



OPEN

Incorporating the field border effect to reduce the predicted uncertainty of pollen dispersal model in Asia

Yuan-Chih Su¹, Cheng-Bin Lee², Tien-Joung Yiu³ & Bo-Jein Kuo^{1,4,5}✉

The presence of the field border (FB), such as roadways or unplanted areas, between two fields is common in Asian farming system. This study evaluated the effect of the FB on the cross-pollination (CP) and predicted the CP rate in the field considering and not considering FB. Three experiments including 0, 6.75, and 7.5 m width of the FB respectively were conducted to investigate the effect of distance and the FB on the CP rate. The dispersal models combined kernel and observation model by calculating the parameter of observation model from the output of kernel. These models were employed to predict the CP rate at different distances. The Bayesian method was used to estimate parameters and provided a good prediction with uncertainty. The highest average CP rates in the field with and without FB were 74.29% and 36.12%, respectively. It was found that two dispersal models with the FB effect displayed a higher ability to predict average CP rates. The correlation coefficients between actual CP rates and CP rates predicted by the dispersal model combined zero-inflated Poisson observation model with compound exponential kernel and modified Cauchy kernel were 0.834 and 0.833, respectively. Furthermore, the predictive uncertainty was reducing using the dispersal models with the FB effect.

Maize (*Zea mays* L.) is one of the most important genetically modified (GM) crops globally, and the area of GM maize accounts for 32% of the total GM crop area¹. Although GM crops are crucial in the global crop production, the safety of GM products is still debated between producers and consumers, even though GM products have been on the market for two decades². The labeling threshold was established to protect non-GM crops and products admixed by the adventitious presence (AP) of GM content³.

The European Food Safety Authority indicated that the risks of GM plants in regard to persistence and invasiveness have to be assessed⁴. The outcrossing with wild or weedy relatives is a way to transfer the GM content to the environment^{5,6}. The hybridization of GM crops and wild relatives has been studied in the past^{7,8}. The study has indicated that the GM maize will hybridize with landrace⁹. The escape of GM content to the environment may reduce of diversity of flora and make weed control harder⁴. Therefore, the assessment of environment risk of GM crops should be concerned.

When conventional crop production cannot yield food sufficiently for the entire population, producers and consumers must seek solutions from other crop production systems. GM crops provide a solution for feeding the growing population¹⁰. Therefore, the coexistence between GM and non-GM crop production systems has been studied in past 2 decades. Coexistence denotes that farmers can choose between organic, non-GM, and GM crop production systems, but the GM contents of the products must meet the labeling standard¹¹. A practical strategy of coexistence would minimize the economic loss and cost.

Pollen dispersal is the main source of AP in the maize production. Not only the pollen dispersal of maize but also the other species, such as tree, had been studied¹². There are many factors (e.g. meteorology, topography) that can affect the pollen dispersal^{13–15}. The emission of pollen is affected by temperature and humidity¹⁶. The pollen emission period is longer when the temperature is low. The onset of pollen emission will be delayed when the humidity is high. The size of pollen influenced the pollen setting velocity, but the genetic variation among pollens

¹Department of Agronomy, National Chung Hsing University, Taichung 402202, Taiwan, ROC. ²Division of Biotechnology, Taiwan Agricultural Research Institute, Taichung 41362, Taiwan, ROC. ³Tainan District Agricultural Research and Extension Station, COA, Executive Yuan, Chiayi 613014, Taiwan, ROC. ⁴Innovation and Development Center of Sustainable Agriculture (IDCSA), National Chung Hsing University, Taichung 402202, Taiwan, ROC. ⁵Pervasive AI Research (PAIR) Labs, Hsinchu 30010, Taiwan, ROC. ✉email: bjkuo@dragon.nchu.edu.tw

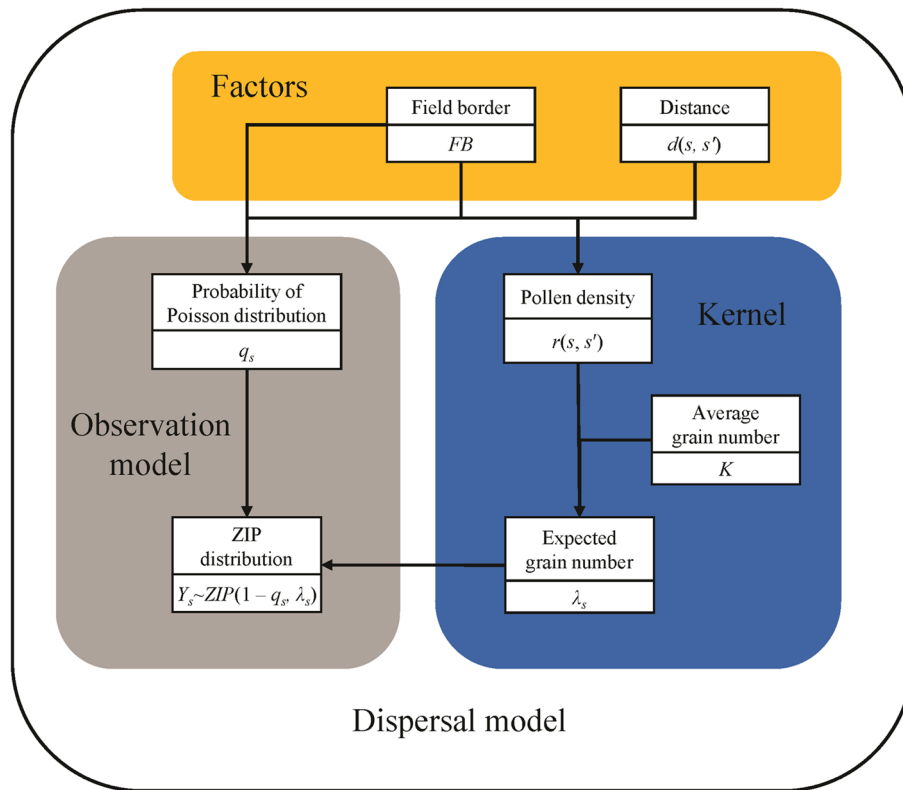


Figure 1. Graphical summary of the establishment of the dispersal model using ZIP distribution observation model as an example.

from the same plant and different plants was found¹⁷. The maize has a relative high setting velocity because of the pollen size. In general, the result of pollen dispersal varies widely with the environments and the plants.

Setting isolation distances to avoid the AP caused by the pollen dispersal is a common containment strategy in numerous countries. However, a fixed isolation distance is not suitable for different cultivation scenarios and may reduce the profit of production. In Asia, most fields are small and close to each other. The fields are usually separated by the field border (FB), such as roadways between two fields or unplanted areas. Under this kind of cropping system, farmers must pay a higher cost of coexistence than extensive farming systems. Studies have indicated that the main factors affecting the AP in a small field were the synchronicity of flowering and isolation distance¹⁸. The flowering time isolation is an alternative strategy to cope with the limitation of the field size¹⁹. However, the separation of flowering time may cause an unfavorable growing season²⁰. Accurate prediction of the AP can provide the useful information for the establishment of the isolation distances. Numerous pollen dispersal models including exponential, log/square, log/log, Gaussian plume, and two-step model were developed to predict the AP caused by the pollen dispersal^{21–23}.

