

Estimation de la fonction de dispersion du *Plum pox virus* par les pucerons vecteurs

François Bonnot, Mathilde Lefort, Sylvie Dallot,
Gaël Thébaud, Emmanuel Jacquot

La sharka, une des maladies les plus graves des fruits à noyaux

- *Plum pox virus* (PPV- genre *Potyvirus*)
- Hôtes : *Prunus* cultivés (abricotier, pêcher, prunier) et sauvages
 - Baisse du rendement et de la qualité des fruits
 - Coût des mesures de contrôle de la maladie
- Deux modes de dissémination du virus
 - Propagation de matériel végétal contaminé et échanges commerciaux
 - Par les pucerons sur le mode non persistant
 - Au moins 20 espèces de pucerons (qui ne colonisent pas les *Prunus*)
 - Processus d'acquisition-transmission rapide
 - Les arbres deviennent infectieux après une période de latence de distribution mal connue



Distances de dispersion de la sharka

- **En France, lutte obligatoire contre la maladie dans les vergers :**

- Prospections visuelles pour identifier les arbres symptomatiques
- Arrachage des arbres symptomatiques



- **Une bonne connaissance des distances de dispersion par les pucerons est nécessaire pour optimiser les stratégies de surveillance et de contrôle.**

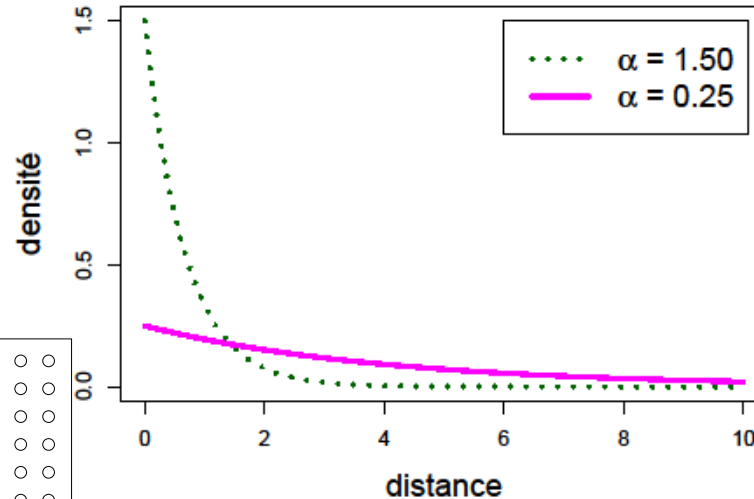
- Analyses préliminaires basées sur la distribution spatiotemporelle des arbres malades dans les vergers (Dallot *et al.*, 2003 and 2004).



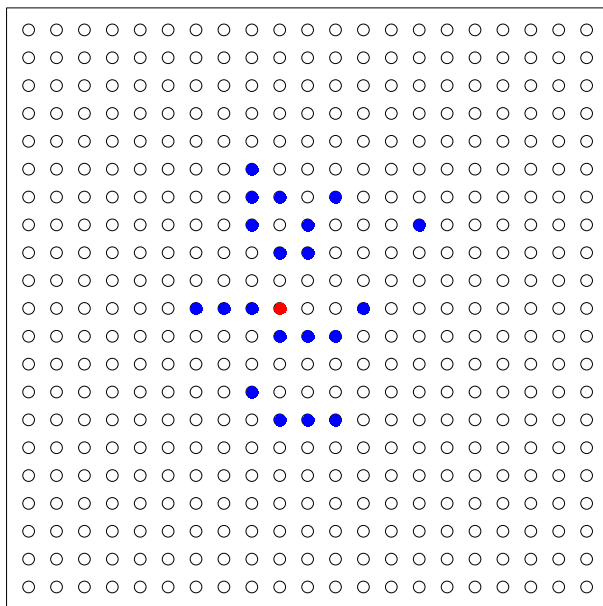
Modélisation spatiotemporelle stochastique

Fonction de dispersion

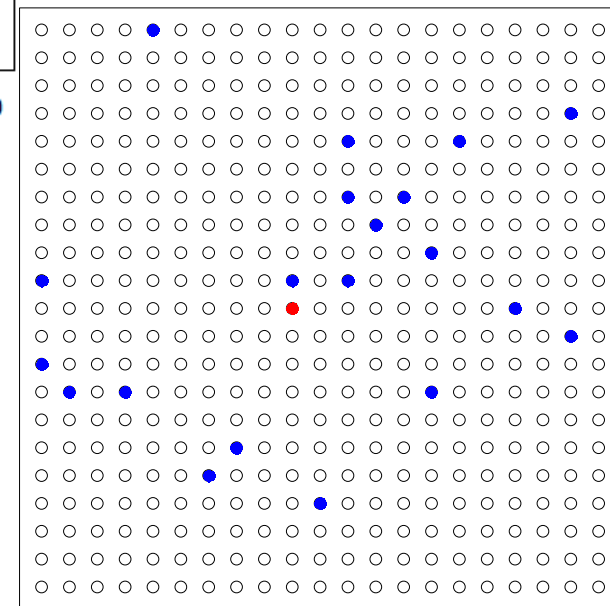
Densité $f(d)$ décrivant la probabilité qu'un arbre infectieux contamine un arbre sain situé à distance d



