



HAL
open science

Modeling of gene ow by a Bayesian approach: A new perspective for decision support

Arnaud Bensadoun, Herve Monod, Frédérique Angevin, David Makowski,
Antoine Messean

► To cite this version:

Arnaud Bensadoun, Herve Monod, Frédérique Angevin, David Makowski, Antoine Messean. Modeling of gene ow by a Bayesian approach: A new perspective for decision support. GMCC 2013, Nov 2013, Lisbon, Portugal. pp.28. hal-01019193

HAL Id: hal-01019193

<https://hal.science/hal-01019193>

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling of gene flow by a Bayesian approach: A new perspective for decision support

Arnaud Bensadoun

Joint work with:

Hervé Monod, Frédérique Angevin, David Makowski and Antoine Messéan

GMCC 2013 - Lisbon
12-15 November 2013



Outline

- 1 Introduction
 - Context
 - Objectives
- 2 Materials and Methods
 - Dispersal model
 - Observation model
 - Types of data
- 3 Results
 - Parameter estimation
 - Overall variability
 - Goodness of fit
 - Uncertainty quantification
- 4 Discussion and Perspectives

Outline

- 1 Introduction
 - Context
 - Objectives
- 2 Materials and Methods
 - Dispersal model
 - Observation model
 - Types of data
- 3 Results
 - Parameter estimation
 - Overall variability
 - Goodness of fit
 - Uncertainty quantification
- 4 Discussion and Perspectives

Context and Challenges

EU Commission provides recommendations for coexistence measures:

- To base decisions on scientific results.
- To take specific measures depending on the species and varieties and regional context.

How to set up coexistence measures in practice?

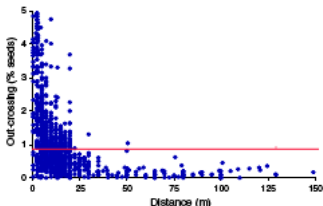
Context and Challenges

EU Commission provides recommendations for coexistence measures:

- To base decisions on scientific results.
- To take specific measures depending on the species and varieties and regional context.

How to set up coexistence measures in practice?

- Statistical approach.



⇒ Easy to implement from a regulatory point of view.

⇒ Difficult adaptation to local contexts → Not proportional to the risk

Figure: Cross pollination rate as a function of distance. *Source:* Riesgo et al., 2010

Context and Challenges

EU Commission provides recommendations for coexistence measures:

- To base decisions on scientific results.
- To take specific measures depending on the species and varieties and regional context.

How to set up coexistence measures in practice?

- Statistical approach.
- **Modelling approach** considering the effects of landscape characteristics, climate and agricultural practices to set up coexistence measures adapted to local climatic and agronomic factors.

⇒ Proportionality principle.

⇒ More difficult to implement under a regulatory framework.

Context and Challenges

EU Commission provides recommendations for coexistence measures:

- To base decisions on scientific results.
- To take specific measures depending on the species and varieties and regional context.

How to set up coexistence measures in practice?

- Statistical approach.
- **Modelling approach** considering the effects of landscape characteristics, climate and agricultural practices to set up coexistence measures adapted to local climatic and agronomic factors.

⇒ Proportionality principle.

⇒ More difficult to implement under a regulatory framework.

Challenges

⇒ Adapt coexistence rules to the diversity of possible situations.

⇒ Incorporate variability and uncertainty in decision making.

Knowledge and Objectives

- Knowledge about the risks of mixing and in particular on gene flow.
⇒ Databases gathering results of field experiments (e.g. SIGMEA).
- Generic models of pollen dispersal and Cross Pollination Rate (CPR) at the landscape scale, for instances:
 - ⇒ MAPOD (Angevin et al., 2008) for maize.
 - ⇒ GeneSys (Colbach et al., 2001) for rapeseed.

Knowledge and Objectives

- Knowledge about the risks of mixing and in particular on gene flow.
⇒ Databases gathering results of field experiments (e.g. SIGMEA).
- Generic models of pollen dispersal and Cross Pollination Rate (CPR) at the landscape scale, for instances:
⇒ MAPOD (Angevin et al., 2008) for maize.
⇒ GeneSys (Colbach et al., 2001) for rapeseed.

Objectives of the study

- Integrate variability and uncertainty in decision making.
 - Probabilistic prediction rather than single value:
⇒ Associate confidence interval to any CPR prediction.
 - Quantify prediction uncertainty.
- Enable adaptability of models to information level.
 - Development of simplified and flexible models.
 - Quantify the benefit of adding input data.



Relevance of the Bayesian approach

Outline

- 1 Introduction
 - Context
 - Objectives
- 2 **Materials and Methods**
 - Dispersal model
 - Observation model
 - Types of data
- 3 Results
 - Parameter estimation
 - Overall variability
 - Goodness of fit
 - Uncertainty quantification
- 4 Discussion and Perspectives

Individual Dispersal Functions

Individual dispersal function: $\gamma(d_s)$

A probability density function describing the probability that a pollen grain fertilises an ovule at a distance d_s from its emission point.