The present study conducted three experiments to simulate different Asian cultivation scenarios. Different dispersal models were established to predict the CP rate. Specifically, the FB effect was included in the dispersal models to adapt to the Asian farming system. Additionally, the predicted uncertainty was evaluated using the Bayesian estimation method.

Methods

Dispersal models. In this study, the dispersal model consists of two parts, namely, kernel and observation model (Fig. 1). The main purpose of the kernel was employed to estimate the proportion of pollen dispersed from location s' to location s and calculate the expected number of CP grains. The observation model used the expected number of CP grains as a parameter and described the number of CP grains at location s (Y_s) by a specific distribution in the following:

$$Y_s \sim f(y_s | \theta_s), \quad (1)$$

where f indicates the probability density function (PDF) of the specific distribution. The θ_s is the parameter vector of the distribution. This study constructed eight different dispersal models combined with two observation models, two kernels, and two conditions of the field border (FB) effect (Table 1). The details of the kernels and observation models were described in the following subsections.

Observation model	Kernel	Field border effect	Model code
Poisson	Compound exponential	–	PExpoN
Poisson	Compound exponential	+	PExpoB
Poisson	Modified Cauchy	–	PCauchyN
Poisson	Modified Cauchy	+	PCauchyB
Zero-inflated poisson	Compound exponential	–	ZExpoN
Zero-inflated poisson	Compound exponential	+	ZExpoB
Zero-inflated poisson	Modified Cauchy	–	ZCauchyN
Zero-inflated poisson	Modified Cauchy	+	ZCauchyB

Table 1. List of dispersal models constructed in this study.

Kernels. The kernel indicates the probability when the pollen emitted at location s' and would fall down at location s . It can be expressed as $\gamma(s, s')$, where s' is the source location closest to location s . Numerous kernels have been used to describe various dispersal phenomena²⁴. The output of the kernel represents the donor pollen density of location s . In order to calculate the expected number of CP grains, the donor pollen density is multiplied by the average total grain number described as follows:

$$\lambda_s = K \times \gamma(s, s'), \quad (2)$$

where λ_s and K indicate the expected number of CP grains at location s and the average number of grains per cob, respectively. The effect of the FB was introduced into the kernel to suit to the small-scale farming system in Asia. This study assumed that the relation between the pollen density at the first recipient row and the width of the FB displayed an exponential decrease^{25,26}. To evaluate the improvement of the kernel with the FB effect, the kernels without the FB effect were also established in this study.

The compound exponential kernel (γ_{Expo}) has been used in the previous pollen dispersal study²⁷. Our study introduced the FB effect into this kernel. Therefore, the form of the compound exponential kernel can be expressed as follows:

$$\gamma_{\text{Expo}}(s, s') = \begin{cases} K_c \exp(-a_1 d^*(s, s')) \exp(-k\sqrt{FB}), & \text{if } d^*(s, s') \leq D \\ K_c \exp(-a_1 D - a_2(d^*(s, s') - D)) \exp(-k\sqrt{FB}), & \text{if } d^*(s, s') > D, \end{cases} \quad (3)$$

where K_c , a_1 , a_2 , k , D are the parameters of the kernel. $d^*(s, s')$ indicates the shortest distance between locations s' and s in which the width of the FB has been subtracted. In the compound exponential kernel without the FB effect, the exponential term of the FB effect was removed and the $d^*(s, s')$ was replaced directly by the shortest distance between s' and s .

The second kernel applied in this study was the modified Cauchy kernel (γ_{Cauchy}) which was based on the PDF of the Cauchy distribution and the concept of compound distribution. The modified Cauchy kernel is represented as follows:

$$\gamma_{\text{Cauchy}}(s, s') = \begin{cases} \frac{2\beta}{\pi[\beta^2 + d^{*2}(s, s')]^2} \exp(-k\sqrt{FB}), & \text{if } d^*(s, s') \leq D \\ \frac{2\beta}{\pi[\beta^2 + D^2 + c_1(d^*(s, s') - D)^2]} \exp(-k\sqrt{FB}), & \text{if } d^*(s, s') > D, \end{cases} \quad (4)$$

where the β indicates the decline rate of the curve. Parameters of k and D are same as the compound exponential kernel. c_1 indicates the relative slow decrease of pollen density at further distances. Similarly, in the modified Cauchy kernel without the FB effect, the term of the FB effect was removed and the $d^*(s, s')$ was replaced directly by the shortest distance between s' and s in which the row spacing (0.75 m) had been subtracted.

Observation models. Because of the high proportions of zero value observations, the present study assumed that the CP grain count followed the zero-inflated Poisson (ZIP) distribution to account for zero-excess condition²⁸. The ZIP distribution was first proposed by Lambert²⁹, and several studies had applied the ZIP distribution to deal with the CP data^{27,30}. The ZIP distribution consists of a Dirac distribution in zero and a Poisson distribution. Therefore, the distribution of CP grain count at location s (Y_s) can be expressed as follows:

$$Y_s \sim \text{ZIP}(1 - q_s, \lambda_s), \quad (5)$$

where q_s indicates the probability of an observation following a Poisson distribution, and λ_s is the parameter of Poisson distribution calculated by Eq. (2). Furthermore, the parameter q_s can be assumed to depend on the shortest distance between the recipient and donor plants. The border effect is also included in the estimation of q_s because it is related to the distance effect. The relationship among distance, border, and the q_s can be described using the following logistic function:

$$q_s = \frac{1}{1 + \exp(b_1 - b_2 d^*(s, s'))}, \quad (6)$$

where b_1 and b_2 are the parameters of the logistic function. The $d^*(s, s')$ was the shortest distance between s' and s in the version of dispersal models without the FB effect. The Poisson distribution was also used as an observation model for comparison with the ZIP observation model.

Experimental and meteorological data collection. The pollen dispersal data were collected from experiments performed in 2009 and 2010 at the geographic coordinates 23° 47' N, 120° 26' E, and an altitude of 20 m. These experiments were coded as 2009-1, 2009-2, and 2010-1, respectively. The experiment 2009-2 was divided into 2009-2A (without the FB) and 2009-2B (with the FB) based on the presence of the FB. The different layouts of the field experiments were designed to investigate the effect of the FB. Two commercial glutinous maize varieties, black pearl (purple grain) and Tainan No. 23 (white grain), were selected as the pollen donor and pollen recipient, respectively. The distance between the plants in a row was 25 cm, whereas the distance between the rows was 75 cm. The recipient plots consisted of 82 and 91 rows in 2009 and 2010 experiments, respectively.