$\alpha=1.50$



$\alpha=0.25$



Objectifs

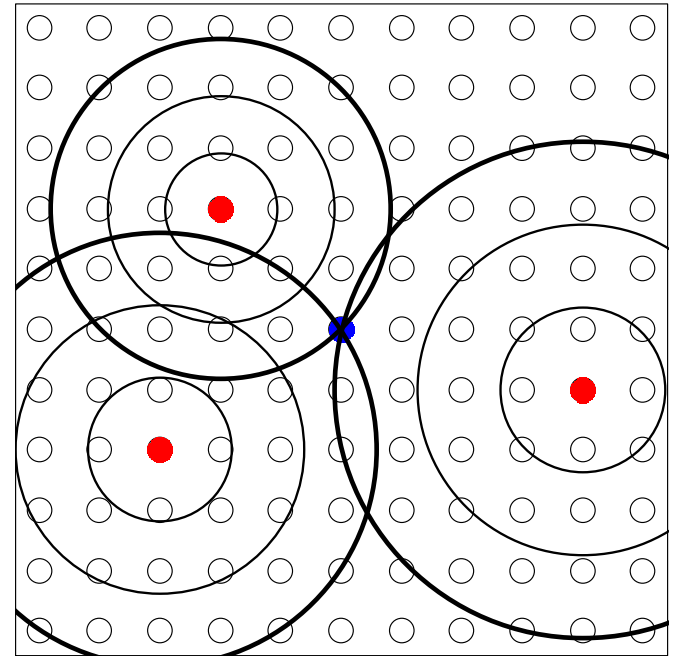
- Implémenter une méthode d'estimation de α à partir de cartes d'une maladie observée à différentes dates (Gibson 1997)
- Tester la méthode existante
- Comparer avec une variante de l'algorithme
- Adapter la méthode d'estimation
 - aux spécificités du pathosystème sharka et des mesures de contrôle
 - aux spécificités du jeu de données disponible

Données épidémiologiques sur le PPV

- 157 parcelles de pêcher, 1000 arbres par parcelle
- Environnement sain
- Détection et arrachage des arbres symptomatiques chaque année pendant 6 années consécutives

Cadre méthodologique

- Les arbres infectieux exercent un **potentiel infectieux** sur les arbres sains.
- Ce potentiel infectieux est défini comme la somme des fonctions de dispersion des arbres infectieux.
- Un arbre sain est infecté avec une probabilité proportionnelle au potentiel infectieux.
- Un arbre infecté devient immédiatement infectieux.

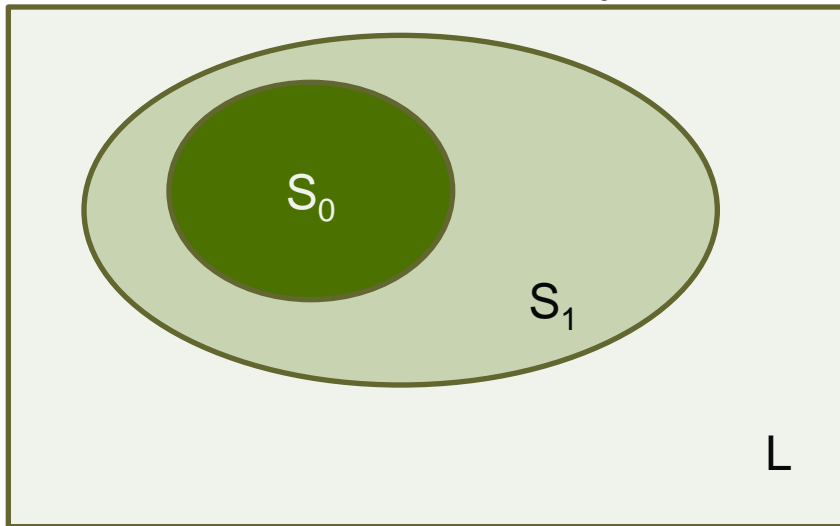


$$P(x \text{ devient infecté pendant } [t, t + \Delta t]) = k(t) \sum_{y \text{ infectieux}} f_{\alpha}(|y - x|) \Delta t$$

Potentiel infectieux en x

Cadre méthodologique

2 dates d'observation t_0 et t_1



- L = ensemble des arbres de la parcelle
- S_0 = arbres infectieux au temps t_0
- S_1 = arbres infectieux au temps t_1
- $S_1 \setminus S_0$ = nouveaux arbres infectés entre t_0 et t_1
- $L \setminus S_1$ = arbres restés sains au temps t_1

L'estimation de α est basée sur la probabilité de l'évènement :

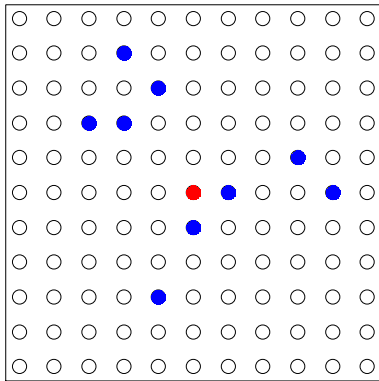
E = « Tous les arbres de $S_1 \setminus S_0$ deviennent infectieux avant tous les arbres de $L \setminus S_1$ »

Vraisemblance de α : $G(\alpha) = P(E|\alpha)$

Calcul de $G(\alpha)$

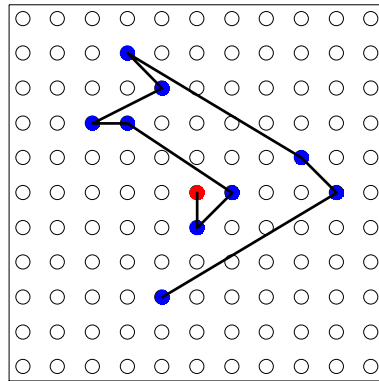
$$G(\alpha) = P(E|\alpha) = P[t_\alpha(x) < t_\alpha(y), \forall x \in S_1 \setminus S_0, \forall y \in L \setminus S_1] \quad t_\alpha(x) = \text{date d'infection de } x$$

