worst, although the plume shape parameters *ax*, *ay* and *az* of the dispersal function were fitted for optimal performance (Tab. 3).

Transfer function, data from 2001 (wind direction in classes, u, ϕ and $\overline{u}, \phi$)

As for prediction with wind data as time series, the cross-pollination predictions also had lower model efficiencies, together with more bias in 2001 compared to 2000. The simulation results still lie within the range of variability of the observations. The distribution of residuals for wind direction in classes (Fig. 9 for wind velocity classes, u, ϕ, and Fig. 10 for mean wind velocity, $\overline{u}, \phi$) are similar to the results with the time series in Figure 4, with overestimation at larger distances, and failure to simulate the high value east-northeast and at larger distances north of the GM plot (Figs. 9 and 10).

Transfer function, data from 2001 (mean wind direction and wind velocity, $\overline{u}, \overline{\phi}$)

As for the 2000 data set, predictions based on the main wind direction and velocity in 2001 were highly biased. However in contrast to the 2000 data set, model efficiency was also lower for this scenario. Orthogonally to and upwind of the main wind direction, all cross-pollination rates were largely underestimated. Only downwind of the GM plot, bias was not systematic (Fig. 11B). The low model efficiency can be explained by the wind conditions in 2001. As mentioned before, in 2001 the wind blew from different directions at low velocities. Prediction of pollen dispersal using only main wind direction therefore increased systematic errors. The dispersal parameters in 2001 were estimated to be much higher than in 2000, because in 2001 cross-pollination rates were similar in all directions (Tab. 3).
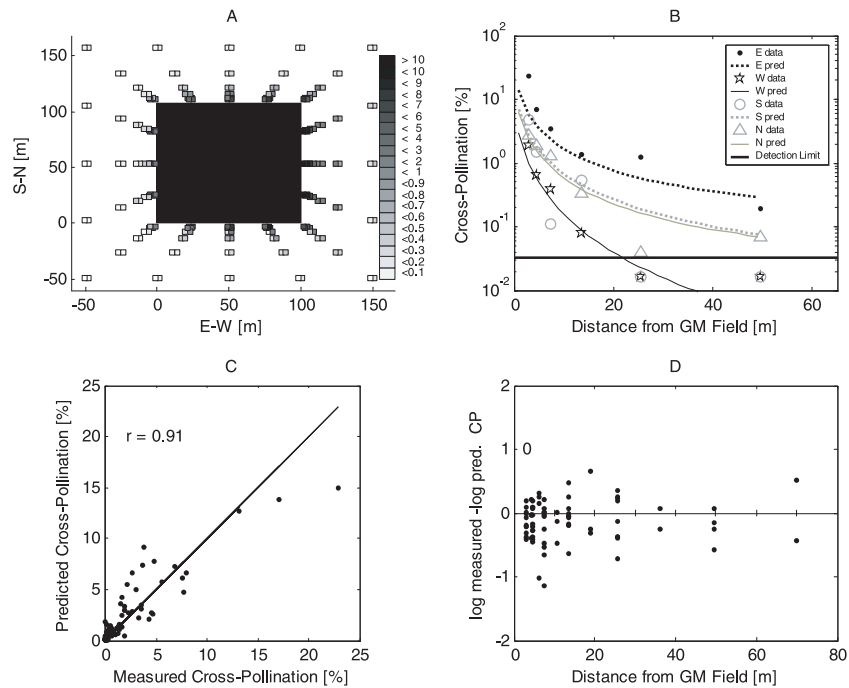
Figure 6. Cross-pollination rates for transfer function model with data from 2000. Meteorological data is wind direction and wind velocity in classes. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.
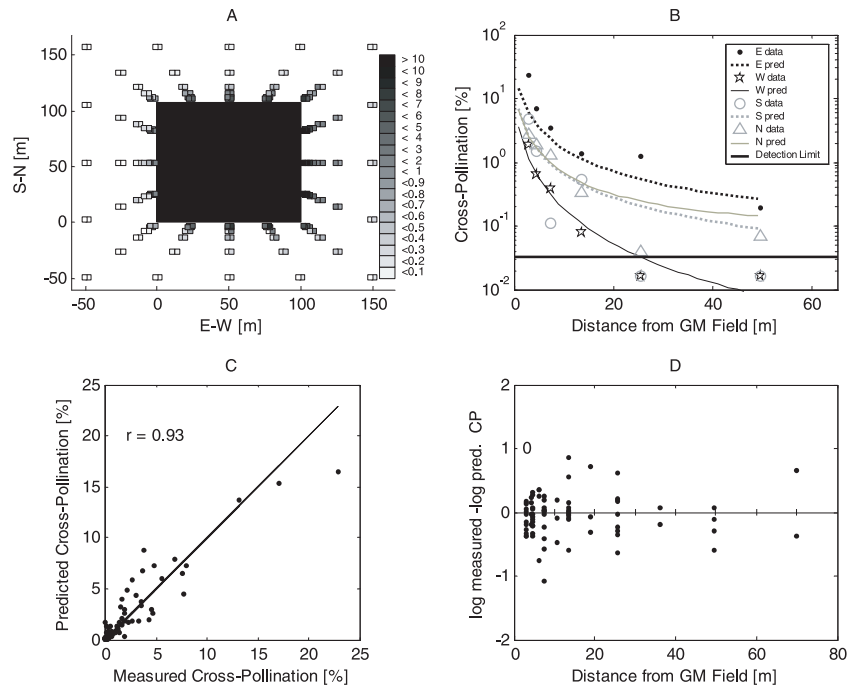