The CP rate was determined based on the differences in grain color on recipient cobs as a result of the xenia effect³¹. In the sampling framework, the whole field was divided into many grids and corn samples were collected from each grid in the whole field. The CP rate of each grid was calculated using the method presented in a previous study³² and defined as:

$$\text{CP}(\%) = \left[\frac{\sum_{i=1}^n \text{Cob}_i}{(n \times K)} \right], \quad (7)$$

where Cob_i and n indicate i th cob and total number of cobs in the grid, respectively. K is the average grain number per cob. Meteorological data were collected from the meteorological station at geographic coordinates 23° 35' N, 120° 27' E, and an altitude of 20 m. The detailed experimental setup was described in our previous study³³. The study complies with relevant institutional, national, and international guidelines and legislation.

Statistical analyses. All statistical analyses were performed using SAS (Statistical Analysis System, version 9.4). The dispersal model parameters were estimated by two methods. First, the nonlinear model estimation was conducted by PROC NLMIXED to evaluate the fitting and predictive abilities of dispersal models. Then the dispersal models with the observation model performed better fitting ability were re-estimated using the Bayesian estimation method to assess the uncertainty by PROC MCMC. In the Bayesian method, the noninformative prior distribution was used to estimate all parameters (Supplementary Table S1). The iteration of Markov Chain was 500,000 times and the burn-in was set to 450,000 iterations. In order to reduce the autocorrelations in the chain, the thinned value was set to 25.

The validation method used in this study was the threefold cross-validation for the results of both estimation methods. The data from three experiments were combined and randomly partitioned into three sub-datasets. To avoid the heterogeneity of the different field designs and distances among sub-datasets, the observations from the same field design and same distance were considered as a group, and then partitioned into three parts. Each sub-dataset contained one part of all groups. At each validation run, two sub-datasets were selected as the training set, and the remaining one was used for validation.

The fitting ability of the dispersal models was evaluated based on two criteria, namely, Akaike information criterion (AIC), Deviance, and coefficient of determination (R^2). The smaller values of AIC or deviance indicate a better fitting. The higher R^2 value represents a better fitting performance. The correlation coefficient (r) between the predicted and actual CP rates was used to assess the predictive ability. The deviance information criterion (DIC) was used to evaluate the performance of dispersal model fitting for the Bayesian estimation. The criterion values calculated from three training and validation sets were averaged to assess the overall results. The uncertainty of the model parameter was quantified by the standard deviation (SD) of parameter posterior distribution. The 95% credible intervals of posterior predictive distribution constructed by the 2.5th and 97.5th percentiles of 200,000 samples generated from the posterior predictive distribution were used to assess the predictive uncertainty. Furthermore, to assess the zero-excess condition, the percentage of observed zero CP grain events was compared with the Poisson probability of the zero CP grain event. A zero-excess condition occurred if the observed percentage was higher than the Poisson probability³⁴.

Results

Cross-pollination. The highest CP rate of each experiment was observed in the first row of the recipient plot (Fig. 2). Compared with the average CP rate at the first row, the average CP rate declined by 50% in the second or third recipient row. In the 2009-1 experiment, the average CP rates at 0.75 and 1.5 m were 27.24% and 9.35%, respectively. The average CP rates of the first row in the 2009-2A and 2009-2B experiments were 74.29% (0.75 m) and 36.12% (6.75 m), respectively. The average CP rates decreased by 50% in the third row in the 2009-2A and 2009-2B experiments. The average CP rate in the first row (7.5 m) was 27.58% and dropped to 11.97% in the second row (8.25 m) in the 2010-1 experiment. In three experiments, the CP rate drastically decreased in the first few rows. However, the CP rate still could be observed at the farthest sampling points.

Field border effect and zero-excess. In the 2009-2A and 2009-2B experiments, the difference in the average CP rate at a distance of 6.75 m indicated that the CP rate was different between the field with and without

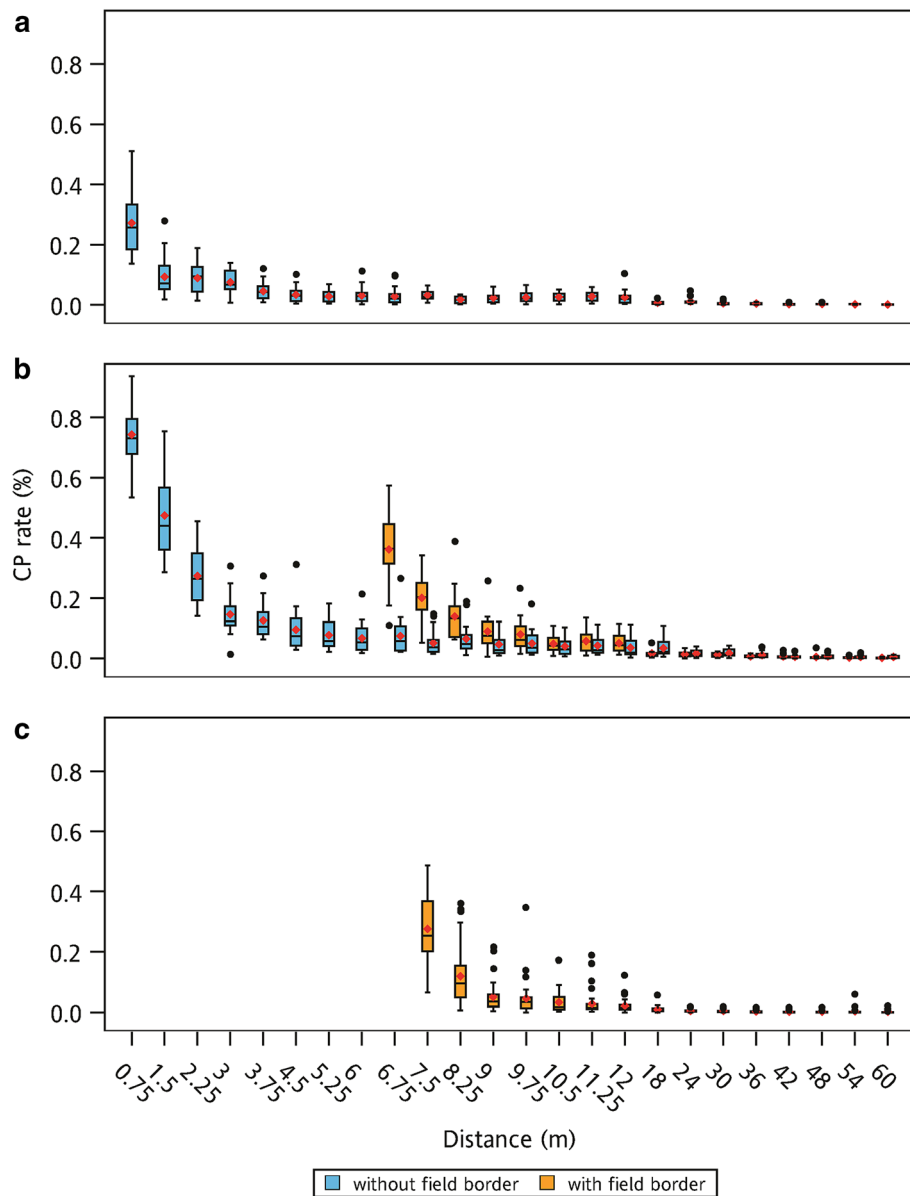


Figure 2. Box plots of cross-pollination (CP) rate of (a) 2009-1 experiment, (b) 2009-2 experiment, and (c) 2010-1 experiment. The red diamonds and black dots represent the mean and outliers, respectively.