Données observées



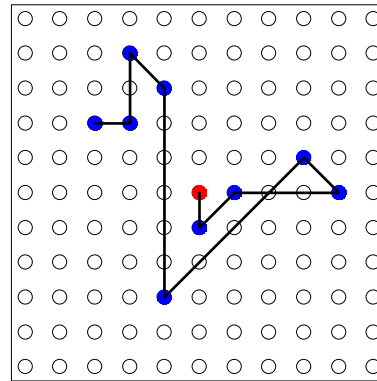
$G(\alpha)?$

ω_1 : Ordre 1



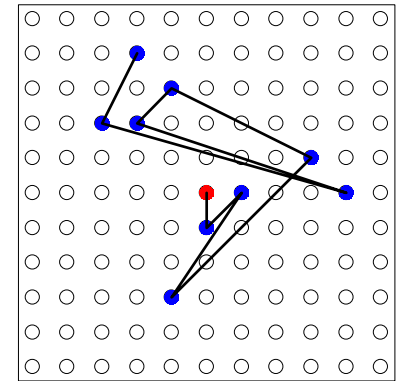
$P_{\omega_1}(\alpha)$

ω_2 : Ordre 2



$P_{\omega_2}(\alpha)$

ω_N : Ordre N



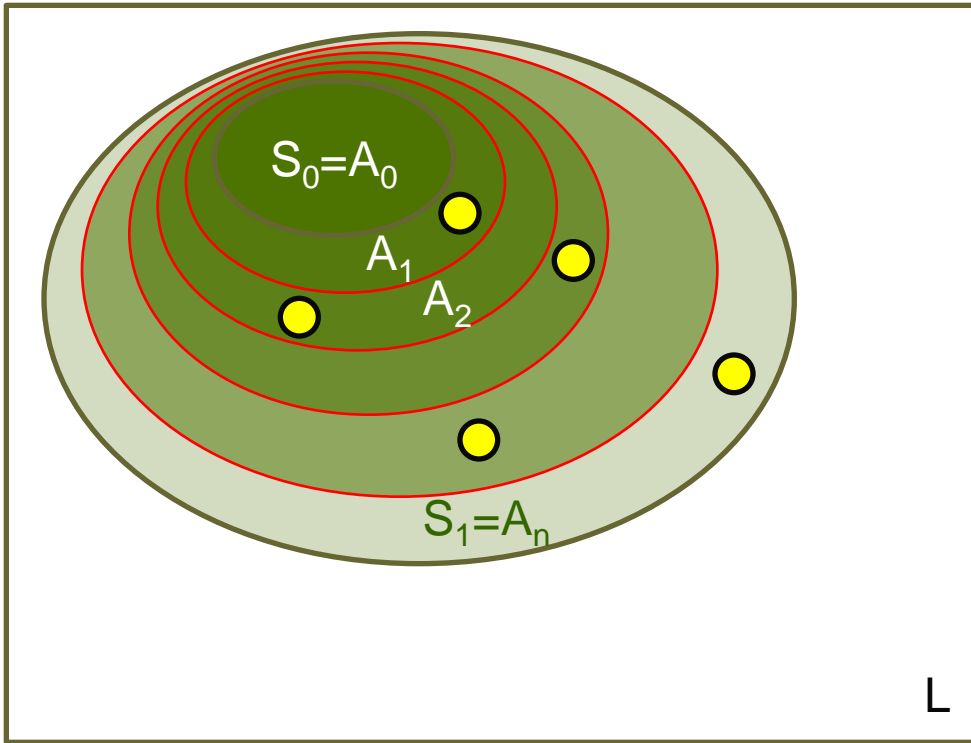
$P_{\omega_N}(\alpha)$

$$G(\alpha) = \sum_{\omega} P_{\omega}(\alpha)$$

Propriété: $G(\alpha)$ ne dépend pas de $k(t)$

Calcul de $P_\omega(\alpha)$

$$S_0 = A_0 \rightarrow A_1 \rightarrow \dots \rightarrow A_i \rightarrow \dots \rightarrow A_n = S_1 \quad A_{i+1} = A_i \cup \{x_{i+1}\}$$



$$P_\omega(\alpha) = \prod_{i=0}^{n-1} P(A_i \rightarrow A_{i+1} | \alpha)$$

$$P(A_i \rightarrow A_{i+1} | \alpha) = \frac{C(x_{i+1} | A_i)}{\sum_{y \in L \setminus A_i} C(y | A_i)}$$

$$C(x | A) = \sum_{y \in A} f_\alpha(|y - x|)$$

Potentiel infectieux exercé par A sur x

Estimation de α pour deux dates d'observation

$$G(\alpha) = \sum_{\omega} P_{\omega}(\alpha) \text{ non calculable numériquement (somme de } n! \text{ termes)}$$



Utilisation de méthodes de Monte Carlo

1) Tirage aléatoire des ordres (Gibson 1997)

Si l'ordre ω est tiré avec la probabilité $g(\omega)$:

$$E\left[\frac{P_{\omega}(\alpha)}{g(\omega)}\right] = G(\alpha)$$

La moyenne de m termes $\frac{P_{\omega}(\alpha)}{g(\omega)}$ obtenus par tirages équiprobables de ω

$$\left(g(\omega) = \frac{1}{n!}\right) \text{ est donc un estimateur de } G(\alpha).$$