Figure 7. Cross-pollination rates for transfer function model with data from 2000. Meteorological data is wind direction in classes and mean wind velocity. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.
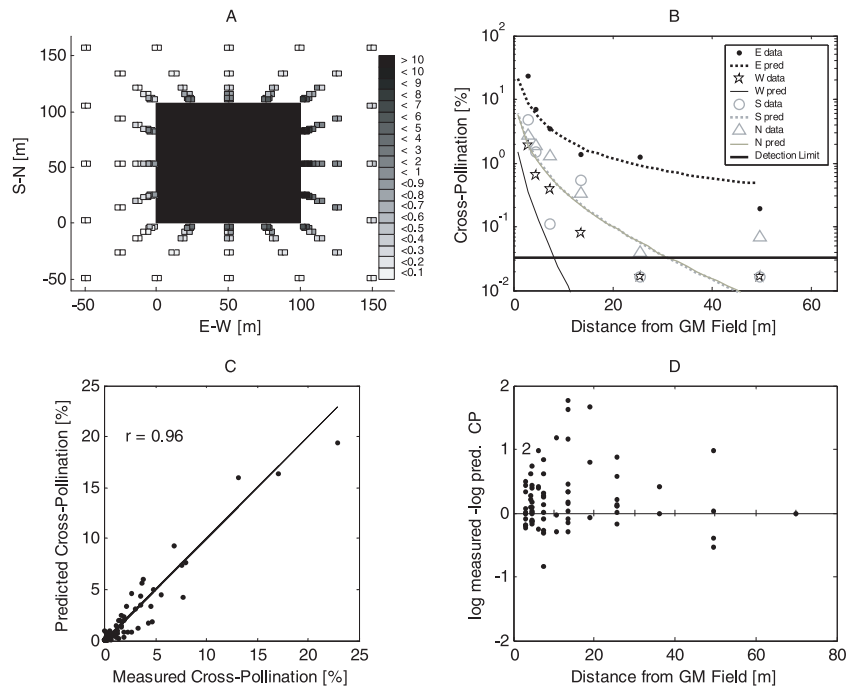
Figure 8. Cross-pollination rates for transfer function model with data from 2000. Meteorological data is main wind direction and mean wind velocity. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.
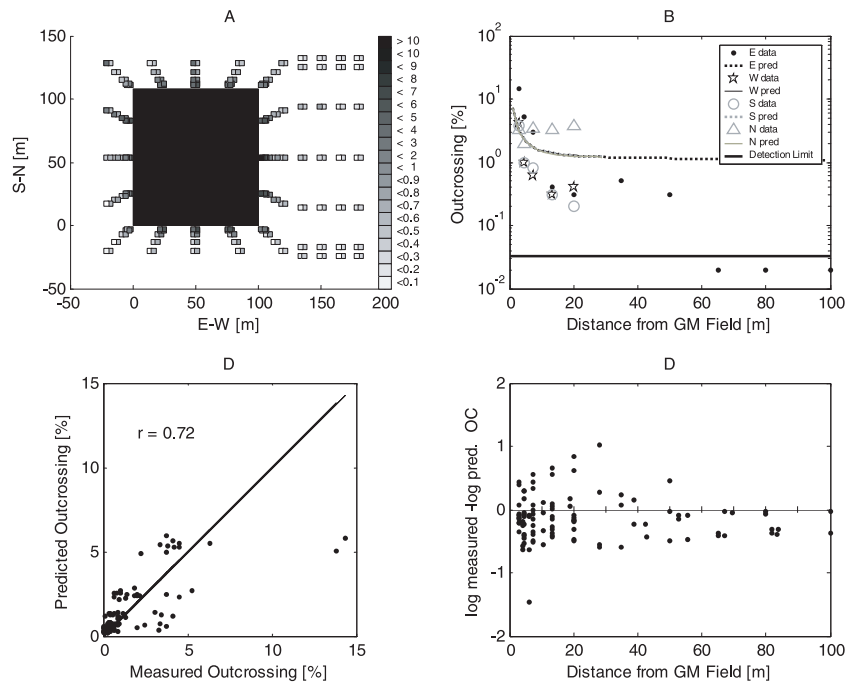

Figure 9. Cross-pollination rates for transfer function model with data from 2001. Meteorological data is wind direction and wind velocity in classes. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.

Figure 10. Cross-pollination rates for transfer function model with data from 2001. Meteorological data is wind direction in classes and mean wind velocity. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.
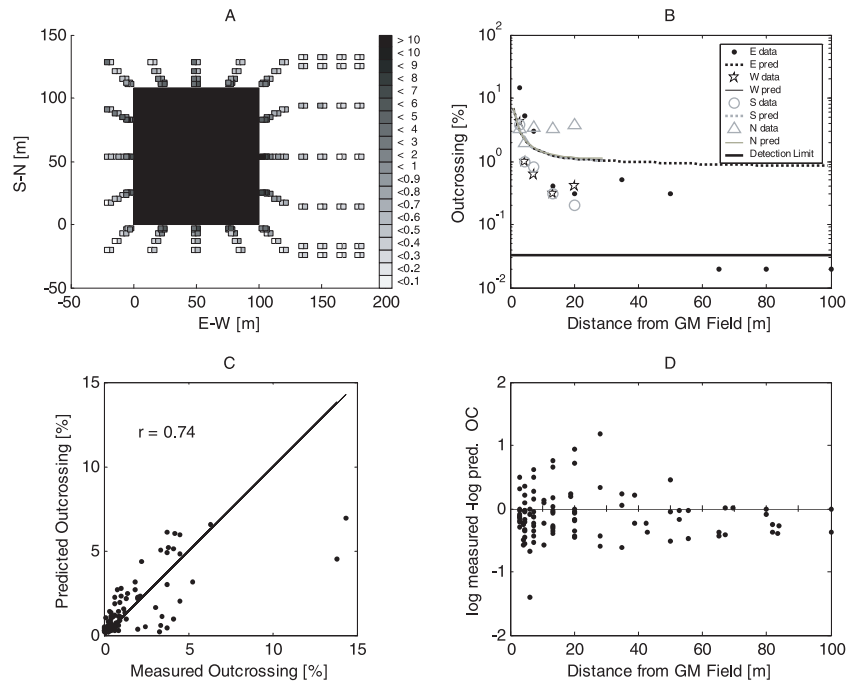