Individual Dispersal Functions

$$\gamma(d_s) = \begin{cases} K \times e^{-a_1 d_s} & d_s < D \\ K \times e^{-a_1 D - a_2 d_s} & d_s \geq D \end{cases} \quad (1)$$

(Source: Damgaard et al., 2005)

Where:

- $d_s = \sqrt{x^2 + y^2}$

⇒ **Distance model.**

Individual Dispersal Functions

$$\gamma(d_s) = \begin{cases} K \times e^{-a_1 d_s} & d_s < D \\ K \times e^{-a_1 D - a_2 d_s} & d_s \geq D \end{cases} \quad (1)$$

(Source: Damgaard et al., 2005)

Where:

- $d_s = \sqrt{x^2 + y^2}$

⇒ **Distance model.**

- $d_s = \sqrt{x^2 + y^2} \times (1 - \theta_v \cos(\omega - \omega_0))$,

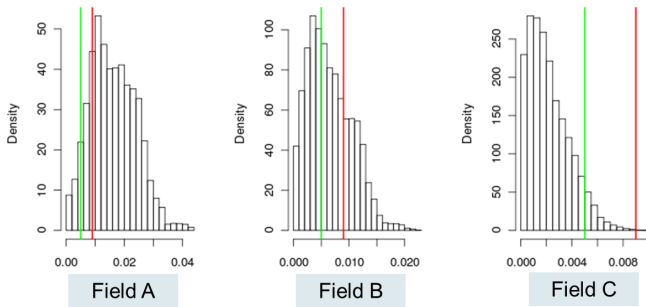
ω_0 the main wind direction,

and ω the angle between the vector donor/receptor and the horizontal axis.

⇒ **Distance+Wind model.**

Bayesian Framework

Model parameters are random variables, so is the prediction:



⇒ Requires the definition of a probabilistic model linking the observations to the model parameters.

Observation model

Natural observation model for counts of rare events

⇒ Poisson model:

$$Y_s \sim \mathcal{P}(K\mu'_s) \quad (2)$$

Observation model

Natural observation model for counts of rare events

⇒ Poisson model:

$$Y_s \sim \mathcal{P}(K\mu'_s) \quad (2)$$

But highly variable observations AND over-representation of zeros

⇒ Zero-Inflated Poisson model:

$$Y_s \sim ZIP(1 - q_s, K\mu'_s) \quad (3)$$

Observation model

Natural observation model for counts of rare events

⇒ Poisson model:

$$Y_s \sim \mathcal{P}(K\mu'_s) \quad (2)$$

But highly variable observations AND over-representation of zeros

⇒ Zero-Inflated Poisson model:

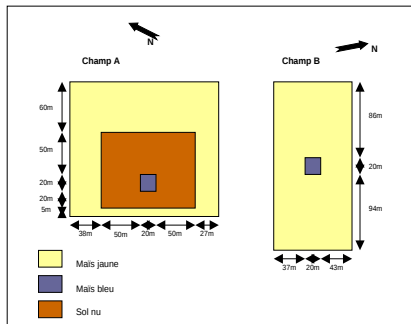
$$Y_s \sim ZIP(1 - q_s, K\mu'_s) \quad (3)$$

Taking into account the remaining variability

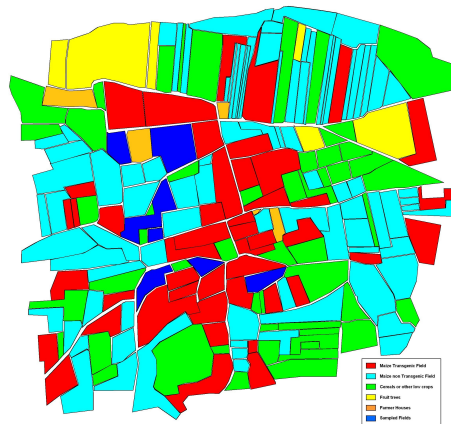
⇒ Normal model for the Poisson model expectation :

$$\mu'_s \sim \mathcal{N}(\gamma(d_s), \sigma^2) \quad (4)$$

Types of data



(Source: Klein et al., 2003)



(Source: Messeguer et al., 2006)

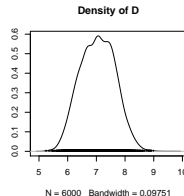
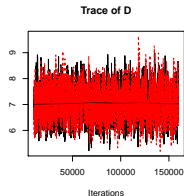
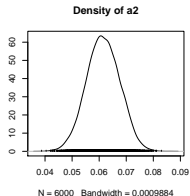
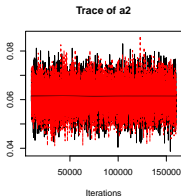
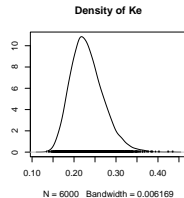
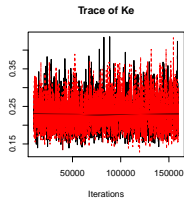
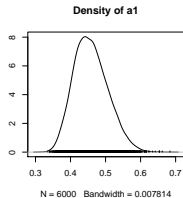
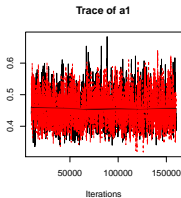
Outline

