

Un modèle hiérarchique bayésien pour évaluer l'efficacité de traitements à partir de données d'intensité et de prévalence

Armand Favrot - INRAE

David Makowski - INRAE

Fabienne Maupas - ITB

Cédric Royer - ITB

Amélie Monteiro - ITB

Audrey Fabarez - ITB

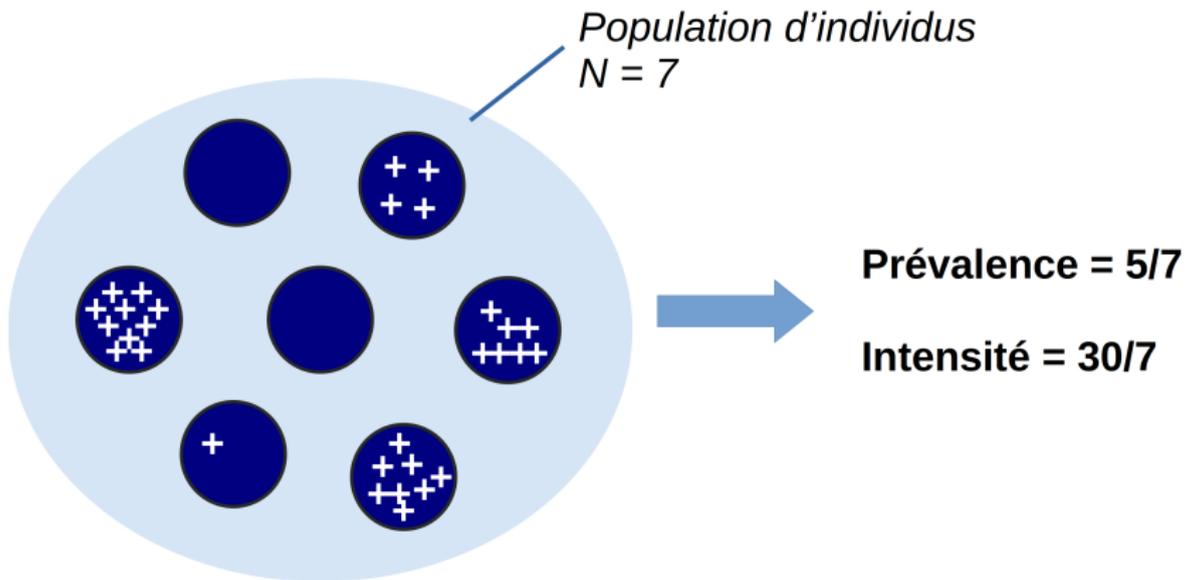
Céline Gouwie - ITB

Elma Raaijmakers - IRS

Kathleen Antoons - IRBAB



- 1 Contexte
- 2 Description des données
- 3 Modélisation
- 4 Conclusion



1 Contexte

2 Description des données

3 Modélisation

4 Conclusion

- Pucerons : vecteur du virus de la jaunisse
→ pertes de rendements importantes
- Néonicotinoïdes très efficace contre les pucerons
- 2019 : néonicotinoïdes interdits pour des raisons environnementales
- 2020 : lourdes pertes ⇒ néonicotinoïdes ré-autorisés à partir de 2021 le temps de trouver des alternatives
→ essais supervisés par l'Institut Technique de la Betterave
→ essais dans d'autres pays européens : Belgique et Pays-Bas
⇒ **Question : quel est le traitement le plus efficace ?**
- 2023 : nouvelle interdiction prononcé par la cour de justice de l'Union Européenne