Figure 11. Cross-pollination rates for transfer function model with data from 2001. Meteorological data is main wind direction and mean wind velocity. A: measured left, predictions right boxes; B: log-transformed transect in four wind directions; C: predictions *versus* observation and correlation coefficient r; D: residuals of log-transformed data against distance from GM field edge.
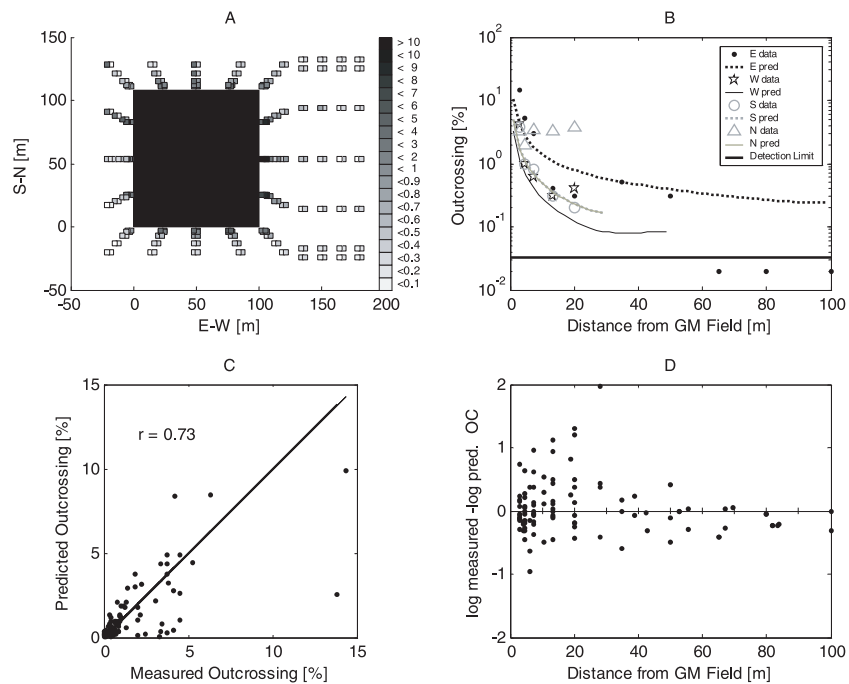
Summary of the effect of the quality of meteorological input data

The results can be summarized as follows.

(i)   Model efficiencies were much higher in 2000 than in 2001, regardless of the model type. This could be due to inadequacy of the models to predict pollen dispersal for low wind velocities, as prevalent in 2001. In retrospect, this could also be ascribed to large observed in-field heterogeneities in plant development in 2001 that were not measured in high spatial resolution, and therefore could not be incorporated into the model.

(ii) The overall model efficiencies of the transfer function model were hardly decreased when wind direction was sorted into classes. This also holds true for assuming a mean wind velocity for the whole flowering period.

(iii) The model version with mean wind direction and velocity for the whole flowering period showed high systematic deviations with underestimation in all directions except the main wind direction, and is therefore not recommended, despite yielding the highest modeling efficiency for the 2000 data.

Competition model parameters and transferability of models

For the competition model with three parameters, to account for the edge effect, the values of the competition factor η_f and η_e were similar in both years (Tabs. 1 and 2). Only the α value, which indicates the range of the edge effect, deviated an order of magnitude between years. For the 2000 data, the TF model parameter estimates were more similar to the PL parameter estimates than for 2001, except for η_e, which generally showed very low variation in 2001. Variation in parameter values was very low between TF model scenarios, with wind direction in classes or as time series. Only parameters for main wind direction over the entire flowering period deviated substantially (Tabs. 1 and 2).

Similarity between parameter estimates was generally high enough to feel confident about the ability of the Lagrangian models together with the competition model to be used to predict the spatial distribution of cross-pollination rates and scenario modeling.

The model concept is well suited to explicitly account for spatial designs of source and receptor plots with several sources in an agricultural landscape, rather than only a single emitter and a single receptor. This is more representative of potential coexistence contexts, however validation data sets are not available (Devos et al., 2005).

It must be concluded that deviations between the estimations of the pollen competition model parameters are not negligible. Transferability to other sites and geometries is therefore not guaranteed, and pollen competition parameters should be calibrated.

Cross-pollination maps

The spatial distribution of predicted cross-pollination rates are presented for the receptor fields as contour maps in Figure 12. The nature and complexity of the different model approaches is mirrored in the maps. The particle Lagrange (PL) model yielded very small contours with highly resolved variability (Fig. 12A). This small-scale variability is caused by the random nature of the PL model. The smoothness of the contours increased with the number of repeated runs. The contour plots from the transfer function (TF) model rendered much smoother contours. Their general patterns however matched those produced by the PL approach (Fig. 12B−D). The patterns of the contour maps of the TF model were similar for all wind input data types, with the exception of mean wind velocity and direction. Here the contours were much simpler and appear artificial. The same observations held true for the predictions in 2001 (not shown here).

CONCLUSION

For using the models to predict GM cross-pollination from a central maize plot into an adjacent surrounding conventional maize plot we arrive at the following conclusions.

(i)   Similar results were obtained for pollen transport models with particle Lagrange (PL) and the much simpler Lagrangian transfer function (TF) approach.