FB (Fig. 1). The average CP rates of experiments 2009-2A and 2009-2B at 6.75 m were 7.43% and 36.12%, respectively. The field without FB had a larger reduction in the CP rate compared with the field with FB. The difference in the CP rate between experiments 2009-2A and 2009-2B was observed within 30 m.

In this study, the zero-excess situations at different distances were also examined. We evaluated the level of the zero-excess of each experiment by the proportion of zero-excess condition of whole field. The probability of zero CP of each distance was calculated and compared with the proportion of zero CP event of actual data. The zero-excess levels differed among three experiments (Supplementary Table S2). The zero CP grain event mainly occurred at the distance > 10 m. The 2010-1 experiment exhibited the highest zero-excess level, and the percentage of the zero-excess condition was 60%. However, the zero-excess levels in other experiments were relatively low (2009-1: 20.8%; 2009-2A: 12.5%; 2009-2B: 6.25%).

Fitting and predictive abilities of dispersal models. This study constructed eight dispersal models including different observation models, kernels, and with and without the FB effect. According to the AIC, deviance, and R^2 , the dispersal models with the FB effect had a better fitting ability than the dispersal models without the FB effect (Table 2). The fitting ability of modified Cauchy kernel without the FB effect was superior to that of the compound exponential kernel without the FB effect. However, the compound exponential kernel with the FB effect displayed a better fitting ability than the modified Cauchy kernel with the FB effect. The dispersal

Model code	AIC	Deviance	R ²	r
PExpoN	47,162 ± 669	35,014 ± 463	0.579 ± 0.025	0.76 ± 0.036
PExpoB	36,647 ± 784	24,497 ± 504	0.703 ± 0.006	0.833 ± 0.013
PCauchyN	44,227 ± 861	32,081 ± 737	0.582 ± 0.028	0.762 ± 0.033
PCauchyB	37,475 ± 1113	25,327 ± 833	0.696 ± 0.007	0.834 ± 0.009
ZExpoN	45,856 ± 639	33,207 ± 432	0.574 ± 0.025	0.76 ± 0.036
ZExpoB	35,780 ± 784	22,899 ± 548	0.701 ± 0.007	0.834 ± 0.012
ZCauchyN	43,608 ± 994	31,141 ± 1065	0.579 ± 0.022	0.763 ± 0.031
ZCauchyB	36,588 ± 1053	23,922 ± 880	0.691 ± 0.011	0.833 ± 0.01

Table 2. The average mean and standard deviation of Akaike information criterion (AIC), deviance, coefficient of determination (R²), and correlation coefficient (r) of dispersal models obtained using threefold cross-validation.

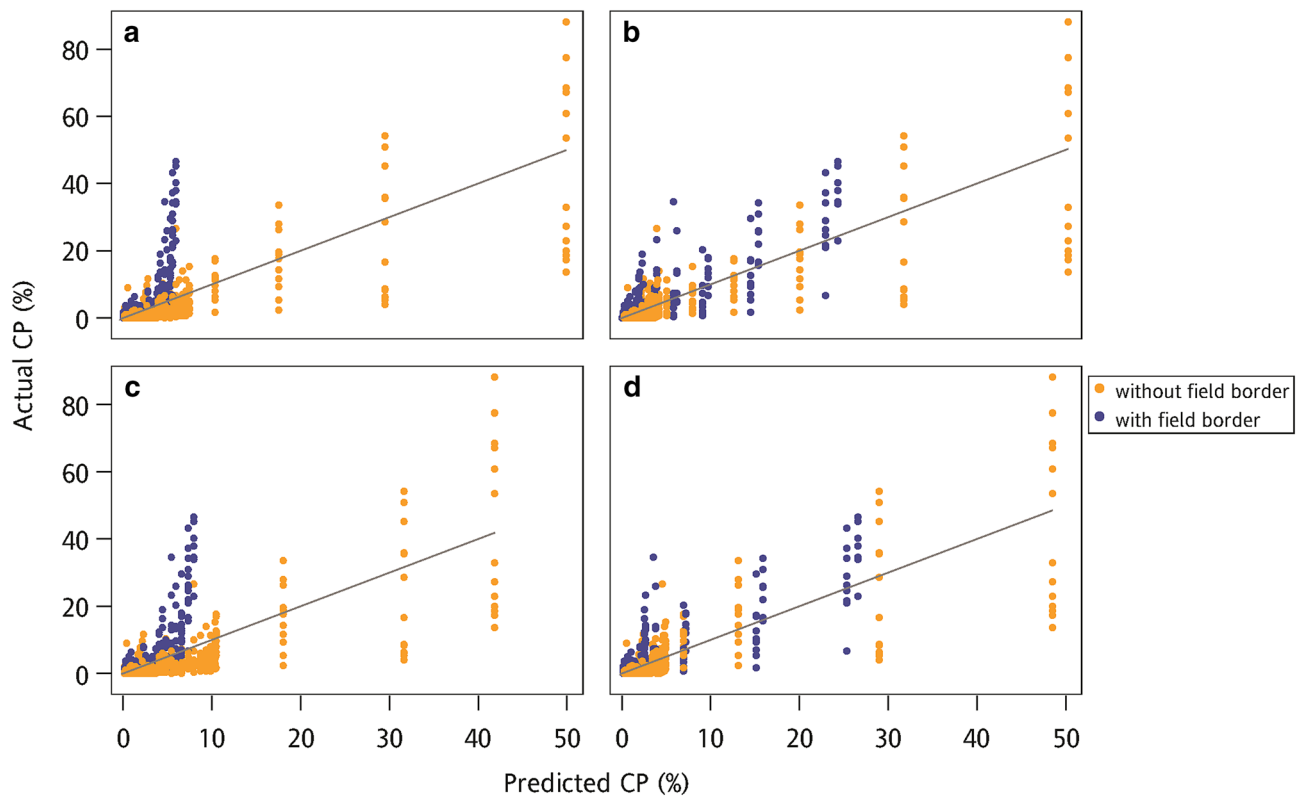


Figure 3. Scatter plots of the actual cross-pollination (CP) rate from the validation data and predicted CP rate using the (a) ZExpoN, (b) ZExpoB, (c) ZCauchyN, and (d) ZCauchyB models to fit the training data. The gray line is the 1:1 reference line.

models with ZIP observation model performed a better fitting ability than the dispersal models with Poisson observation model. Overall, the ZExpoB model showed the best fitting ability and the ZCauchyB was the second. Additionally, the results of the predictive ability and fitting ability among the dispersal models were similar. However, the difference of the predictive ability of dispersal models with Poisson and ZIP observation models was negligible. The PCauchyB and ZExpoB models performed the best predictive ability and both models had a same correlation coefficient of 0.834.