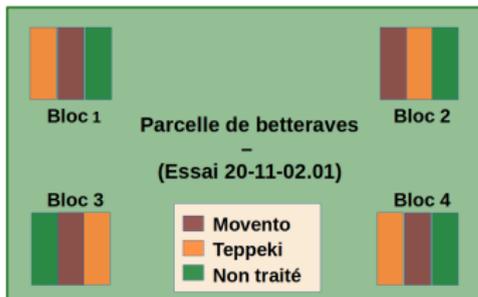
1 Contexte

2 Description des données

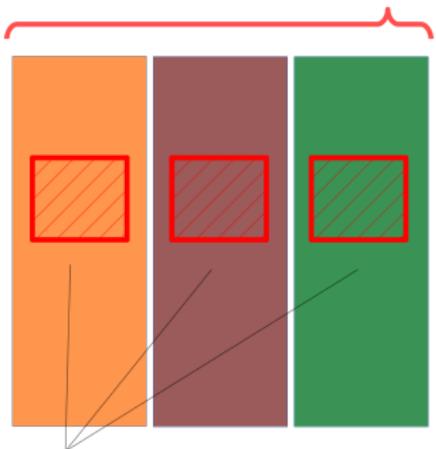
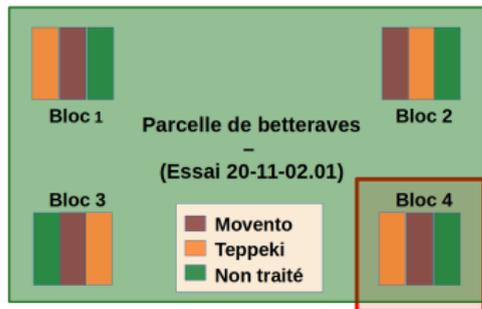
3 Modélisation

4 Conclusion

Data description

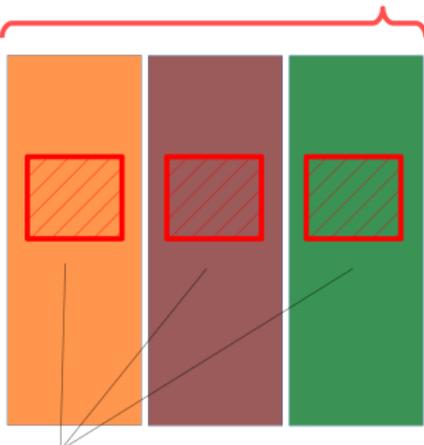
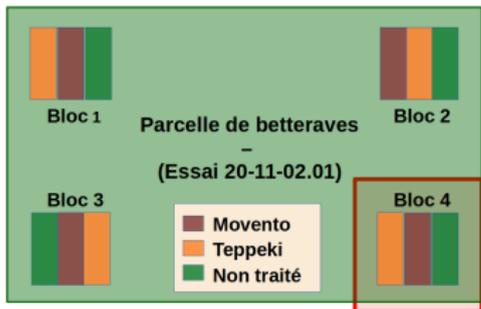