Mais variance importante si transmission à courte distance (Gibson)

Estimation de α pour deux dates d'observation

2) Méthode MCMC 1 (Gibson 1997)

Si la distribution *a priori* de α est uniforme : $P(\alpha|E) = \sum_{\omega} P(\alpha, \omega|E) \propto G(\alpha)$

Démarche : construire une chaîne de Markov $(\alpha_1, \omega_1), (\alpha_2, \omega_2), \dots, (\alpha_m, \omega_m)$
 dont la distribution stationnaire est $P(\alpha, \omega|E)$

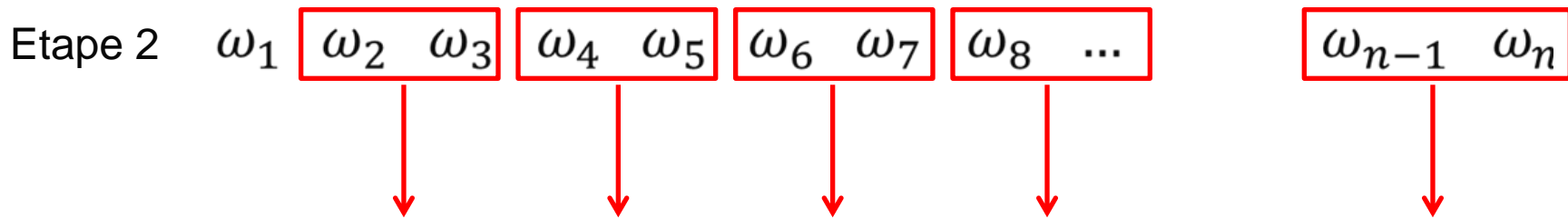
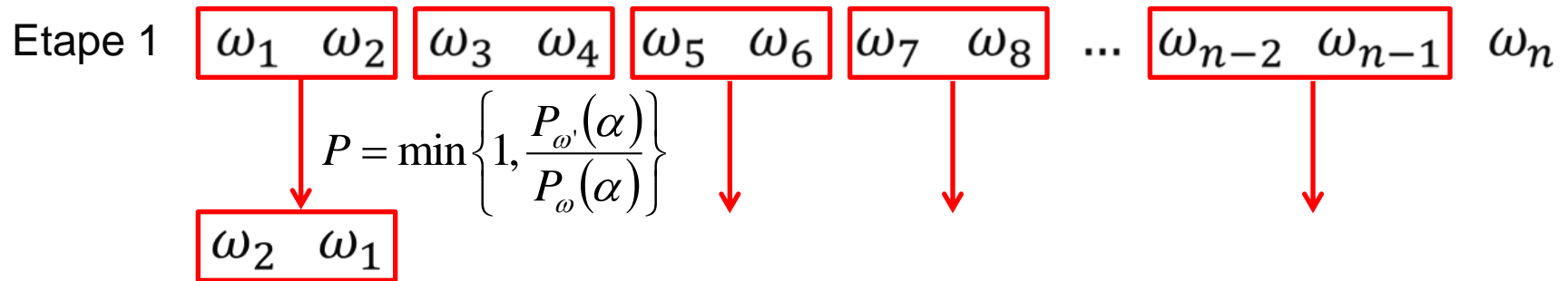
$$\hat{P}(\alpha|E) = \frac{1}{m} \sum_{i=1}^m P(\alpha|\omega_i, E) \quad \text{permet d'estimer } \alpha$$

Estimation de α pour deux dates d'observation

Méthode MCMC 1 : Construction de la chaîne de Markov

Obtention de $(\alpha_{i+1}, \omega_{i+1})$ à partir de (α_i, ω_i)

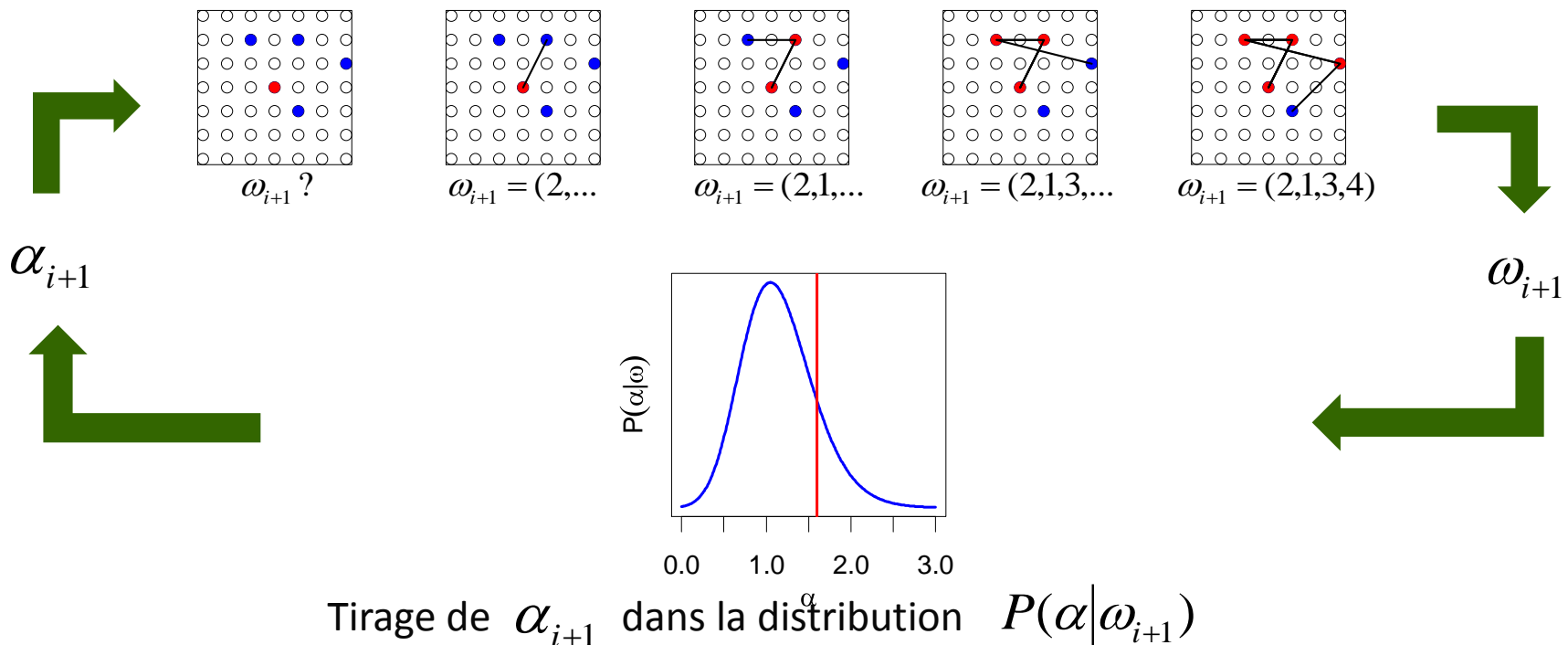
- Tirage de α_{i+1} dans $P(\alpha|\omega_i, E)$
- Obtention de ω_{i+1} à partir de ω_i au moyen d'un processus de « mutation » :