- 1 Introduction
 - Context
 - Objectives
- 2 Materials and Methods
 - Dispersal model
 - Observation model
 - Types of data
- 3 Results
 - Parameter estimation
 - Overall variability
 - Goodness of fit
 - Uncertainty quantification
- 4 Discussion and Perspectives

Parameter estimation

- Method: Monte Carlo Markov Chain (MCMC)
- Software for estimation and inference: JAGS
- Convergence to posterior distribution: OK

$$\gamma(d_s) = \begin{cases} K \times e^{-a_1 d_s} & d_s < D \\ K \times e^{-a_1 D - a_2 d_s} & d_s \geq D \end{cases}$$



Overall variability

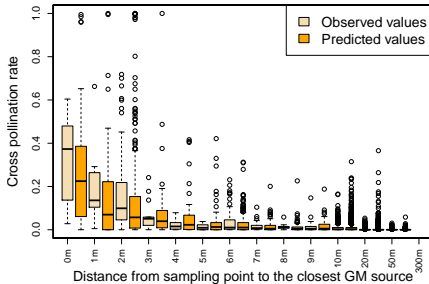


Figure: Boxplots of observed and predicted cross-pollination rates as a function of distance to the closest GM source.

⇒ Both models reflect the overall variability well.

⇒ The observation model ($ZIP + Normal$ expectation) is sufficient to reproduce this variability.

Goodness of fit

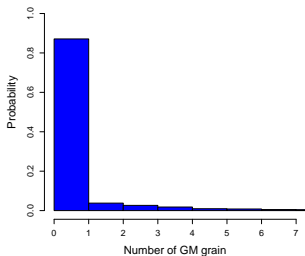
Criterion	Model	
	Distance	Distance + Wind
<i>CRPS</i>	-2.720	-2.400 ✓
<i>Correlation</i>	0.696	0.746 ✓
<i>Efficiency</i>	0.465	0.536 ✓
<i>RMSE</i>	14.234	13.254 ✓
<i>ROC</i>	0.813	0.863 ✓

Table: Criteria values calculated with mean predictions for the two models

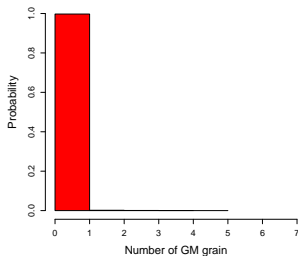
⇒ **Distance+Wind** model outperforms **Distance** model either on absolute error (RMSE, efficiency, correlation) or on quality of decision (ROC).

Uncertainty quantification

**Histogram of a
Distance model Prediction**



**Histogram of a
Distance+Wind model Prediction**



- Both models predict the observed value as the more likely.
- **Distance+Wind** model gives more precise predictions.

⇒ Incorporate wind effect reduce uncertainty prediction.

⇒ Better description of the situation increases the accuracy of prediction.

Outline

- 1 Introduction
 - Context
 - Objectives
- 2 Materials and Methods
 - Dispersal model
 - Observation model
 - Types of data
- 3 Results
 - Parameter estimation
 - Overall variability
 - Goodness of fit
 - Uncertainty quantification
- 4 Discussion and Perspectives

Discussion and Perspectives

Discussion

- High uncertainty and/or unavailability of input that must be taken into account.
- High uncertainty in predictions that need to be assessed.

Discussion and Perspectives

Discussion

- High uncertainty and/or unavailability of input that must be taken into account.
- High uncertainty in predictions that need to be assessed.

- Bayesian methods allows for probabilistic predictions.
 - ⇒ Overall variability well reflected.
 - ⇒ Provide better understanding of prediction uncertainty.
- Substantial benefit of a better description of the situation.
 - ⇒ Decrease of absolute error.
 - ⇒ Decrease of prediction uncertainty.
- Probabilistic predictions provide better insight into the risk for decision makers.
 - ⇒ Decisions may adapt to the level of risk aversion.

Discussion and Perspectives

Discussion

- High uncertainty and/or unavailability of input that must be taken into account.
- High uncertainty in predictions that need to be assessed.

- Bayesian methods allows for probabilistic predictions.
 - ⇒ Overall variability well reflected.
 - ⇒ Provide better understanding of prediction uncertainty.
- Substantial benefit of a better description of the situation.
 - ⇒ Decrease of absolute error.
 - ⇒ Decrease of prediction uncertainty.
- Probabilistic predictions provide better insight into the risk for decision makers.
 - ⇒ Decisions may adapt to the level of risk aversion.

Perspectives

- Extension to *Multi-source* situations.
- Incorporate the effect of flowering delay.
- Refine the individual dispersal function.

Thanks for listening
Any questions?



Life is like a Markov chain, you never know what you're gonna get