Data description

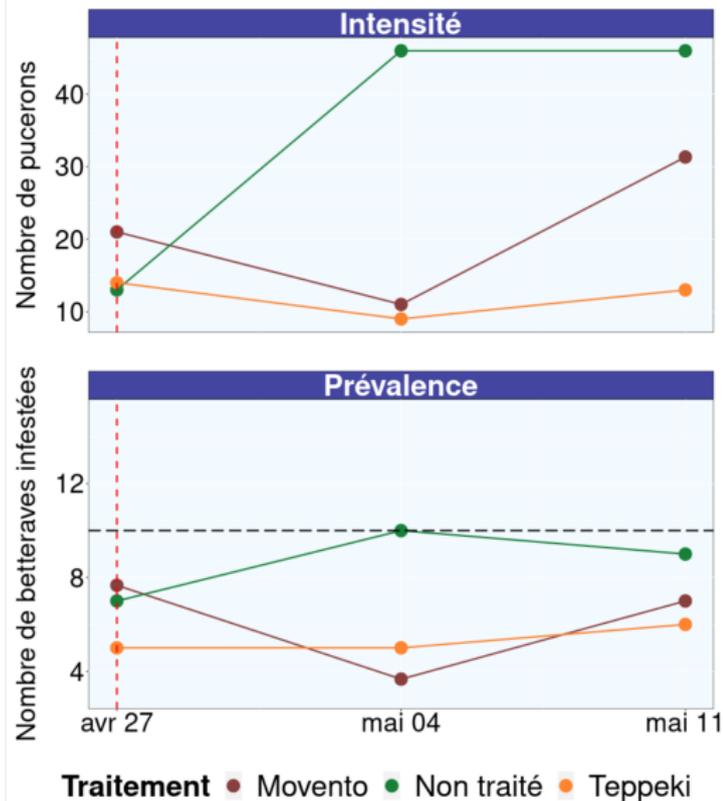


Echantillon de N = 10 betteraves

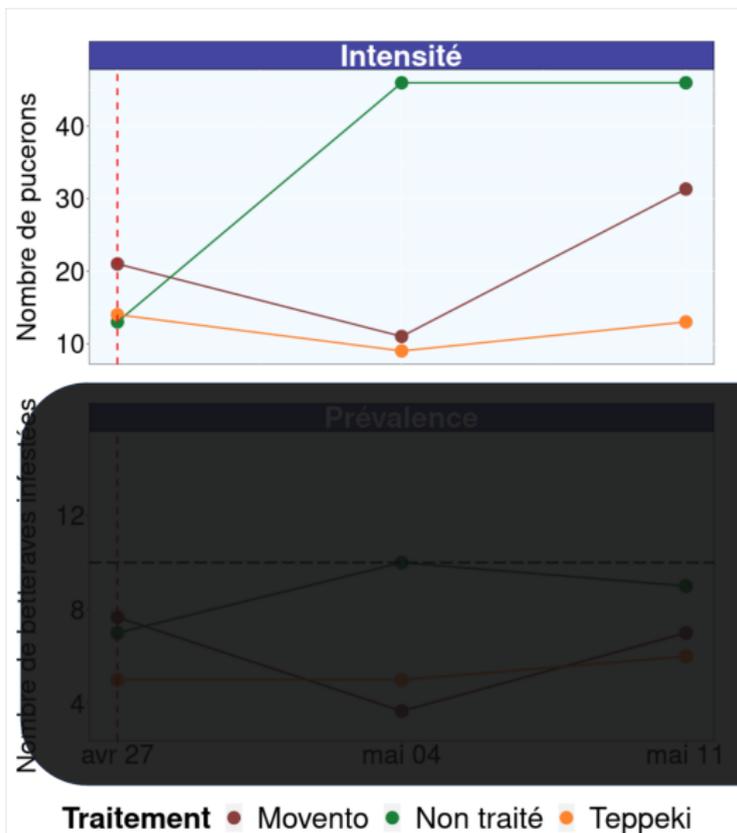
Data description



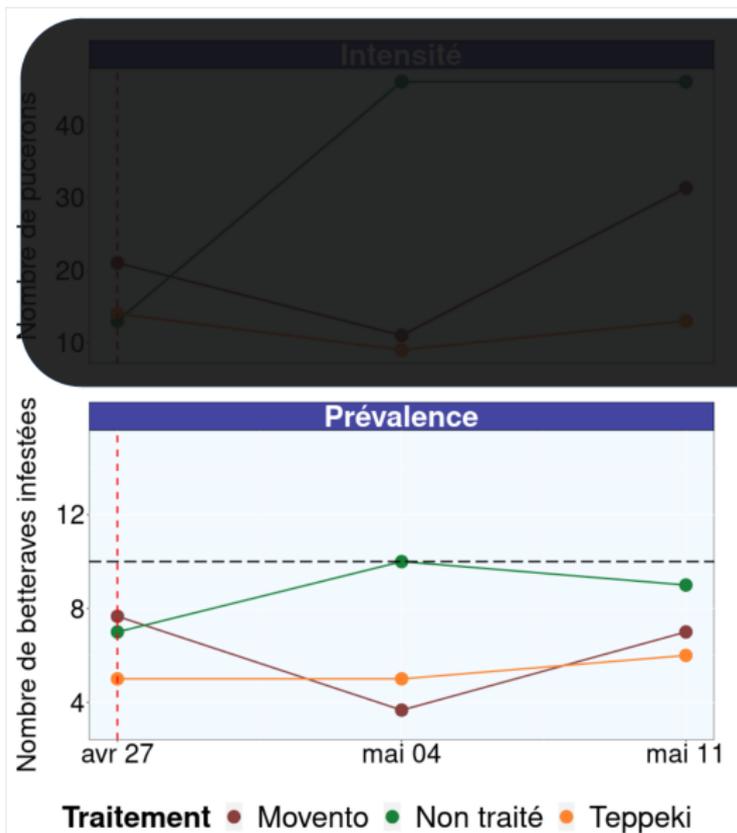
Echantillon de N = 10 betteraves



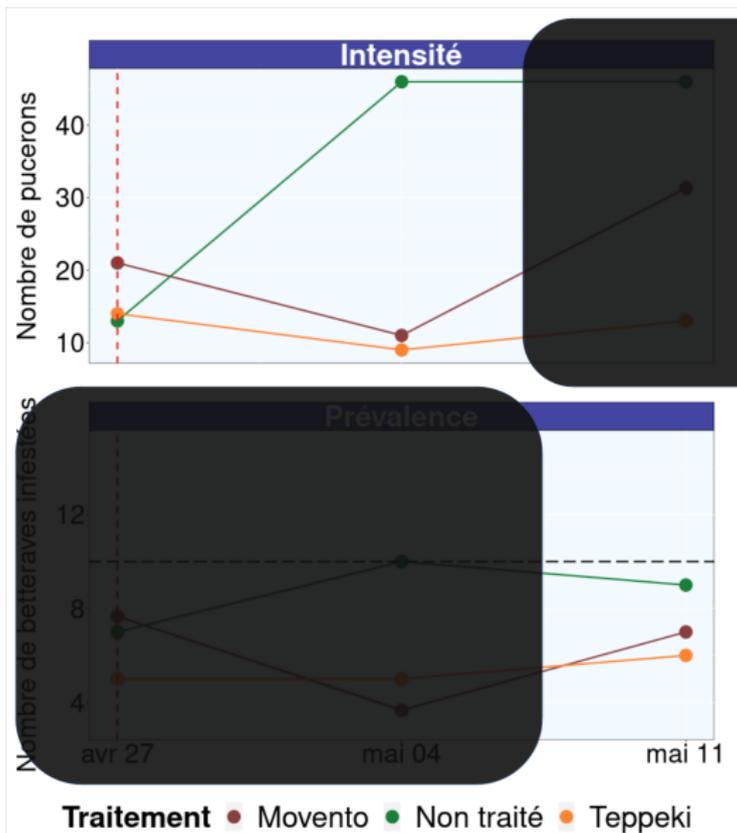
Data description



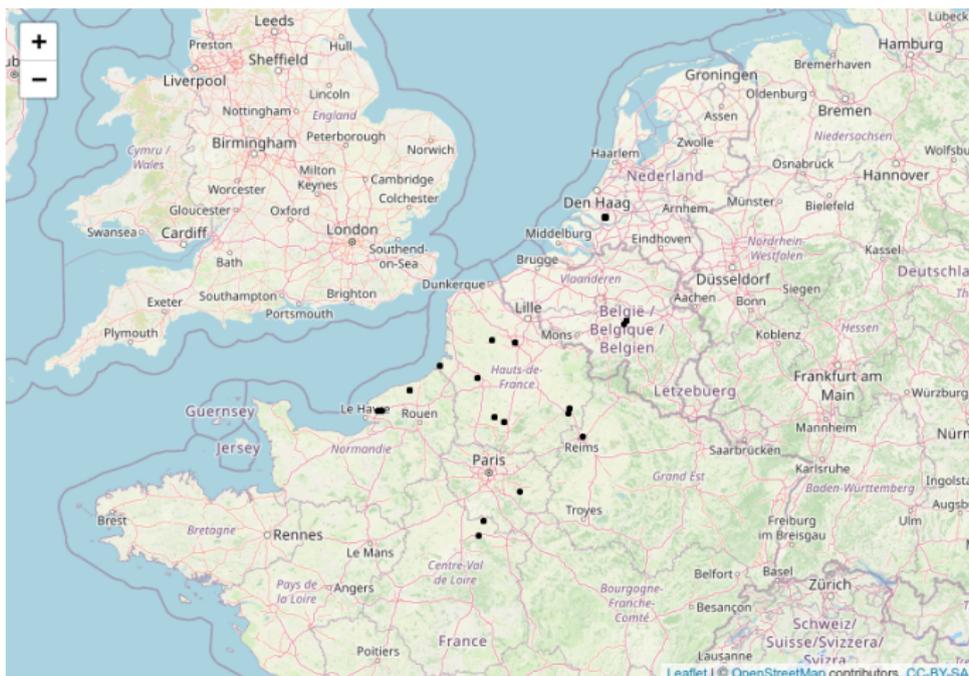
Data description



Data description



Carte des essais



- 1 Contexte
- 2 Description des données
- 3 Modélisation**
- 4 Conclusion

- Y_{ijkt} : Intensité

- Y_{ijkt} : Intensité

- i : essai

- j : traitement

- k : bloc

- t : temps

- Y_{ijkt} : Intensité
- Z_{ijkt} : Prévalence

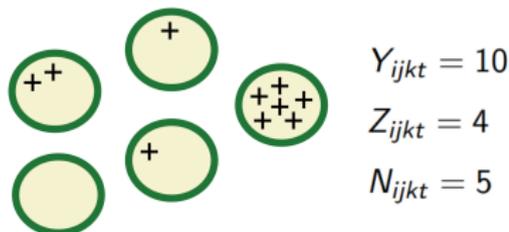
- i : essai
- j : traitement
- k : bloc
- t : temps

- Y_{ijkt} : Intensité
- Z_{ijkt} : Prévalence
- N_{ijkt} : Nombre de betterave dans l'échantillon

- i : essai
