



**HAL**  
open science

## Modélisation des flux de gènes à l'échelle du paysage par approche bayésienne. Application à l'aide à la décision pour la coexistence entre cultures OGM et non OGM

Arnaud Bensadoun, Antoine A. Messean, Hervé Monod

### ► To cite this version:

Arnaud Bensadoun, Antoine A. Messean, Hervé Monod. Modélisation des flux de gènes à l'échelle du paysage par approche bayésienne. Application à l'aide à la décision pour la coexistence entre cultures OGM et non OGM. Journées de l'école doctorale ABIES, Feb 2013, Paris, France. 48 p. hal-02804573

**HAL Id: hal-02804573**

**<https://hal.inrae.fr/hal-02804573>**

Submitted on 5 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.

Modélisation des flux de gènes à l'échelle du paysage par  
approche bayésienne.  
Application à l'aide à la décision pour la coexistence entre  
cultures OGM et non OGM.

**Arnaud Bensadoun**

Directeurs:

Antoine Messéan, Hervé Monod

En collaboration avec:

Frédérique Angevin, David Makowski

Journées ABIES

26-27 février 2013



# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - Modélisation
- 3 Résultats
  - Estimation
  - Prédications
- 4 Discussions

# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - Modélisation
- 3 Résultats
  - Estimation
  - Prédications
- 4 Discussions

## Contexte de la coexistence

- Plusieurs types de cultures existent :  
Conventionnel, biologique ou biotechnologique (OGM).
- Chaque agriculteur doit pouvoir choisir le mode de production qu'il souhaite.  
⇒ Dispersion de pollen  
⇒ Risque de pollinisation croisée
- Le consommateur doit pouvoir choisir le produit qu'il souhaite consommer en connaissance de cause.

## Contexte de la coexistence

- Plusieurs types de cultures existent :  
Conventionnel, biologique ou biotechnologique (OGM).
- Chaque agriculteur doit pouvoir choisir le mode de production qu'il souhaite.  
⇒ Dispersion de pollen  
⇒ Risque de pollinisation croisée
- Le consommateur doit pouvoir choisir le produit qu'il souhaite consommer en connaissance de cause.

### Enjeux

⇒ Assurer la séparation des filières compte tenu des risques de présence fortuite de transgène dans des lots considérés comme non OGM.

## Règlementation et distance d'isolement

Commission Européenne formule des recommandations pour l'établissement de mesures de coexistence :

## Règlementation et distance d'isolement

Commission Européenne formule des recommandations pour l'établissement de mesures de coexistence :

- Fonder ses décisions sur des résultats scientifiques
- Prendre des mesures spécifiques en fonction des espèces et variétés mais aussi du contexte régional

## Règlementation et distance d'isolement

Commission Européenne formule des recommandations pour l'établissement de mesures de coexistence :

- Fonder ses décisions sur des résultats scientifiques
- Prendre des mesures spécifiques en fonction des espèces et variétés mais aussi du contexte régional

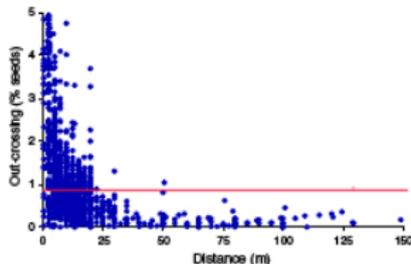


Fig.: Taux de pollinisation croisée en fonction de la distance à la source OGM  
Source : Riesgo et al., 2010

## Règlementation et distance d'isolement

Commission Européenne formule des recommandations pour l'établissement de mesures de coexistence :

- Fonder ses décisions sur des résultats scientifiques
- Prendre des mesures spécifiques en fonction des espèces et variétés mais aussi du contexte régional

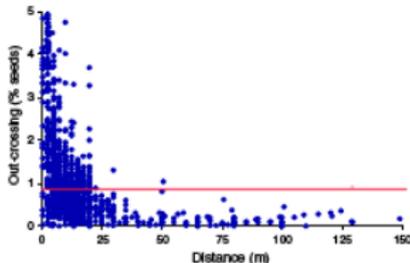


Fig.: Taux de pollinisation croisée en fonction de la distance à la source OGM  
Source : Riesgo et al., 2010

## Enjeux

- ⇒ Adapter les règles de coexistence à la diversité des situations possibles.
- ⇒ Intégrer la variabilité et les incertitudes dans la prise de décision.