According to the fitting result of dispersal models, the dispersal models with the ZIP observation model were the preferred models for describing the CP data. Therefore, the dispersal models with the ZIP observation model were selected to investigate the difference of predicted CP rate between the dispersal models with and without the FB effect. The scatter plot of the actual CP rate from the validation data versus the predicted CP rate calculated using four dispersal models fitted with the training data was represented in Fig. 3. It was found that the dispersal model without the FB effect seriously underestimated some CP rates from the field with FB (Fig. 3). In contrast, the predicted CP rate from the dispersal model with the FB effect distributed around the reference line evenly. The result indicated that the dispersal models with the FB effect had a great predictive ability under the conditions of the field with and without FB.

Parameter	Model code			
	ZExpoN		ZExpoB	
	Mean	SD	Mean	SD
K_c	0.8318	0.0174	0.6760	0.0091
a_1	0.6823	0.0131	0.6073	0.0065
a_2	0.0781	0.0042	0.0506	0.0007
k	n/a	n/a	0.3552	0.0060
D	3.3954	0.0488	2.8480	0.0437
b_1	2.5743	0.1226	2.7621	0.1103
b_2	0.0086	0.0039	0.0275	0.0034

Table 3. The average mean and standard deviation (SD) of parameter posterior distribution of ZExpoN and ZExpoB models obtained using threefold cross-validation.

Parameter	Model code			
	ZCauchyN		ZCauchyB	
	Mean	SD	Mean	SD
β	1.3608	0.0196	1.1000	0.0122
c_1	0.1467	0.0028	0.2832	0.0038
k	n/a	n/a	0.0632	0.0016
D	2.4030	0.0257	3.5803	0.0401
b_1	3.0979	0.1448	2.7795	0.1102
b_2	0.0927	0.0044	0.0266	0.0034

Table 4. The average mean and standard deviation (SD) of parameter posterior distribution of ZCauchyN and ZCauchyB models obtained using threefold cross-validation.

Predicted uncertainty of dispersal models. The parameters of dispersal models with the ZIP observation model were re-estimated by the Bayesian estimation method. According to the DIC, the ZExpo model appeared the best fitting ability, with the ZCauchy model following (Supplementary Table S3). The comparison result of fitting ability between dispersal models estimated by Bayesian method was similar to the result by non-linear estimation.

For the parameter estimation of the Bayesian method, the average mean values of the parameter posterior distribution between ZExpoN and ZExpoB models were comparable (Table 3). However, the ZExpoB model had a small SD of the parameter posterior distribution than the ZExpoN model. This result indicated the posterior distribution of parameters between these two models were similar but the parameter uncertainty of ZExpoB model was smaller than that of ZExpoN model. The ZCauchyB model also performed a small parameter uncertainty, but only a half of the SD of the parameter posterior distribution was reduced after adding the FB effect (Table 4). The predicted uncertainty was assessed under the field conditions with and without FB and quantified by the 95% credible interval of posterior predictive distribution of the CP rate. The posterior predictive distribution of the CP rate was generated by the dispersal models fitted with the training data. Under the field condition without FB, dispersal models with the FB effect illustrated a relatively narrow 95% credible interval, especially at the distance between 3 and 10 m (Fig. 4). The narrow 95% credible intervals indicated a smaller predicted uncertainty. The predicted CP rates calculated by four dispersal models were closed to the observed CP rates of the data points without FB from the validation data. For the field condition with a 7.5 m width FB, ZExpoB and ZCauchyB models performed a smaller predicted uncertainty at the distance between 7.5 m and 18 m than ZExpoN and ZCauchyN model (Fig. 5). The predicted CP rates were compared to the observed CP rates from the validation data that only included the data points with the 7.5 m FB. In addition, the dispersal models without the FB effect underestimated the CP rates at distances of 7.5 m and 8.25 m. These results showed that the dispersal models with the FB effect could reduce the predicted uncertainty at the field conditions with and without FB.

Discussion

This study conducted three experiments with different field designs to investigate the effect of FB on CP results. The main difference between the field designs was with or without the setting of the FB. A study by Della Porta et al.²¹ indicated that maize plants were a more efficient barrier against the CP than the unplanted area. Because the pollen was obstructed by maize plants, the CP rate at the first row in the plot with the FB was lower than the CP rate at the first row in the plot without the FB. Under the 0.9% threshold level, fields with the unplanted area (25.9 m) required a longer isolation distance than fields with the planted area (17.5 m), because the intervening plants between source and recipient fields compete with their own pollens. This difference between the planted and unplanted areas was also observed in a study on oilseed rape³⁵. Our findings also revealed a similar