- j : traitement
- k : bloc
- t : temps

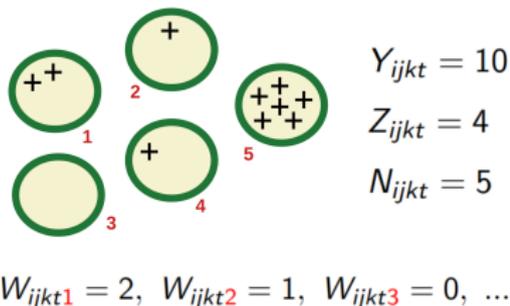
- Y_{ijkt} : Intensité
- Z_{ijkt} : Prévalence
- N_{ijkt} : Nombre de betterave dans l'échantillon

- i : essai
- j : traitement
- k : bloc
- t : temps



- Y_{ijkt} : Intensité
- Z_{ijkt} : Prévalence
- N_{ijkt} : Nombre de betterave dans l'échantillon

- i : essai
- j : traitement
- k : bloc
- t : temps



W_{ijkt} : Nombre de pucerons sur chaque betterave (**non observé**)

Notations

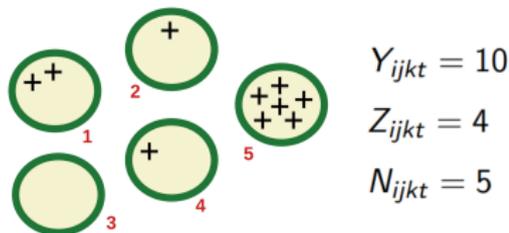
- Y_{ijkt} : Intensité
- Z_{ijkt} : Prévalence
- N_{ijkt} : Nombre de betterave dans l'échantillon

- i : essai

- j : traitement

- k : bloc

- t : temps



$$W_{ijkt1} = 2, W_{ijkt2} = 1, W_{ijkt3} = 0, \dots$$

W_{ijkt_s} : Nombre de pucerons sur chaque betterave (**non observé**)