## Connaissances et Besoins

- Connaissances sur les risques de mélange et en particulier sur les flux de gènes.  
⇒ Bases de données de résultats d'expérimentations au champ
- Modèles génériques de la dispersion du transgène à l'échelle des paysages :  
modèles MAPOD sur maïs (Angevin et al., 2008) et GeneSys sur colza (Colbach et al., 2001)

## Connaissances et Besoins

- Connaissances sur les risques de mélange et en particulier sur les flux de gènes.  
⇒ Bases de données de résultats d'expérimentations au champ
- Modèles génériques de la dispersion du transgène à l'échelle des paysages : modèles MAPOD sur maïs (Angevin et al., 2008) et GeneSys sur colza (Colbach et al., 2001)

## Objectifs de la thèse

- Elaboration de modèles simplifiés
- Quantification de l'incertitude des prédictions
- Intégration de la variabilité et des incertitudes dans la prise de décision
- Valorisation de données hétérogènes à l'échelle du paysage



**Intérêt de l'approche bayésienne**

# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - Modélisation
- 3 Résultats
  - Estimation
  - Prédications
- 4 Discussions

## Dispositifs Expérimentaux

Les flux de gènes sur maïs ont été étudiés selon deux types de dispositif :

- Une seule parcelle émettrice (une ou plusieurs parcelles réceptrices). Dispositif le plus courant
- Dispositifs multisources (plusieurs émetteurs) à l'échelle du paysage dans des situations réelles de coexistence.

## Dispositifs Expérimentaux

Les flux de gènes sur maïs ont été étudiés selon deux types de dispositif :

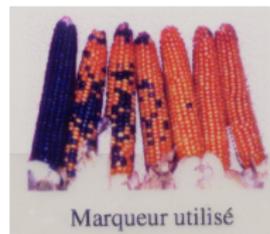
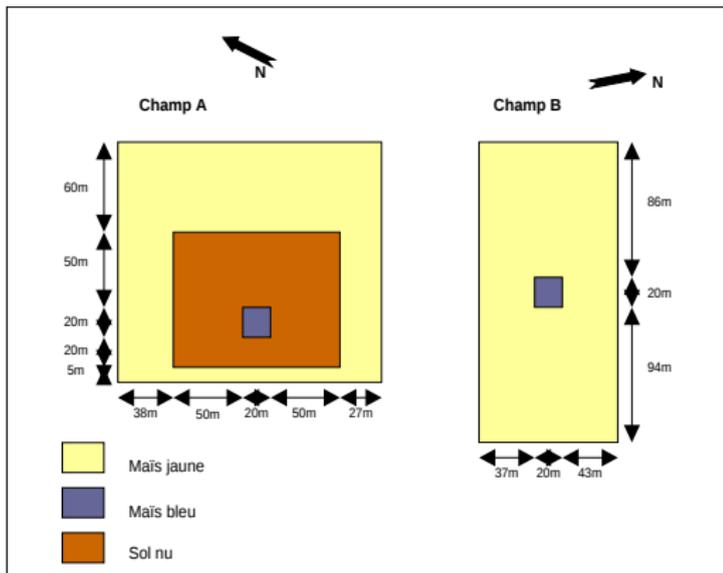
- Une seule parcelle émettrice (une ou plusieurs parcelles réceptrices). Dispositif le plus courant
- Dispositifs multisources (plusieurs émetteurs) à l'échelle du paysage dans des situations réelles de coexistence.