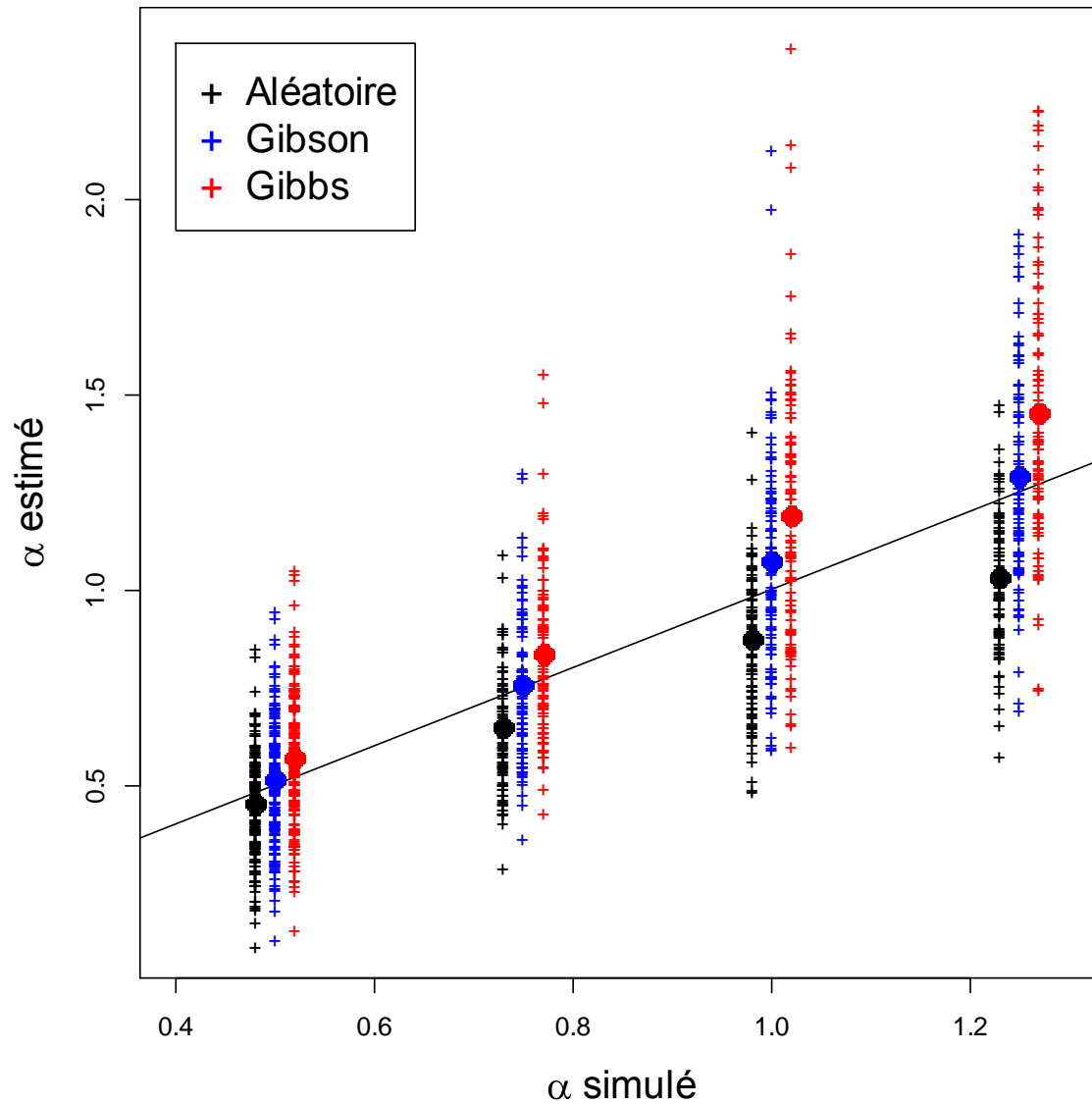
Estimation de α pour deux dates d'observation