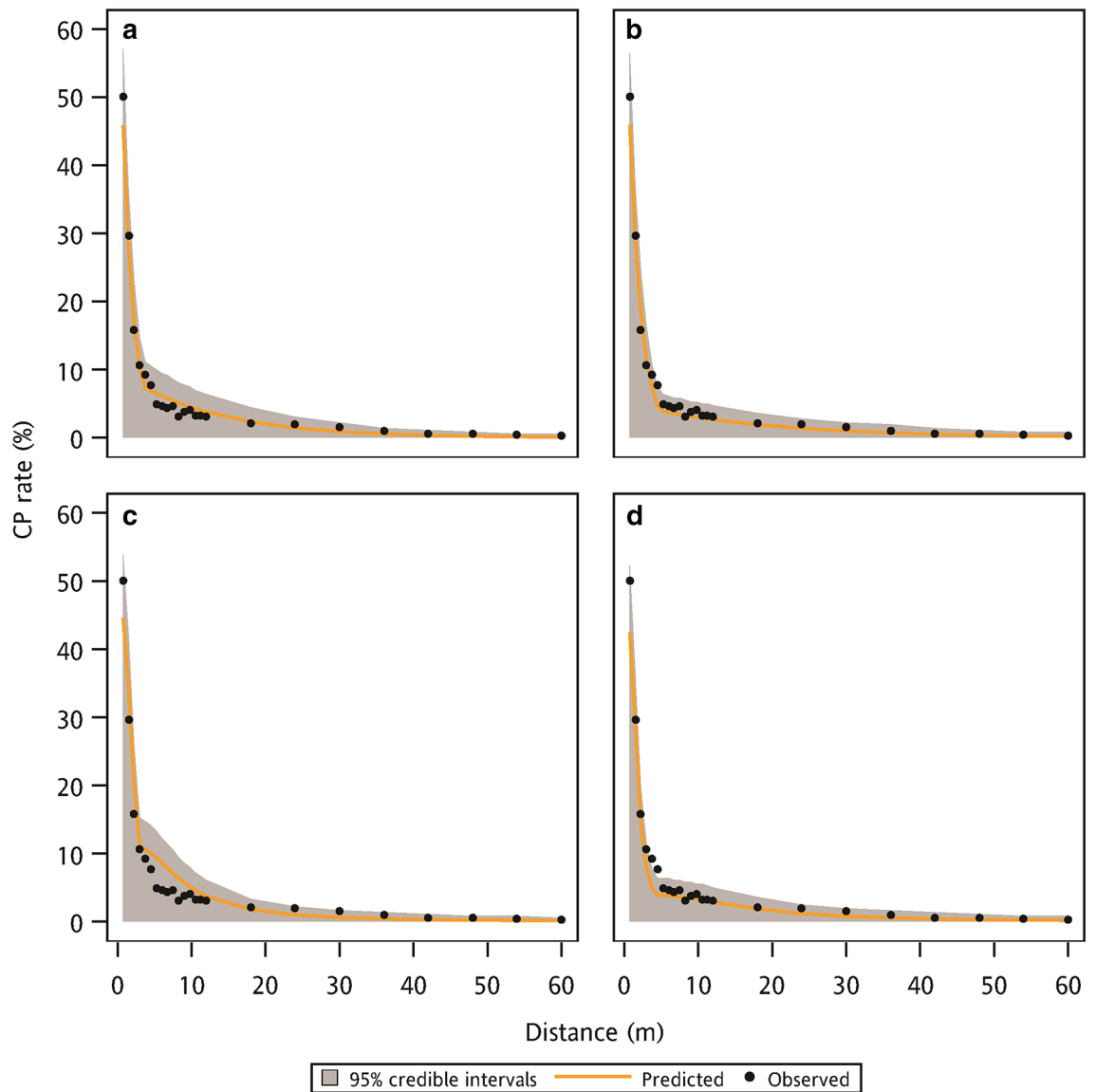


Figure 4. The observed CP rate from validation data (without FB condition) and the predicted CP rate with 95% credible intervals using the (a) ZExpoN, (b) ZExpoB, (c) ZCauchyN, and (d) ZCauchyB models to fit the training data.

declining trend in CP rates in fields with and without FB. Furthermore, the result of the oilseed rape study also supported the hypothesis that the border and buffer zone should be treated as different factors in small-scale field condition studies³⁵. Moreover, they found if the dispersal model without the FB effect, the CP rate was underestimated. This is consistent with the result of our study. The CP results of the pollen dispersal experiment and the predictive ability of the fitted model usually depend on the arrangement of the source and recipient fields^{23,36}. The self-protection index proposed by Melé et al.³⁷ demonstrated that recipient fields with a large area and a small perimeter displayed a higher CP protection. Because of the self-protection ability, the CP results in a large-scale experiment are not affected by the presence of the FB effect. However, the CP results of our study showed a notable FB effect. Therefore, the fields with and without the FB effect would result in different CP trends in the small-scale experiment.

When the number of zero in the data is more than the expected number of zero in Poisson distribution, the problem of zero-excess occurs. The zero-excess characteristic of the dispersal data has been investigated in some studies^{38,39}. Because the recipient plots were set downwind of the source plots, the excess level of the zero value differed in the experiments in our study. This arrangement of the recipient plot can explain the low zero-excess level in some of our experiments. The obvious fluctuation of the level of zero-excess among experiments probably due to the high variation of CP grain number at the long distance. According to the result of Bensadoun et al.²⁷, the CP grain number variation was higher at a long distance in the downwind direction plots. The higher level of zero-excess of experiment 2010-1 might result from the sampling bias.

The dispersal models were constructed by combining different observation models and kernels in this study. The zero-inflated models, such as the ZIP mode and zero-inflated negative binomial model, were usually used