Par construction : $Y_{ijkt} = \sum_{s=1}^{N_{ijkt}} W_{ijkt_s}$ et $Z_{ijkt} = \sum_{s=1}^{N_{ijkt}} \mathbf{1}_{W_{ijkt_s} > 0}$

$$W_{ijkt} \sim \mathcal{P}(\lambda_{ijkt})$$

$$\log \lambda_{ijkt} = \alpha_0 + \beta_{0i} + \gamma_{0j} + (\alpha_1 + \gamma_{1j}) X_{it} + u_{ij} + \epsilon_{ijkt}$$

$$W_{ijkt} \sim \mathcal{P}(\lambda_{ijkt})$$

$$\log \lambda_{ijkt} = \alpha_0 + \beta_{0i} + \gamma_{0j} + (\alpha_1 + \gamma_{1j}) X_{it} + u_{ij} + \epsilon_{ijkt}$$

Avec $\gamma_{00} = \gamma_{10} = 0$ (identifiabilité) et

$$\begin{cases} \beta_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\ u_{ij} \sim \mathcal{N}(0, \chi^2) \\ \epsilon_{ijkt} \sim \mathcal{N}(0, \eta^2) \end{cases}$$

$$W_{ijkt} \sim \mathcal{P}(\lambda_{ijkt})$$

$$\log \lambda_{ijkt} = \alpha_0 + \beta_{0i} + \gamma_{0j} + (\alpha_1 + \gamma_{1j}) X_{it} + u_{ij} + \epsilon_{ijkt}$$