3) Méthode MCMC 2 : Echantillonneur de Gibbs

- Construction d'une chaîne de Markov $(\alpha_1, \omega_1), (\alpha_2, \omega_2), \dots, (\alpha_m, \omega_m)$:
 - Itération 1 : α_1 fixé
 - Itération i : Tirage de l'ordre ω_{i+1} dans la distribution $P(\omega|\alpha_i)$

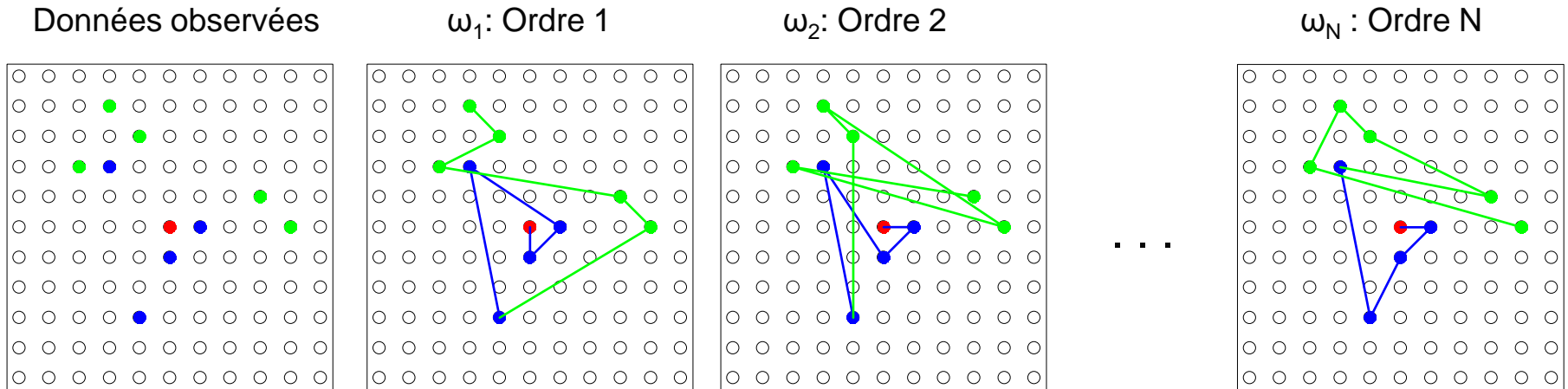


Comparaison des 3 méthodes d'estimation de α



Extension à plus de deux dates d'observation

- Seuls les ordres compatibles avec les cartes successives sont considérés



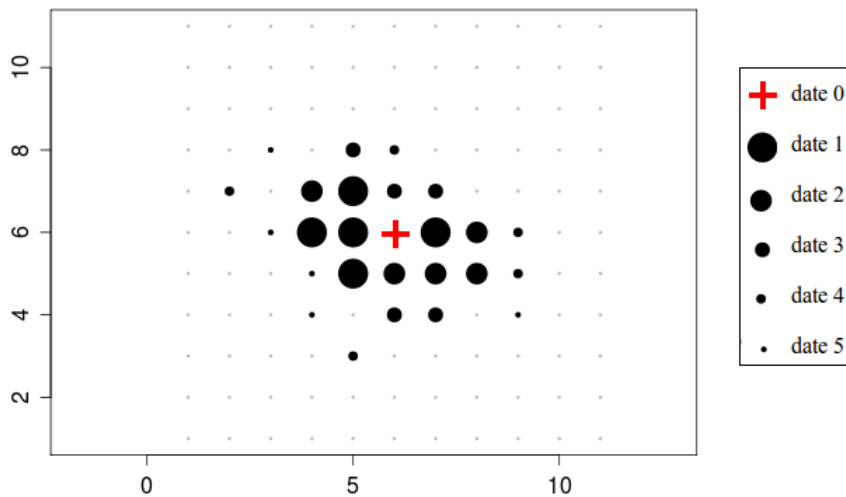
$$N = n_1! \times n_2!$$

$$G(\alpha) = \sum_{\omega} P_{\omega}(\alpha)$$

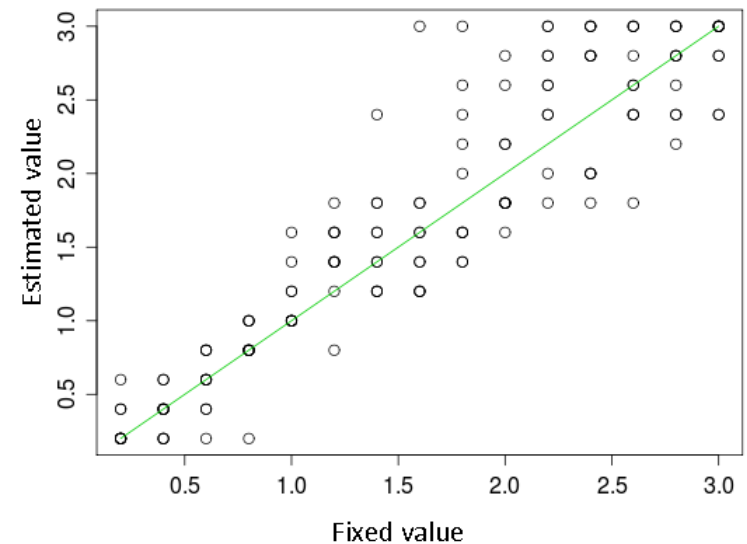
- Construction de la chaîne de Markov $(\alpha_1, \omega_1), (\alpha_2, \omega_2), \dots, (\alpha_N, \omega_N)$