(ii) The model performance varied strongly between the two years considered. For the first year of the experiment, the measured cross-pollination rates could be well predicted by both models. In the second year, the goodness of fit decreased considerably. This is mainly due to in-field heterogeneity in flowering and different meteorological conditions.

(iii) Using wind statistics instead of the time series, the TF model performances were hardly worse, as long as wind direction was considered in frequency of 10° steps. Systematic errors occurred slightly more frequently, with a tendency towards underestimation
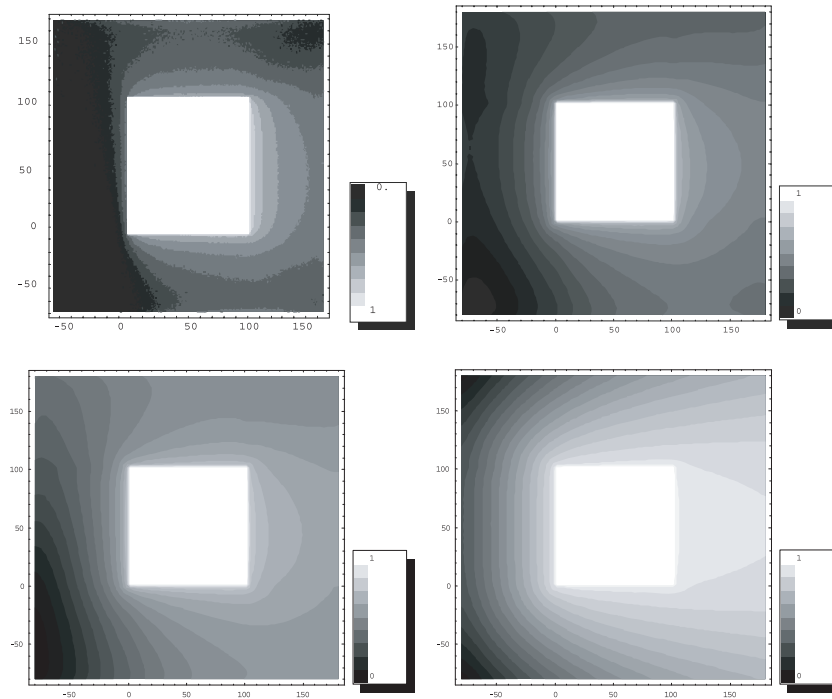
Figure 12. Contour maps of predicted cross-pollination rates for 2000 using different qualities of meteorological input data.
A: PL time series; B: TF model time series; C: wind direction and velocity in classes; D: mean wind direction and velocity. The central GM donor plot is shown in white.

of cross-pollination in all but the main wind direction, because pollen transport in the main wind direction received too heavy weighting.

(iv) Systematic underestimation of cross-pollination in other directions than the main wind direction increased strongly when pollen dispersal was simulated using main wind direction and mean wind velocity for the entire flowering period. Therefore it can be concluded that wind direction needs to be considered for pollen dispersal predictions.

(v) To improve prediction of gene flow from wind-pollinated plants, it is not necessary to use sophisticated pollen transport models, but to improve the description of plant development, *e.g.* spatio-temporal distribution of male and female flowering phases.

(vi) Variability of the fitted competition parameters was low, mainly within the same order of magnitude. Larger differences in parameter values were found only when main wind direction was considered. Transfer of the models to other sites seems possible; however they should be tested with different sized experiments under different meteorological conditions.

MATERIALS AND METHODS

Experimental design and data acquisition

Cross-pollination data from farm scale field trials, which were carried out at the Federal Biological Research Centre for Agriculture and Forestry (BBA) in Braunschweig, Germany, in 2000 and 2001, were used for model application and verification (Meier-Bethke and Schiemann, 2003). The model predictions were compared with measured cross-pollination rates in conventionally bred maize (pollen recipient) at distances up to 100 m around a source plot of isogenic but herbicide tolerant GM maize (pollen donor) of about 1 ha (approx. 100 m × 108 m). Around the GM maize plot a maize-free strip of about 2 m was established. The maize was planted in rows at intervals of 0.75 m with a distance of approx. 0.16 m between the plants in a row. The plants grew to an average height of 2.45 m, topped by tassels (male inflorescences) of about 30 cm, so most of the pollen was emitted at heights of about 2.30 to 2.60 m. The mean maize ear height (female inflorescence) was located at 0.85 m. Main flowering periods were estimated for the whole field. Considerable in-field heterogeneity in flowering times was observed, especially in 2001. Due to the lack of more
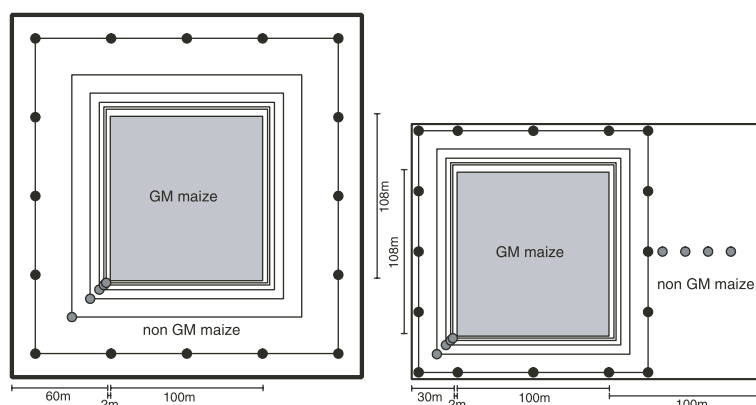
Figure 13. Experimental design in 2000 (A = left) and 2001 (B = right). GM field is grey, non-GM field within thick line. Thin lines mark sampling distances, with black points marking end points of transects, and grey points marking exemplary transect sampling points.

specific information on this heterogeneity, flowering was considered to be homogeneous in the models.