Observations :

- $y(d)$  ou  $y(x, y)$  : nombre de grains contenant le transgène en un point
- $\bar{N}$  : Nombre moyen de grains par épi

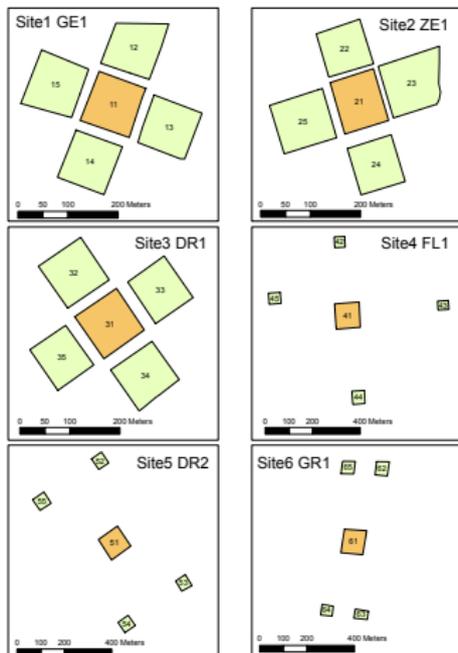


## Dispositif "Mono-source" (France)



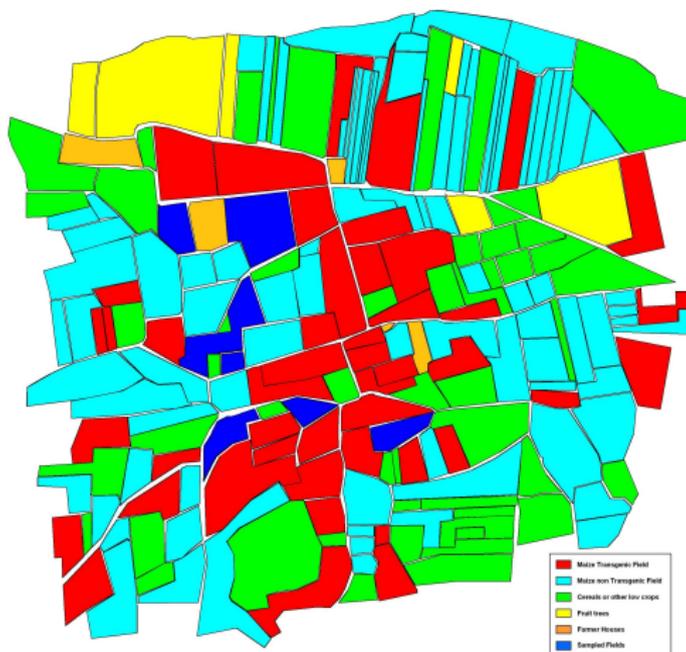


## Dispositif “Mono-source” (Pays-Bas)



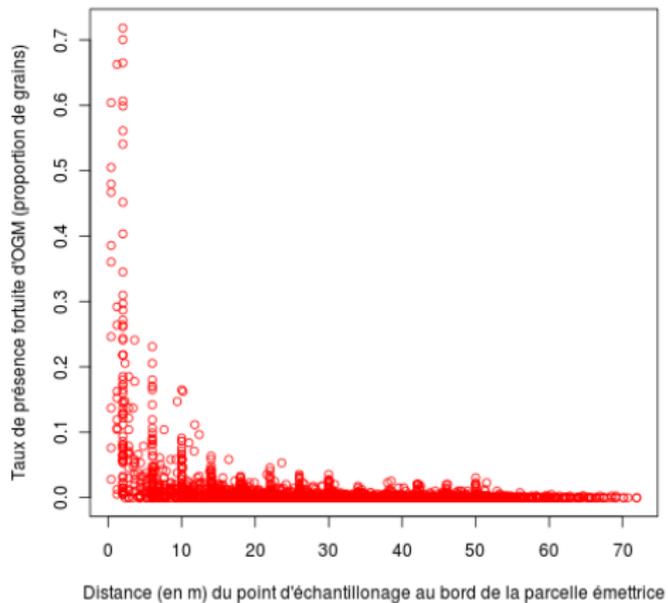


## Dispositif "Multi-sources" (Espagne)





## Exemple - Données : Essai Montargis 1998



# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - **Modélisation**
- 3 Résultats
  - Estimation
  - Prédications
- 4 Discussions