Extension à plus de deux dates d'observation

Données simulées avec α fixé



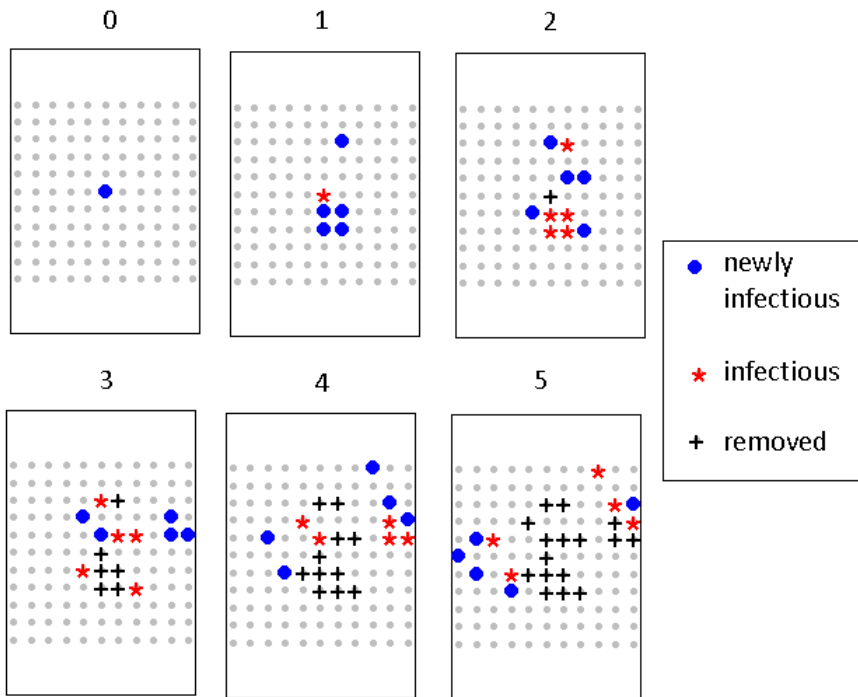
76% des estimations diffèrent de moins de 0.4 unités des valeurs fixées



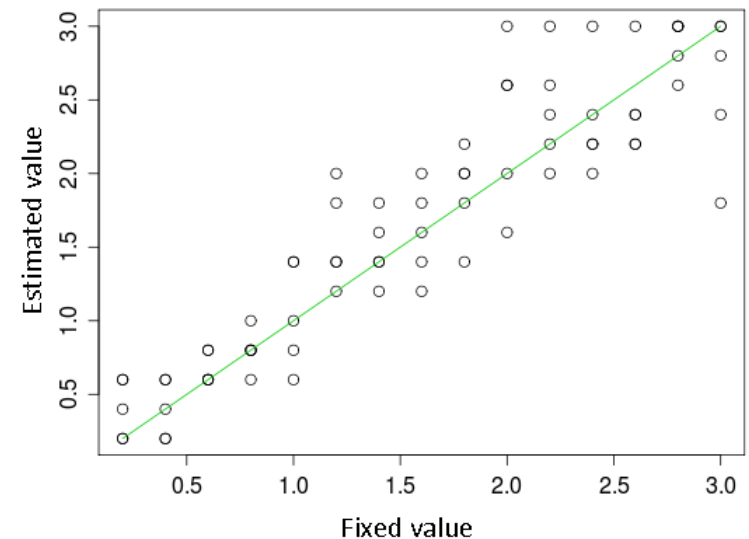
Prise en compte de l'arrachage des arbres