Avec $\gamma_{00} = \gamma_{10} = 0$ (identifiabilité) et

$$\begin{cases} \beta_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\ u_{ij} \sim \mathcal{N}(0, \chi^2) \\ \epsilon_{ijkt} \sim \mathcal{N}(0, \eta^2) \end{cases}$$

$$Y_{ijkt} = \sum_{s=1}^{N_i} W_{ijkt} \quad \Rightarrow \quad Y_{ijkt} \sim \mathcal{P}(N_i \lambda_{ijkt})$$

$$Z_{ijkt} = \sum_{s=1}^{N_i} \mathbf{1}_{W_{ijkt} > 0} \quad \Rightarrow \quad Z_{ijkt} \sim \mathcal{B}(N_i, 1 - \exp(-\lambda_{ijkt}))$$

$$W_{ijkt} \sim \mathcal{P}(\lambda_{ijkt})$$

$$\log \lambda_{ijkt} = \alpha_0 + \beta_{0i} + \gamma_{0j} + (\alpha_1 + \gamma_{1j}) X_{it} + u_{ij} + \epsilon_{ijkt}$$

Avec $\gamma_{00} = \gamma_{10} = 0$ (identifiabilité) et

$$\begin{cases} \beta_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\ u_{ij} \sim \mathcal{N}(0, \chi^2) \\ \epsilon_{ijkt} \sim \mathcal{N}(0, \eta^2) \end{cases}$$

$$Y_{ijkt} = \sum_{s=1}^{N_i} W_{ijkt} \quad \Rightarrow \quad Y_{ijkt} \sim \mathcal{P}(N_i \lambda_{ijkt})$$

$$Z_{ijkt} = \sum_{s=1}^{N_i} \mathbf{1}_{W_{ijkt} > 0} \quad \Rightarrow \quad Z_{ijkt} \sim \mathcal{B}(N_i, 1 - \exp(-\lambda_{ijkt}))$$

Efficacité des traitements : $\text{Ef}_{jt} = \left(1 - \exp(\gamma_{0j} + \gamma_{1j} \times X_t)\right) \times 100$

⇒ Inférence bayésienne avec rjags

Lois à priori :

- $\alpha, \beta, \gamma : \mathcal{N}(0, 10^3)$
- $\sigma^2, \chi^2, \eta^2 : \mathcal{U}(0, 10)$

Nombre d'itérations : 200000

⇒ Calcul de la moyenne a posteriori de Ef_{jt} et de son interval de crédibilité

```
for (i in 1:n_obs){  
  Y[i] ~ dpois(N[i] * lambda[i])  
  Z[i] ~ dbinom(pi[i], N[i])  
  
  log(lambda[i]) = beta0[ESSAI[i]] + gamma0[TRAITEMENT[i]] +  
                  (alpha1 + gamma1[TRAITEMENT[i]]) * TEMPS[i] +  
                  u[ESSAI_TRAITEMENT[i]] + epsilon[i]  
  
  pi[i] = 1 - exp(- lambda[i])  
  epsilon[i] ~ dnorm(0, inv_eta)  
}
```