## Fonction de dispersion individuelle

*Fonction de dispersion individuelle :*

$\gamma_{\theta}(x - x', y - y')$  représente la probabilité qu'un grain de pollen émis en  $(x', y')$  tombe et fertilise une plante en  $(x, y)$ .

## Fonction de dispersion individuelle

*Fonction de dispersion individuelle :*

$\gamma_{\theta}(x - x', y - y')$  représente la probabilité qu'un grain de pollen émis en  $(x', y')$  tombe et fertilise une plante en  $(x, y)$ .

$\iint_{(x', y')} \gamma_{\theta}(x - x', y - y') dx' dy'$  représente donc le nombre d'ovules en  $(x, y)$  fécondés par un grain de pollen émis dans un carré  $dx' dy'$

## Fonction de dispersion individuelle

*Fonction de dispersion individuelle :*

$\gamma_{\theta}(x - x', y - y')$  représente la probabilité qu'un grain de pollen émis en  $(x', y')$  tombe et fertilise une plante en  $(x, y)$ .

$\iint_{(x', y')} \gamma_{\theta}(x - x', y - y') dx' dy'$  représente donc le nombre d'ovules en  $(x, y)$  fécondés par un grain de pollen émis dans un carré  $dx' dy'$

**Or** observations :  $y(x, y)$  nombre de grains en  $(x, y)$  fécondés par une source marquée  
⇒ Résultante de dispersion individuelle d'un grand nombre de plantes

## Modèle de dispersion

L'espérance du nombre de grains marqués  $\mu_\theta$  en un point  $(x, y)$  est reliée à la fonction de dispersion individuelle moyennant quelques hypothèses.

## Modèle de dispersion

L'espérance du nombre de grains marqués  $\mu_\theta$  en un point  $(x, y)$  est reliée à la fonction de dispersion individuelle moyennant quelques hypothèses.

$$\mu_\theta(x, y) = \bar{N} \times \frac{\iint_{(x', y') \in A} \gamma_\theta(x - x', y - y') dx' dy'}{\iint_{(x', y') \in A} \gamma_\theta(x - x', y - y') dx' dy' + \iint_{(x', y') \in B} \gamma_\theta(x - x', y - y') dx' dy'}$$

Avec  $A$  : Champ OGM

$B$  : Champ non OGM

## Paradigme Bayésien

$$P(\theta|Y) \propto \pi(\theta) \times \mathcal{L}(Y|\theta)$$

## Paradigme Bayésien

$$P(\theta|Y) \propto \pi(\theta) \times \mathcal{L}(Y|\theta)$$

Nécessite :

- Des données  $Y$
- Des connaissances *a priori*  $\pi(\cdot)$  sur les paramètres  $\theta$
- Une fonction de vraisemblance  $\mathcal{L}(Y|\theta)$

⇒ Requiert la définition d'un modèle probabiliste reliant les observations aux paramètres du modèle.

## Modèles d'observations

Observation de comptages d'évènements relativement rares  
⇒ Modèle de Poisson

$$Y \sim \mathcal{P}(\mu') \quad (1)$$

## Modèles d'observations

Observation de comptages d'évènements relativement rares

⇒ Modèle de Poisson

$$Y \sim \mathcal{P}(\mu') \quad (1)$$

Mais observations très variables ET sur-représentation de zéros

⇒ Modèle de Poisson Zéro Inflaté

$$Y \sim ZI\mathcal{P}(1 - q, \mu') \quad (2)$$

⇔

$$Y|Z = 0 \sim \delta(\{0\})$$

$$Y|Z = 1 \sim \mathcal{P}(\mu')$$

## Modèles d'observations

Observation de comptages d'évènements relativement rares

⇒ Modèle de Poisson

$$Y \sim \mathcal{P}(\mu')$$
 (1)