The source plot was surrounded by concentric squares of sampling points, which were spaced at approximate 22.5° intervals or lines of equidistant sampling points. Sampling points were located at distances of 3, 4.5, 7.5, 13.5, 25.5 and 49.5 m distance in 2000, and at 3, 4.5, 7.5, 13.5, 20, 35, 50, 65, 80 and 100 m in 2001 from the edge of the GM field (Fig. 13). At each sampling point, about 60 ears were sampled. From a sub-sample of 1.5 kg of shelled seeds, the target was to test at least 3000 seeds (actual minimum 2497) for herbicide resistance (Meier-Bethke and Schiemann, 2003). The maize seeds were germinated in the greenhouse, and after 10 days the seedlings were sprayed with the herbicide Glufosinate, and after 3 weeks the surviving seedlings were counted (*i.e.* genetically modified seedlings that carried the transgene conferring resistance to the herbicide). The theoretical minimal measured cross-pollination rate was 0.033%. However, due to sampling variability, the probability for detection of cross-pollination was greater than 0.95 only for cross-pollination rates above 0.1%.

Meteorological data

A meteorological station was positioned immediately adjacent to the experimental site. Wind direction and wind velocity were measured at a single point 2 m high at 1-h intervals. Further available data include relative humidity, precipitation and air temperature. Mean wind velocity was 1.4 m.s^{-1} and 0.7 m.s^{-1} in 2000 and 2001, respectively. A summary of the wind measurements during the flowering period in 2000 and 2001 is shown in Figure 14.
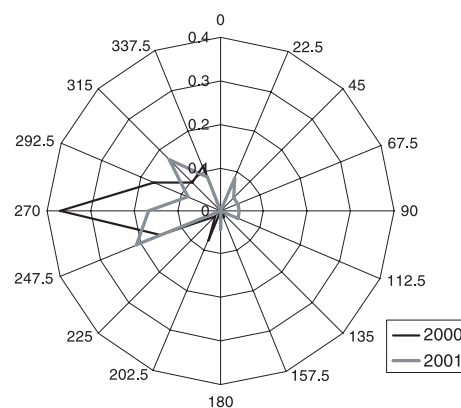

Figure 14. Frequency of hourly wind directions during flowering periods in 2000 and 2001.

Cross-pollination models

Modeling concept

The cross-pollination, *i.e.* gene flow *via* pollen, of anemophilous higher plants like maize comprises several processes at different scales. Once a pollen grain has been emitted from a donor plant, it is subject to atmospheric dispersion (Aylor et al., 2003). During this phase the pollen grain can be treated as any other particle, and its movement described in terms of classical fluid dynamics transport theories. In this study, two different Lagrangian approaches were employed, a stochastic Lagrange particle model (PL), and a model based on explicit probability density functions for pollen transfer probability from a point source to a point sink (TF), also known as individual dispersal function (Klein et al., 2003). The stochastic particle model is based on the tracking of large ensembles

of individual particles (VDI, 2000). General conclusions can only be drawn from the statistics of the ensemble. The advantage of this approach is that dispersal can be simulated for each weather situation during anthesis. Meteorological data have to be provided in high resolution to this end.

The alternative approach, TF, is based on the explicit formulation of the concentration statistics in terms of the stochastic properties of the paths of ensembles of particles. Transport of a particle from a location (x', y', z') to a location (x, y, z) is mediated *via* a probability density or transfer function $Q(x, y, z \mid x', y', z')$ (Seinfeld, 1986). The transfer function depends on the statistics of the wind field during pollination. The total amount of pollen that reaches a certain location is then derived by the integral over all sources.

After atmospheric transport, the pollen eventually reaches the upper layer of the canopy of the recipient crop. Close to and inside the canopy, it competes with the pollen of the recipient population to fertilize (Luna et al., 2001). The probability of a cross-pollination event is thus not simply proportional to the number of donor pollen grains. Among other factors, such as the phenology of the recipient population, j, it depends on the position of the plant within the stand.

It should be stressed that the data obtained in cross-pollination experiments do not reflect single dispersal events. They reflect rather the history of meteorological and phenological conditions during the entire female flowering period of the receptor population.

Particle Lagrange model

The particle Lagrange (PL) model for atmospheric dispersion of the VDI (2000) was used in this study. The change of position of a particle during a time step of length τ is determined by the mean wind velocity $\vec{V}$, the turbulent velocity $\vec{u}$ and an additional velocity vector $\vec{U}$ due to gravitational settling:

$$\vec{x}_{new} = \vec{x}_{old} + \tau \left(\vec{V}(\vec{x}_{old}) + \vec{u} + \vec{U} \right). \quad (1)$$

The turbulent velocity u is updated during each time step *via* a Markov process, which is governed by the tensor Ψ and the stochastic velocity increment $\vec{w}$. The components of the tensor Ψ depend on the atmospheric diffusivity tensor K. The components of K are related to the local meteorological conditions (VDI, 2000):

$$\vec{u}_{new} = \Psi . \vec{u}_{alt} + \vec{w}. \quad (2)$$

Input data of the model are the meteorological conditions, such as wind velocity and wind direction, the geometry of the source, thermal stability of the atmosphere, and

roughness lengths of the terrain. Here the latter two were kept constant throughout. We assumed a constant pollen release height of 2.45 m, and the end point of simulation where we considered pollen as deposited was 0.85 m, corresponding to the male and female inflorescences, respectively. The PL model was used to predict pollen dispersal for the whole flowering period using the hourly time series of wind direction and wind velocity.

Lagrange transfer function model

The core of the Lagrange transfer function (TF) model is the transition probability density function or individual dispersal function (Klein et al. 2003) $Q(x, y, z \mid x', y', z')$ for the transport of a particle from location (x', y', z') to (x, y, z).