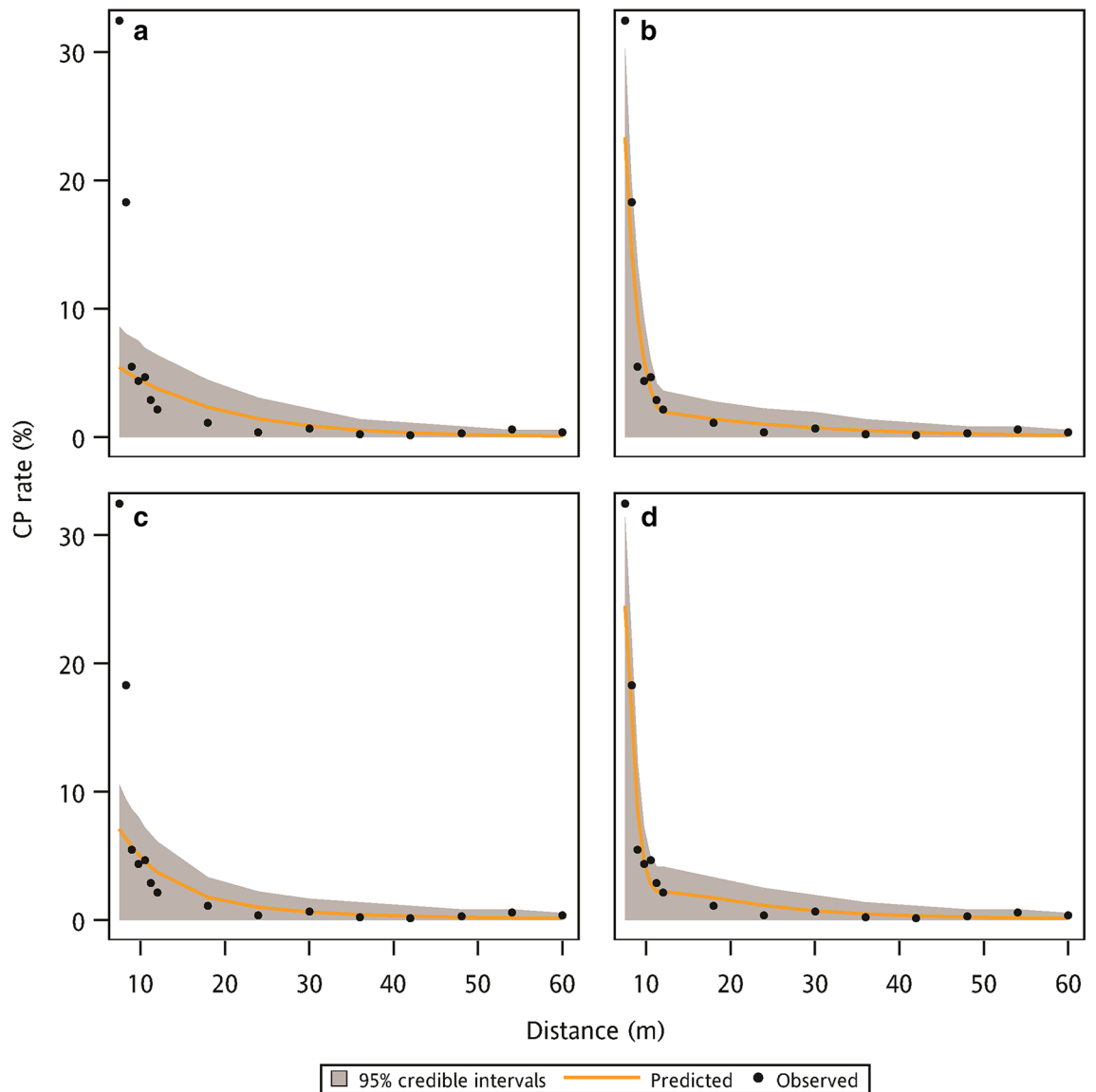


Figure 5. The observed CP rate from validation data (with 7.5 m FB condition) and the predicted CP rate with 95% credible intervals using the (a) ZExpoN, (b) ZExpoB, (c) ZCauchyN, and (d) ZCauchyB models to fit the training data.

to describe the zero-excess count data^{27,30}. Bensadoun et al.²⁷ reported that the cross validation result of the dispersal model with the ZIP observation model ($R^2 = 0.582$) was better than that of the Poisson observation model ($R^2 = 0.511$). However, in our study, the dispersal models with the Poisson and ZIP observation models had a negligible difference in the R^2 values. For the models of PExpoN and ZExpoN, R^2 values were close to the 0.582. It was concluded that this negligible difference resulted from the low zero-excess level in some experimental data.

Numerous fat-tailed kernels have been used in dispersal studies, such as Cauchy⁴⁰, log-sech³⁰, and $2Dt$ ⁴¹. Paradis et al.⁴² reported that the Cauchy kernel predicted long-distance dispersal events more frequently than the exponential model. Studies have used only the scale parameter in the Cauchy kernel to describe the dispersal result^{42,43}. The present study modified the Cauchy kernel and retained the location parameter of Cauchy kernel to introduce the FB effect. The FB effect enabled the modified Cauchy kernel to be more suitable to fit the data for the field designs with and without FB. In this study, the FB effect was also introduced into the compound exponential kernel to compare with the modified Cauchy kernel. When incorporating the FB effect into the dispersal models, the predictive ability was increased. The dispersal models with the modified Cauchy kernel ($r = 0.833$) and the compound exponential kernel ($r = 0.834$) displayed favorable results in the CP rate prediction. As described in our previous study, adding the FB effect into the models also improved the predictive ability of models³³. In another study, the highest R^2 of the fitting result of log/log model was 0.91, when the field included a 17 m empty space between the source and recipient²¹. For a farm scale study, R^2 of the two-step model was 0.997²³. Although aforementioned models had a great fitting ability, the predictive ability should be further validated by a suitable validation procedure. More importantly, those models might fail to predict the CP rate because of only considering the distance effect.

The uncertainty was rarely mentioned in previous pollen dispersal studies. The study indicated that when the wind effect was considered in the dispersal model, the predicted uncertainty was decreased⁴⁴. Similarly, the FB effect was incorporated into the dispersal models to reduce the predicted uncertainty in our study. Therefore, this finding indicates that adding some effects which could improve the ability of dispersal model to adapt to various scenarios also might diminish the predicted uncertainty.

Conclusion

The setting of isolation distance has relative high cost in small-scale farming system despite that the isolation distance is the main strategy of the coexistence. Small-scale farming systems with the FB and the planted area display notable differences in CP rates. In fact, FBs can be regarded as a short isolation distance. Therefore, the combination of plant date and FB may be a measure instead of the isolation distance in small-scale farming system. Because of the difference of CP rates between the FB and planted area, predictions of CP rates should account for the FB effect and make a distinction between the FB and planted area in small-scale fields. Incorporating FB effect into the dispersal models improved both the fitting ability and the predictive ability. In addition, the dispersal models with FB effect predicted the CP rates with a relatively small uncertainty. The introducing of Bayesian estimation can not only evaluate the predictive uncertainty, but also provide a prediction with a probabilistic support.

Data availability

The data presented in this study are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.13370756.v1>.

Received: 4 July 2021; Accepted: 1 November 2021

Published online: 12 November 2021

References

1. ISAAA. *ISAAA Brief 55-2019: Executive Summary Biotech Crops Drive Socio-Economic Development and Sustainable Environment in the New Frontier*. (ISAAA, 2019).
2. Pham, N. & Mandel, N. What influences consumer evaluation of genetically modified foods?. *J. Public Policy Mark.* **38**, 263–279 (2019).
3. Ramesar, K., Capell, T., Twyman, R. M., Quemada, H. & Christou, P. Trace and traceability—A call for regulatory harmony. *Nat. Biotechnol.* **26**, 975–977 (2008).
4. Guidance on the environmental risk assessment of genetically modified plants. *EFSA J.* **8**, 1879 (2010).
5. Snow, A. A. Transgenic crops—Why gene flow matters. *Nat. Biotechnol.* **20**, 542 (2002).
6. Warwick, S. I. *et al.* Hybridization between transgenic *Brassica napus* L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) O.E. Schulz. *Theor. Appl. Genet.* <https://doi.org/10.1007/s00122-003-1278-0> (2003).
7. Stewart, C. N., Halfhill, M. D. & Warwick, S. I. Transgene introgression from genetically modified crops to their wild relatives. *Nat. Rev. Genet.* **4**, 806–817 (2003).
8. Raybould, A. F. & Gray, A. J. Genetically modified crops and hybridization with wild relatives: A UK Perspective. *J. Appl. Ecol.* **30**, 199–219 (1993).
9. Quist, D. & Chapela, I. H. Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature* **414**, 541–543 (2001).
10. Zilberman, D., Holland, T. G. & Trilnick, I. Agricultural GMOs—What we know and where scientists disagree. *Sustain.* **10**, 1–19 (2018).
11. Jank, B., Rath, J. & Gaugitsch, H. Co-existence of agricultural production systems. *Trends Biotechnol.* **24**, 198–200 (2006).
12. Adams-Groom, B., Skj oth, C. A., Baker, M. & Welch, T. E. Modelled and observed surface soil pollen deposition distance curves for isolated trees of *Carpinus betulus*, *Cedrus atlantica*, *Juglans nigra* and *Platanus acerifolia*. *Aerobiologia (Bologna)* **33**, 407–416 (2017).
13. Jarosz, N., Loubet, B., Durand, B., Foueillassar, X. & Huber, L. Variations in maize pollen emission and deposition in relation to microclimate. *Environ. Sci. Technol.* **39**, 4377–4384 (2005).
14. Peel, R. G., Kennedy, R., Smith, M. & Hertel, O. Do urban canyons influence street level grass pollen concentrations?. *Int. J. Biometeorol.* **58**, 1317–1325 (2014).
15. Rojo, J. *et al.* Modeling olive pollen intensity in the Mediterranean region through analysis of emission sources. *Sci. Total Environ.* **551–552**, 73–82 (2016).
16. Marceau, A. *et al.* Modelling diurnal and seasonal patterns of maize pollen emission in relation to meteorological factors. *Agric. For. Meteorol.* **151**, 11–21 (2011).
17. Di-Giovanni, F., Kevan, P. G. & Nasr, M. E. The variability in settling velocities of some pollen and spores. *Grana* **34**, 39–44 (1995).
18. Messeguer, J. *et al.* Pollen-mediated gene flow in maize in real situations of coexistence. *Plant Biotechnol. J.* **4**, 633–645 (2006).
19. Nieh, S. C., Lin, W. S., Hsu, Y. H., Shieh, G. J. & Kuo, B. J. The effect of flowering time and distance between pollen source and recipient on maize. *GM Crops Food* **5**, 287–295 (2014).
20. Devos, Y. *et al.* Coexistence of genetically modified and non-GM crops in the European Union: A review. *Sustain. Agric.* **29**, 203–228 (2009).
21. DellaPorta, G. *et al.* Maize pollen mediated gene flow in the Po valley (Italy): Source-recipient distance and effect of flowering time. *Eur. J. Agron.* **28**, 255–265 (2008).
22. Loos, C., Seppelt, R., Meier-Bethke, S., Schiemann, J. & Richter, O. Spatially explicit modelling of transgenic maize pollen dispersal and cross-pollination. *J. Theor. Biol.* **225**, 241–255 (2003).
23. Weekes, R. *et al.* A study of crop-to-crop gene flow using farm scale sites of fodder maize (*Zea mays* L.) in the UK. *Transgenic Res.* **16**, 203–211 (2007).
24. Nathan, R., Klein, E. K., Robledo-Arnuncio, J. J. & Revilla, E. Dispersal kernels: Review. In *Dispersal Ecology and Evolution* (eds Clobert, J., Baguette, M., Benton, T. G. & Bullock, J. M.) (Oxford University Press, 2012).
25. Damgaard, C. & Kjellsson, G. Gene flow of oilseed rape (*Brassica napus*) according to isolation distance and buffer zone. *Agric. Ecosyst. Environ.* **108**, 291–301 (2005).
26. Dolphin, C. J., Mosher, G. A., Ambrose, R. P. K. & Ryan, S. J. Meeting the tolerance: How successful is coexistence in commodity corn handling systems. *Am. Soc. Agric. Biol. Eng.* **36**, 777–784 (2020).