Mais observations très variables ET sur-représentation de zéros

⇒ Modèle de Poisson Zéro Inflaté

$$Y \sim ZI\mathcal{P}(1 - q, \mu')$$
 (2)

⇔

$$Y|Z = 0 \sim \delta(\{0\})$$

$$Y|Z = 1 \sim \mathcal{P}(\mu')$$

Gestion de la variabilité supplémentaire

⇒ Modèle normale pour l'espérance

$$\mu' \sim \mathcal{N}(\mu_\theta, \sigma^2)$$
 (3)

$$\mu' \sim \mathcal{N}(\mu_\theta, \mu_\theta \times \sigma^2)$$
 (4)

## Comparaison/Sélection de modèles

Plusieurs modèles d'observations pour  $Y$  : Poisson et Poisson Zéro Inflaté

Plusieurs modèles pour l'espérance  $\mu'$  : Fixe, Normale et Normale à variance proportionnelle

Plusieurs fonctions de dispersion individuelle  $\gamma_\theta$  : Exponentielle, 2Dt, NIG

## Comparaison/Sélection de modèles

Plusieurs modèles d'observations pour  $Y$  : Poisson et Poisson Zéro Inflaté

Plusieurs modèles pour l'espérance  $\mu'$  : Fixe, Normale et Normale à variance proportionnelle

Plusieurs fonctions de dispersion individuelle  $\gamma_\theta$  : Exponentielle, 2Dt, NIG



**Plan factoriel complet**

## Plan d'expérience

Distribution $Y$	Distribution $\mu'$	IDF $\gamma_\theta$
$\mathcal{P}(\mu')$	$\mu$	Exponential
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \sigma^2)$	Exponential
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	Exponential
$\mathcal{P}(\mu')$	$\mu$	2Dt
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \sigma^2)$	2Dt
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	2Dt
$\mathcal{P}(\mu')$	$\mu$	NIG
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \sigma^2)$	NIG
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	NIG
$ZIP(1 - q, \mu')$	$\mu$	Exponential
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \sigma^2)$	Exponential
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	Exponential
$ZIP(1 - q, \mu')$	$\mu$	2Dt
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \sigma^2)$	2Dt
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	2Dt
$ZIP(1 - q, \mu')$	$\mu$	NIG
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \sigma^2)$	NIG
$ZIP(1 - q, \mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	NIG

# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - Modélisation
- 3 Résultats
  - Estimation
  - Prédications
- 4 Discussions

## Estimation des paramètres

Approche : Statistique Bayésienne

Méthode d'estimation : Monte Carlo Markov Chain (MCMC)

Software : JAGS

Convergence : OK

Temps d'estimation : Selon modèle entre 3 et 12h

(Avant optimisation/simplification du calcul de l'intégrale entre 2 et 8 jours!)