The pollen concentration at point x, y, z is then given by the integral over the product of the source density $S(x', y', z')$ and the transition probability density or individual dispersal function Q:

$$c(x, y, z) =$$
$$\iiint\limits_{G} Q\left(x, y, z \mid x', y', z'\right) S\left(x', y', z'\right) \, dx' \, dy' \, dz'. \quad (3)$$

For the transition probability function, a general Gaussian form is used for particle transport from a point source at location (x', y', z') in a stationary wind field with wind velocity u and direction ϕ (Seinfeld, 1986). Loos et al. (2003) showed that the Gaussian plume model predicted maize pollen dispersal with good accuracy for locations further than 7 m from the source field for simple terrains. Starting from this approach, we used the slender plume equation for the individual dispersal function to guarantee the highest comparability to the PL approach. According to Seinfeld (1986) transition probability function Q has the form

$$Q\left(x, y, z \mid x', y', z', u(\phi)\right) = \frac{q}{2\pi \sqrt{a_y a_z} \, r}$$
$$\times \exp\left[\frac{u(\phi)}{a_x} r - (x - x') \cos \phi - (y - y') \sin \phi \right] \quad (4)$$

with

$$r^2 = \left((x - x') \cos \phi + (y - y') \sin \phi\right)^2 + \left(\frac{a_x}{a_y}\right)$$
$$\times \left((y - y') \cos \phi - (x - x') \sin \phi\right)^2 + \left(\frac{a_x}{a_z}\right)(z - z')^2. \quad (5)$$

Equation (4) is valid for a point source in a wind field with constant wind velocity u and wind direction ϕ.
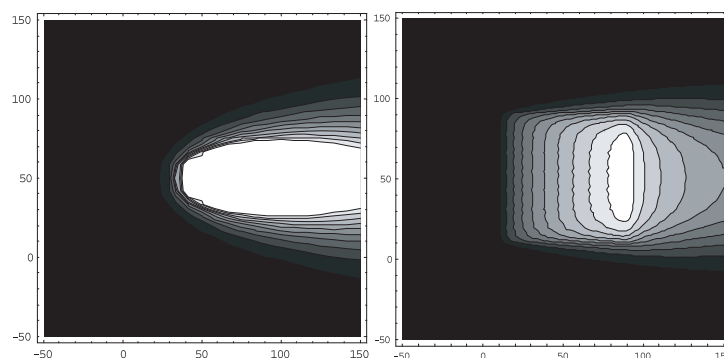
Figure 15. Example pollen deposition concentration contour plots. Wind blew from the left. For A (left) a point source at (50, 50) and B (right) a superposition of point sources which cover the area of a field within the edges (0–100, 0–100).

We approximated the difference in height from z' to z to be 1.6 m for the whole maize field. The pollen transport on even ground is therefore basically a two-dimensional problem. The shape of the plume in Equation (4) is determined by the parameters ax, ay and az, describing the dispersion longitudinally, transversally and vertically to the wind direction, respectively (Seinfeld, 1986).

In reality, the source, *i.e.* the maize field, is not a continuous source but an ensemble of point sources, *i.e.* the single plants. Therefore the integral (3) can be approximated by summing the individual dispersal functions for pollen transport from all point sources, *i.e.* the position of plants $((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$ to the considered point (x, y) (Fig. 15):

$$c_{Pollen}(x, y) = \sum_{i=1}^{n} Q(x, y \mid x_i, y_i) S(x_i, y_i). \quad (6)$$

The number and positions of point sources are chosen at a resolution high enough to represent the maize fields, but not at every plant position to save computing time. Point sources of the GM pollen are situated within the GM field, while the conventional fields contain the point sources for conventional pollen, thus covering the whole GM or conventional plot, respectively.

In each model run, only one wind direction ϕ and one wind velocity u as considered. When different wind directions or velocities were considered, the model was run repeatedly, and the pollen deposited at each sink site (x, y) was simply summed for each considered u-ϕ-combination. The pollen deposition distribution calculated with the single plume equation (Eq. (4)) was weighted according to the frequency of readings with the given u-ϕ-combination. In this way, we could consider wind data as a time series or using aggregate statistics.

Competition of pollen

In the case of maize, pollen grains are released from the male inflorescence (tassel) at the top of the maize plant. The female ear(s) bloom at the middle part of the stem, exposing stigmas (silk). It seems reasonable to describe the competition between donor pollen and recipient pollen on the stigmas of target plants simply by their relative proportion at the target site. However, the contribution of nearby receptor plants within a stand has to be given special attention, because pollen from neighboring plants can reach the target from within the canopy without wind transportation (Aylor et al., 2003). To take this into account, we developed a competition model. The competitive advantage of the receptor pollen was considered by multiplying the receptor density by a weighting factor η. However, at the edge of the receptor field, the number of neighboring plants is lower. Loos et al. (2003) found that the Gaussian model under-predicted outcrossing in the first 7.5 m of the receptor field. To account for this edge effect, the weighting factor was considered a function of the distance, d from the edge of the field in the following way:

$$\eta(d) = \eta_f(1 - e^{-\alpha d}) + \eta_e e^{-\alpha d}. \quad (7)$$

The cross-pollination probability, $P_{out}(x, y)$, was derived from the predicted donor, $c_{donor}(x, y)$, and recipient, $c_{receptor}(x, y)$, pollen concentration at position (x, y) by

$$P_{out}(x, y) = \gamma \frac{c_{donator}(x, y)}{c_{donator}(x, y) + \eta(d) c_{receptor}(x, y)} \quad (8)$$

and will be referred to as the cross-pollination rate in the discussion.