27. Bensadoun, A., Monod, H., Makowski, D. & Messéan, A. A Bayesian approach to model dispersal for decision support. *Environ. Model. Softw.* **78**, 179–190 (2016).
28. Goedhart, P. W., van der Voet, H., Baldacchino, F. & Arpaia, S. A statistical simulation model for field testing of non-target organisms in environmental risk assessment of genetically modified plants. *Ecol. Evol.* <https://doi.org/10.1002/ece3.1019> (2014).
29. Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14 (1992).
30. Su, Y. C. *et al.* Using a zero-inflated model to assess gene flow risk and coexistence of *Brassica napus* L. and *Brassica rapa* L. on a field scale in Taiwan. *Bot. Stud.* **61**, 1–11 (2020).
31. Bulant, C. & Gallais, A. Xenia effects in maize with normal endosperm: I. Importance and stability. *Crop Sci.* **38**, 1517–1525 (1998).
32. Bannert, M. & Stamp, P. Cross-pollination of maize at long distance. *Eur. J. Agron.* **27**, 44–51 (2007).
33. Kuo, B. J., Jhong, Y. S., Yiu, T. J., Su, Y. & Lin, W. Bootstrap simulations for evaluating the model estimation of the extent of cross-pollination in maize at the field-scale level. *PLoS One* **16**, e0249700 (2021).
34. Yang, Z., Hardin, J. W. & Addy, C. L. Testing overdispersion in the zero-inflated Poisson model. *J. Stat. Plan. Inference* **139**, 3340–3353 (2009).
35. Reboud, X. Effect of a gap on gene flow between otherwise adjacent transgenic *Brassica napus* crops. *Theor. Appl. Genet.* **106**, 1048–1058 (2003).
36. Beckie, H. J. & Hall, L. M. Simple to complex: Modelling crop pollen-mediated gene flow. *Plant Sci.* **175**, 615–628 (2008).
37. Melé, E. *et al.* Modeling gene flow distribution within conventional fields and development of a simplified sampling method to quantify adventitious GM contents in maize. *Sci. Rep.* **5**, 1–13 (2015).
38. Bontrager, M. & Angert, A. L. Gene flow improves fitness at a range edge under climate change. *Evol. Lett.* **3**, 55–68 (2019).
39. Stehlik, I., Caspersen, J. P. & Barrett, S. C. H. Spatial ecology of mating success in a sexually polymorphic plant. *Proc. R. Soc. B Biol. Sci.* **273**, 387–394 (2006).
40. VanHoutan, K. S., Pimm, S. L., Halley, J. M., Bierregaard, R. O. & Lovejoy, T. E. Dispersal of Amazonian birds in continuous and fragmented forest. *Ecol. Lett.* **10**, 219–229 (2007).
41. Klein, E. K., Lavigne, C., Picault, H., Renard, M. & Gouyon, P. H. Pollen dispersal of oilseed rape: Estimation of the dispersal function and effects of field dimension. *J. Appl. Ecol.* **43**, 141–151 (2006).
42. Paradis, E., Baillie, S. R. & Sutherland, W. J. Modeling large-scale dispersal distances. *Ecol. Model.* **151**, 279–292 (2002).
43. Krkošek, M., Lauzon-Guay, J. S. & Lewis, M. A. Relating dispersal and range expansion of California sea otters. *Theor. Popul. Biol.* **71**, 401–407 (2007).
44. Bensadoun, A., Monod, H., Angevin, F., Makowski, D. & Messéan, A. Modeling of gene flow by a Bayesian approach: A new perspective for decision support. *J. Agrobiotechnol. Manag. Econ.* **17**, 213–220 (2014).

Acknowledgements

This research was funded by Ministry of Science and Technology under Grant Number 108-2313-B-005-036. The Ministry of Science and Technology provided funding for the publication fee in the form of a grant awarded to B.J.K., through Pervasive AI Research (PAIR) Labs, Taiwan (108-2634-F-005-003), and the Innovation and Development Center of Sustainable Agriculture, from The Featured Areas Research Center Program, within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. The manuscript was edited by Wallace Academic Editing.

Author contributions

B.J.K. contributed to conception, review and editing the draft of manuscript. C.B.L. contributed to conception and performed the formal analysis. T.J.Y. managed the experiment resource and methodology. Y.C.S. performed the formal analysis and wrote the draft of manuscript. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01583-x>.

Correspondence and requests for materials should be addressed to B.-J.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021