# Prédictions

## Modèle de Poisson

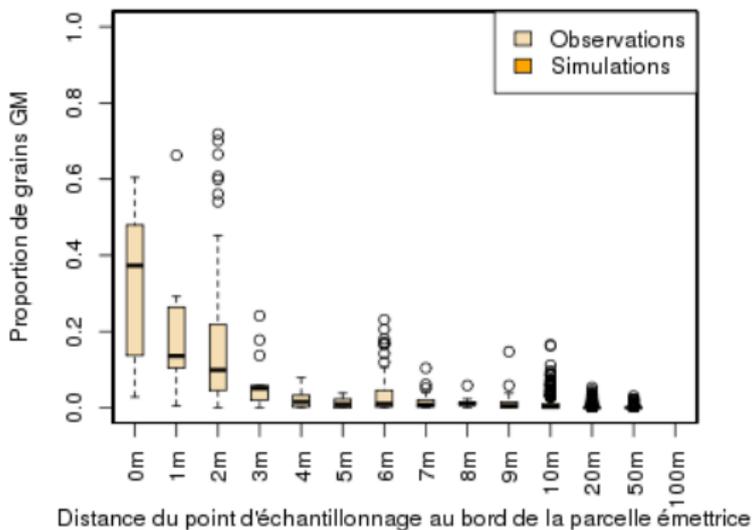


Fig.: Observations en fonction de la distance

# Prédictions

## Modèle de Poisson

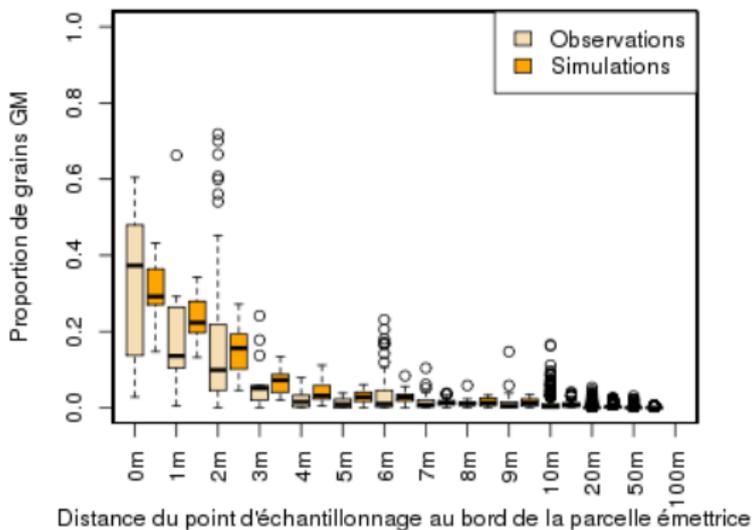


Fig.: Observations en fonction de la distance vs Simulations avec le modèle Poisson 2Dt

# Prédictions

## Modèle de Poisson

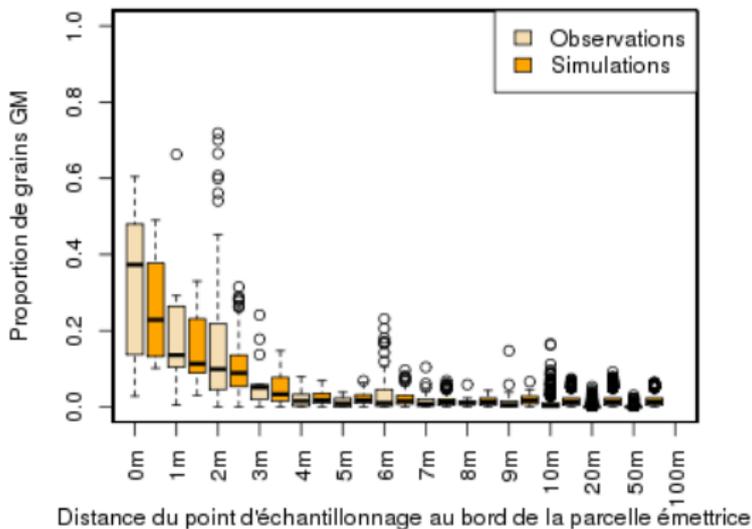


Fig.: Observations en fonction de la distance vs Simulations avec le modèle Poisson 2Dt avec espérance normale