In Equation (8), we also introduced the factor γ to be able to take into account the time overlap of male flowering phases of donor and recipient populations. In this study we considered γ to always take the value of one,

i.e. we assumed the flowering phases were at perfect synchrony. Due to the lack of more detailed information, the pollen source strength of all point sources was also considered to be equal throughout all plots, because the GM maize varieties were isogenic to the varieties in the surrounding conventional plot. Only differences in the densities between the plots would affect the results, because only relative not absolute pollen ratios were considered. Documentation of the flowering phases confirmed that no obvious differences between the plots arose in 2000. However, in-field heterogeneity was observed by botanical rating in 2001, both in the GM plot and recipient plot, but was not incorporated in the model, because the spatial resolution of the measurements was too low for interpolation.

Parameter estimations

Particle Lagrange model

The only parameters to be estimated for the cross-pollination prediction with the particle Lagrange (PL) model were the parameters of the pollen competition sub-model. The parameters pertaining to the transport model, such as the turbulence parameters, were all derived from the meteorological input data, and the roughness length by the use of a meteorological pre-processor sub-model implemented in the model (VDI, 2000). For estimation of the parameters of the competition model, the densities of donor, $c_{donor}(x, y)$, and recipient pollen, $c_{receptor}(x, y)$, were computed for each sampling point (x_k, y_k) and inserted into the competition sub-model.

The goal of the parameter estimations was to find the value of the parameter set $\theta \in \Theta$, where the predicted values of the competition sub-model $P_{out}(x_k, y_k, \theta)$ were overall closest to the observed values of cross-pollination rate, *i.e.* the ratio of number of seeds to total seeds of an ear that are herbicide tolerant, $HT_k/Corn_k$. Because the number of seeds sampled on the *k*th ear, $Corn_k$, is a covariate, a binomial model for the probability of observing *i* herbicide tolerant seeds, HT_k, on the *k*th ear is given by (Klein et al., 2003)

$$P_\theta[HT_k = i] = \\ C^i_{Corn_k} P_{out}(x_k, y_k, \theta)^i [1 - P_{out}(x_k, y_k, \theta)]^{Corn_k - i}, \quad (9)$$

where $C^i_{Corn_k}$ is the binomial coefficient.

Two statistical criteria were used to optimize the parameter set θ. The first criterion was the log likelihood

(Tufto et al., 1997)

$$L_k(\theta; HT_1, ..., HT_k) = L_K(\theta) = \log\left(\prod_{k=1}^{K} P_\theta[HT_k]\right)$$

$$= \sum_{k=1}^{K} \log(P_\theta[HT_k]). \quad (10)$$

Combination of Equations (9) and (10) leads to:

$$L_K(\theta) = \sum_{k=1}^{K} \log(C^{HT_k}_{Corn_k}) + HT_k \log[P_{out}(x_k, y_k, \theta)] \\ + (Corn_k - HT_k) \log[1 - P_{out}(x_k, y_k, \theta)]. \quad (11)$$

The second objective function is the extensively used least-squares (LS) criterion:

$$SS_K(\theta, n_1, ..., n_k) = SS_K(\theta) \\ = \sum_{k=1}^{K} \left[\frac{HT_k}{Corn_k} - P_{out}(x_k, y_k, \theta) \right]^2. \quad (12)$$

With nonlinear regression, we determined the value of θ where $L_K(\theta)$ is maximal, which is defined as the maximum likelihood estimator (MLE) $\hat{\theta}_{MLE}$ and the value of θ where $SS_K(\theta)$ is minimal, and will be referred to as the least sum of square estimator (LSSE) $\hat{\theta}_{LSSE}$. To calculate the results, only the MLE was used, because ML estimations are not biased towards high cross-pollination rates. The LSSE was presented for comparison of both estimation methods.

To compare both parameter estimation methods, we present a general statistical criterion for the overall goodness of fit with the model efficiency (ME), which is defined by

$$ME = 1 - \frac{\sum_{k=1}^{K} \left[\frac{HT_k}{Corn_k} - P_{out}(\theta, x_k, y_k) \right]^2}{\sum_{k=1}^{K} \left[\frac{HT_k}{Corn_k} - \overline{\left(\frac{HT_k}{Corn_k}\right)} \right]^2}. \quad (13)$$

Since the ME criterion considers all measured values linearly, the contribution to the ME is higher for large cross-pollination values.

Lagrange transfer function model

The parameters in the plume Equations (4) and (5) are in principal related to meteorological conditions. However, we used the dispersion parameters a_x, a_y and a_z as shape parameters for the plume. We used the maximum likelihood method to optimize the values of a_x, a_y and a_z

simultaneously with the competition sub-model parameters for the mean wind direction and wind velocity in 2000 and 2001, respectively (Tab. 3). All further simulations considering more than one wind direction, wind velocity combination were run with the same values for a_x, a_y and a_z, and numerical parameter optimization was only performed for the competition model parameters in the same way as for the particle Lagrange model.

ACKNOWLEDGEMENTS

REFERENCES

Angevin F, Klein E, Choimet C, Meynard J, de Rouw A, Sohbi Y (2001) Modélisation des effets des systèmes de culture et du climat sur les pollinisations croisées chez le maïs. INRAFNSEA, pp 21–36, http://www.fnsea.fr/dossiers/ogm/OGM020211e.pdf

Arritt R, Westgate M, Clark C, Fonseca A, Riese J (2003) Development of an adventitious pollen risk assessment model. In Boelt B, ed, 1st European Conference on the Coexistence of Genetically Modified Crops with Conventional and Organic Crops, Research Centre Flakkebjerg, pp 203–205

Aylor DE, Flesch TK (2001) Estimating spore release rates using a Lagrangian stochastic simulation model. *J. Appl. Meteorol.* **40**: 1196–1208

Aylor D, Schultes N, Shields E (2003) An aerobiological framework for assessing crosspollination in maize. *Agric. For. Meteorol.* **119**: 111–129