- Les ordres sont conditionnés à l'ensemble des arbres restants

Données simulées avec α fixé

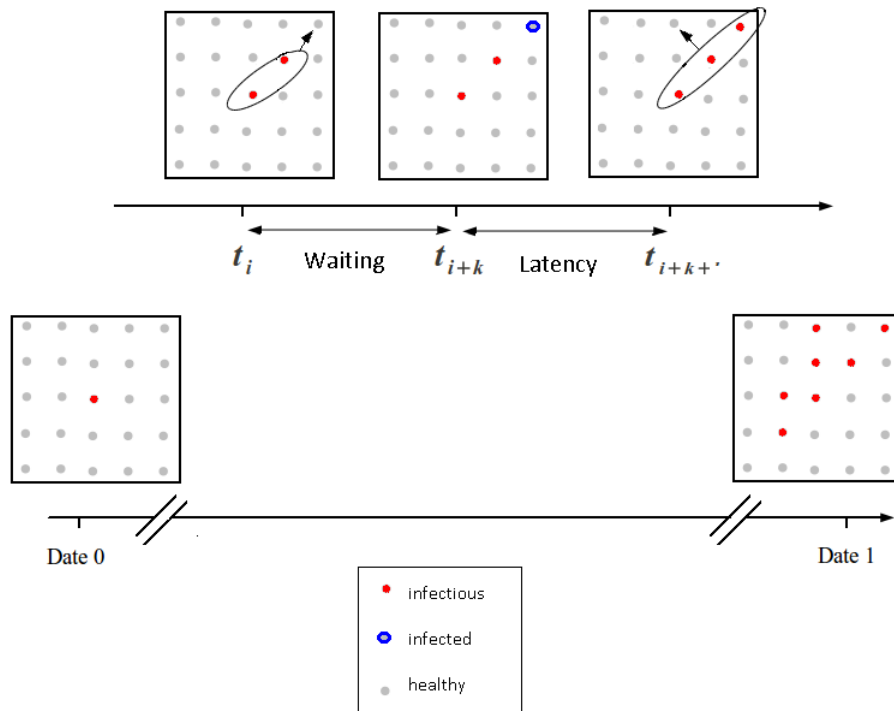


85% des estimations diffèrent de moins de 0.4 unités des valeurs fixées

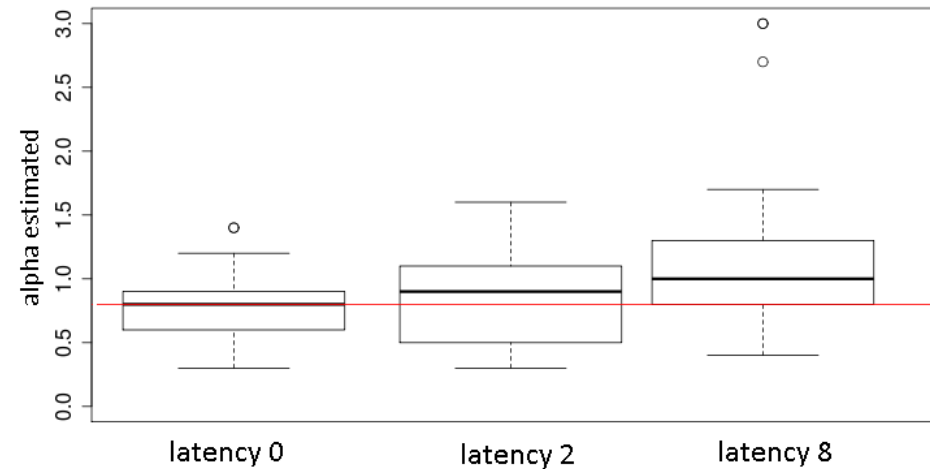


Effet de la latence

Simulation d'un nouvel arbre infecté



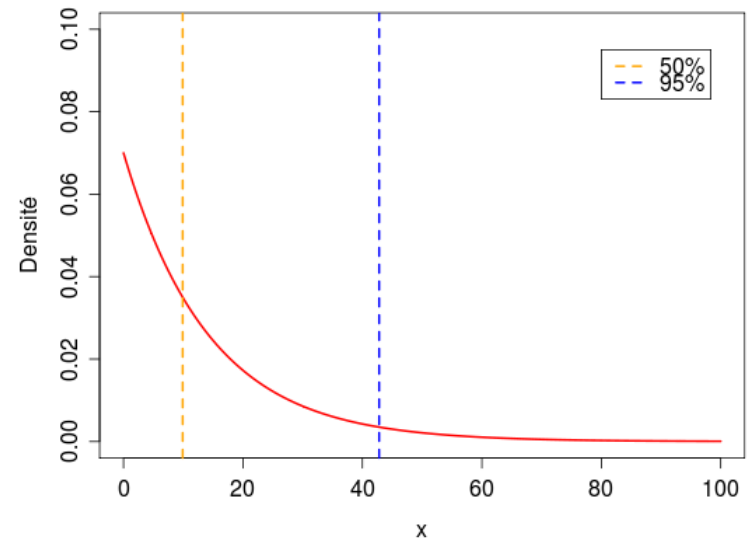
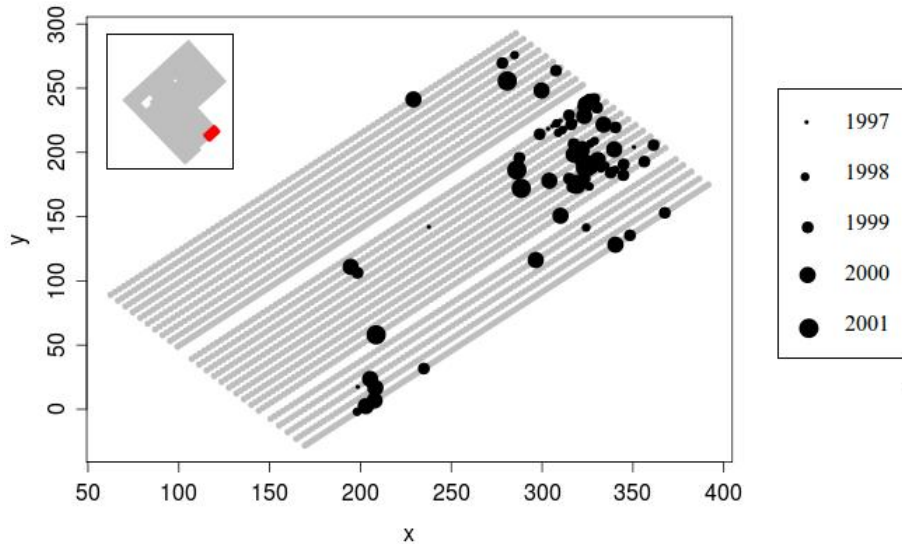
Le biais de l'estimation croît avec la latence



$$\alpha = 0.8$$

Application aux données réelles

50% des évènements de dispersion se produisent à moins de 10 m de la source et 95% à moins de 43 m



Conclusions

- Développement d'une méthode d'estimation de la fonction de dispersion de la sharka adaptée :
 - aux gros jeux de données
 - aux données observées à plusieurs dates
 - aux données incluant des arbres arrachés
- La latence biaise l'estimation

Perspectives

- Affiner la méthode :
 - Améliorer la procédure d'estimation
 - Poursuivre l'analyse de l'influence de la latence
 - Intégrer la probabilité de non-détection
 - Appliquer la méthode au jeu de données