# Prédictions

## Modèle de Poisson

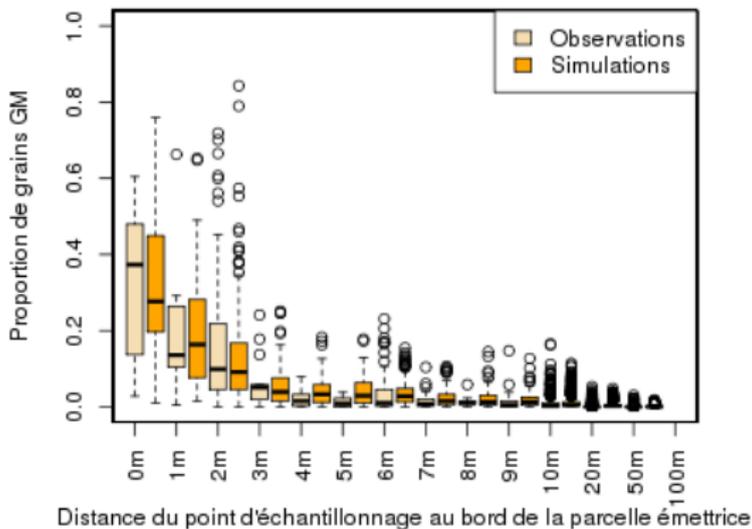


Fig.: Observations en fonction de la distance vs Simulations avec le modèle Poisson 2Dt avec espérance normale à variance proportionnelle

## Choix de modèles

Pour chaque observation  $y_{(x,y)}$  on a une distribution empirique prédictive  $F$

## Choix de modèles

Pour chaque observation  $y_{(x,y)}$  on a une distribution empirique prédictive  $F$

Un bon modèle se caractérise par :

- une moyenne de  $F$  proche de  $y_{(x,y)}$
- une faible variance

## Choix de modèles

Pour chaque observation  $y_{(x,y)}$  on a une distribution empirique prédictive  $F$

Un bon modèle se caractérise par :

- une moyenne de  $F$  proche de  $y_{(x,y)}$
- une faible variance

$$CRPS(F, y_{(x,y)}) = \frac{1}{2} \mathbb{E}_F |Y - Y'| - \mathbb{E}_F |Y - y_{(x,y)}|$$

## Choix de modèles

Pour chaque observation  $y_{(x,y)}$  on a une distribution empirique prédictive  $F$

Un bon modèle se caractérise par :

- une moyenne de  $F$  proche de  $y_{(x,y)}$
- une faible variance

$$CRPS(F, y_{(x,y)}) = \frac{1}{2} \mathbb{E}_F |Y - Y'| - \mathbb{E}_F |Y - y_{(x,y)}|$$

Distribution $Y$	Distribution $\mu'$	IDF $\gamma_\theta$	CRPS
$\mathcal{P}(\mu')$	$\mu$	2Dt	-3.53
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \sigma^2)$	2Dt	-4.86
$\mathcal{P}(\mu')$	$\mathcal{N}(\mu, \mu\sigma^2)$	2Dt	<b>-2.94</b>

# Plan

- 1 Introduction
- 2 Matériels et Méthodes
  - Données
  - Modélisation
- 3 Résultats
  - Estimation
  - Prédictions
- 4 Discussions

# Discussions

- Méthodes bayésiennes
  - Application aux modèles de dispersion individuelle
  - Prise en compte de la variabilité

## Discussions

- Méthodes bayésiennes
  - Application aux modèles de dispersion individuelle
  - Prise en compte de la variabilité
- Sélection de modèle
  - Meilleur modèle?
  - Modèle moyen : *Bayesian Model Averaging*?

## Discussions

- Méthodes bayésiennes
  - Application aux modèles de dispersion individuelle
  - Prise en compte de la variabilité
- Sélection de modèle
  - Meilleur modèle?
  - Modèle moyen : *Bayesian Model Averaging*?
- Prédictions par point vs. prédictions par parcelle

## Discussions

- Méthodes bayésiennes
  - Application aux modèles de dispersion individuelle
  - Prise en compte de la variabilité
- Sélection de modèle
  - Meilleur modèle?
  - Modèle moyen : *Bayesian Model Averaging*?
- Prédictions par point vs. prédictions par parcelle
- Extension au multi-sources
  - Prise en compte du paysage
  - Obtention de covariables à cette échelle
  - Temps de calcul

# Merci pour votre attention

Des questions?

*“La vie c’est comme une chaîne de Markov, on ne sait jamais sur quoi on va tomber !”*