Bock A-K, Lheureux K, Libeau-Dulos M, Nilsagard H, Rodriguez-Cerezo E (2002) Scenarios for co-existence of genetically modified, conventional and organic crops in European agriculture. IPTS-JRC, ftp://ftp.jrc.es/pub/EURdoc/eur20394en.pdf

Devaux C, Lavigne C, Falentin-Guyomarc'h H, Vautrin S, Lecomte J, Klein E (2005) High diversity of oilseed rape pollen clouds over an agro-ecosystem indicates long-distance dispersal. *Mol. Ecol.* **14**: 2269–2280

Devos Y, Reheul D, De Schrijver A (2005) The co-existence between transgenic and nontransgenic maize in the European Union: a focus on pollen flow and cross-fertilization. *Environ. Biosafety Res.* **4**: 71–87

Eastham K, Sweet J (2002) Genetically modified organisms (GMOs): the significance of gene flow through pollen transfer. EEA, http://reports.eea.eu.int/environmental_issue_report_2002_28/en/GMOs%20for%20www.pdf

EU (2003) Regulation (EC) No 1830/2003 of the European Parliament and of the Council of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC. http://europa.eu.int/eurlex/pri/en/oj/dat/2003/1_268/1_26820031018en00240028.pdf

Henry C, Morgan D, Weekes R, Daniels R, Boffey C (2003) Farm scale evaluations of GM crops: monitoring gene flow from GM crops to non-GM equivalent crops in the vicinity: part I: forage maize. DEFRA, http://www.defra.gov.uk/environment/gm/research/pdf/epg_1-5-138.pdf

Ireland DS, Westgate ME, Ashton BA (2001) Combining ISCST3 and AERMOD particulate dispersion models to quantify maize pollen distribution. *ASACSSA-SSSA Annual Meetings*, Charlotte, NC, October 21–25

Jarosz N, Loubet B, Durand B, McCartney A, Foueillassar X, Huber L (2003) Field measurements of airborne concentration and deposition rate of maize pollen. *Agric. For. Meteorol.* **119**: 37–51

Jarosz N, Loubet B, Huber L (2004) Modelling airborne concentration and deposition rate of maize pollen. *Atmos. Environ.* **38**: 5555–5566

Jarosz N, Loubet B, Durand B, Foueillassar X, Huber L (2005) Variations in maize pollen emission and deposition in relation to microclimate. *Environ. Sci. Technol.* **39**: 4377–4384

Klein E, Lavigne C, Foueillassar X, Gouyon P-H, Larédo C (2003) Corn pollen dispersal: quasi-mechanistic models and field experiments. *Ecol. Monogr.* **73**: 131–150

Klein E, Lavigne C, Picault H, Renard M, Gouyon P-H (2006) Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. *J. Appl. Ecology* **43**: 141–151

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 a field. *Theor. Appl. Genet.* **96**: 886–896

Loos C, Seppelt R, Meier-Bethke S, Schiemann J, Richter O (2003) Spatially explicit modelling of transgenic maize pollen dispersal and cross-pollination. *J. Theor. Biol.* **225**: 241–255

Luna VS, Figueroa MJ, Baltazar MB, Gomez LR, Townsend R, Schoper JB (2001) Maize pollen longevity and distance isolation requirements for effective pollen control. *Crop Sci.* **41**: 1551–1557

Ma BL, Subedi KD, Reid LM (2004) Extent of cross-fertilization in maize by pollen from neighboring transgenic hybrids. *Crop Sci.* **44**: 1273–1282

Meier-Bethke S, Schiemann J (2003) Effect of varying distances and intervening maize fields on outcrossing rates of transgenic maize. In Boelt B, ed, 1st European Conference on the Coexistence of Genetically Modified

Crops with Conventional and Organic Crops, Research Center Flakkebjerg, pp 77–78

Nathan R, Perry G, Cronin JT, Strand AE, Cain ML (2003) Methods for estimating long distance dispersal. *Oikos* **103**: 261–273

Raynor GS, Ogden EC, Hayes JV (1972) Dispersion and deposition of corn pollen from experimental sources. *Agron. J.* **64**: 420–427

Seinfeld JH (1986) Atmospheric Chemistry and Physics of Air Pollution. *Wiley*, New York, ISBN: 0-471-82857-2, pp 880–949

Tolstrup K, Andersen S, Boelt B, Buus M, Gylling M, Holm P, Kjellson G, Pedersen S, Oestergard H, Mikkelsen SA (2003) Report from the working group on the co-existence of genetically modified crop with conventional and organic crops. DIAS, http://web.agrsci.dk/gmcc-03/Co_exist_rapport.pdf

Treu R, Emberlin J (2000) Pollen dispersal in the crops maize (*Zea mays*), oilseed rape (*Brassica napus* ssp. *oleifera*), potatoes (*Solanum tuberosum*), sugar beet (*Beta vulgaris* ssp. *vulgaris*) and wheat (*Triticum aestivum*). Soil Association, http://www.soilassociation.org/web/sa/saweb.nsf/librarytitles/GMO14012000/$file/Pollen%20Dispersal%20Report.pdf

Tufto J, Engen S, Hindar K (1997) Stochastic dispersal processes in plant populations. *Theor. Pop. Biol.* **52**: 16–26.

VDI (2000) Environmental meteorology, Atmospheric dispersion models: Particle model. VDI Richtlinien, VDI 3945, Beuth Verlag, Berlin

Weber W, Bringezu T, Broer I, Holz F, Eder J (2005) Koexistenz von gentechnisch verändertem und konventionellem Mais. *Mais* **1**: 1–6

Yamamura K (2004) Dispersal distance of corn pollen under fluctuating diffusion coefficient. *Popul. Ecol.* **46**: 87–101

To access this journal online:
www.edpsciences.org/ebr