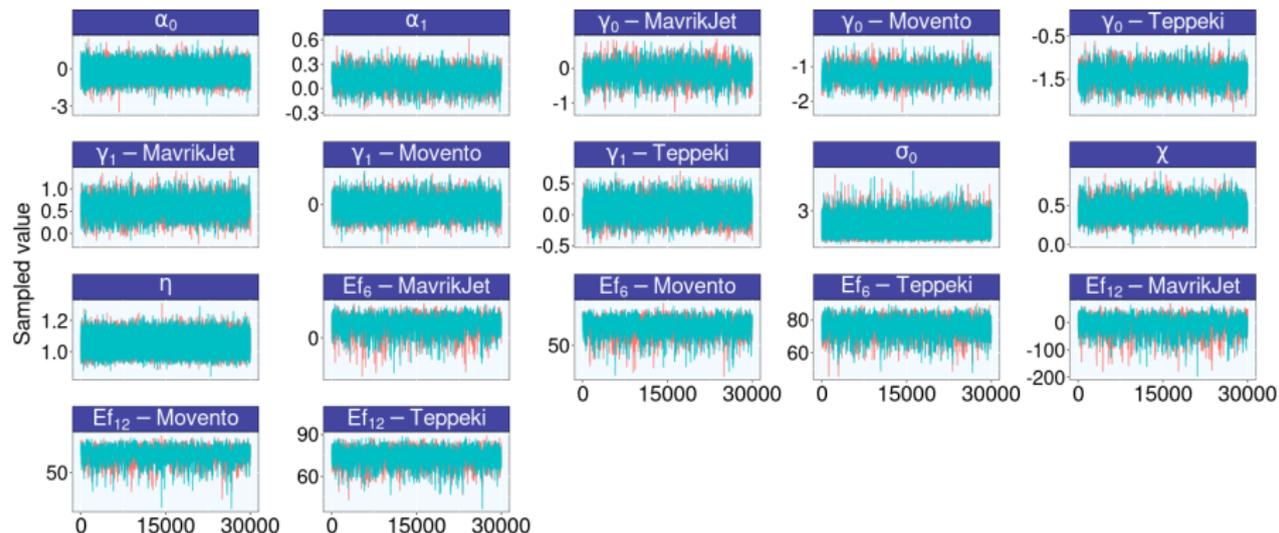
100% Y — 0%Z : intensité disponible dans tous les essais et prévalence non disponible

50% Y — 50%Z : intensité disponible dans la moitié des essais et prévalence disponible dans l'autre moitié

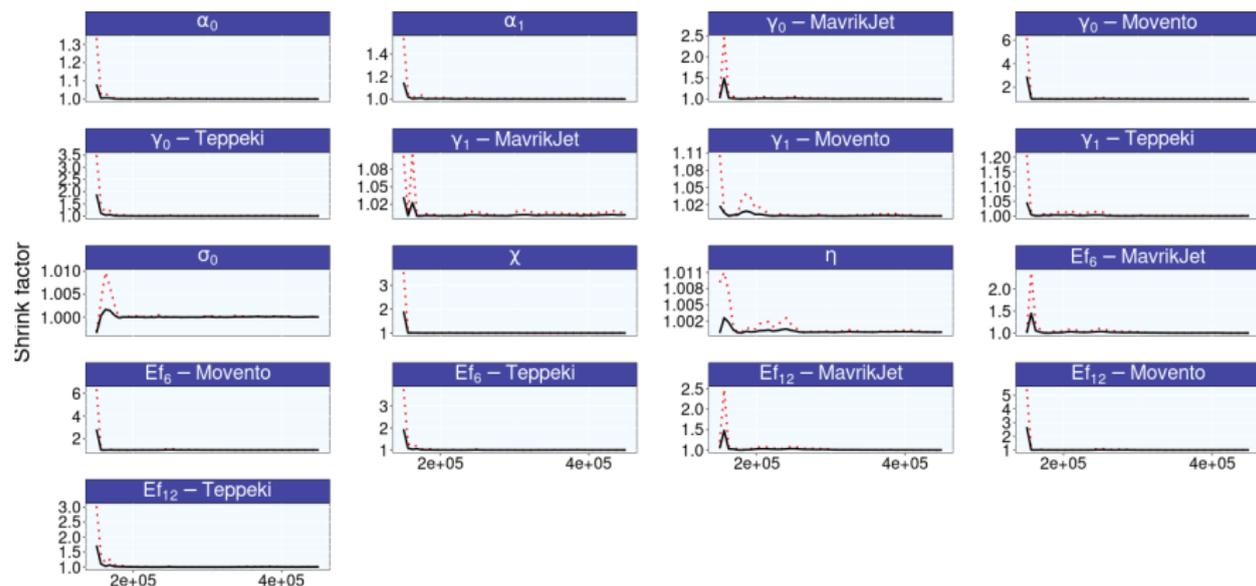
50% Y — 0%Z : intensité disponible dans la moitié des essais et prévalence non disponible

0% Y — 50%Z : intensité non disponible et prévalence disponible dans la moitié des essais

Vérification de la convergence - Inspection des chaînes



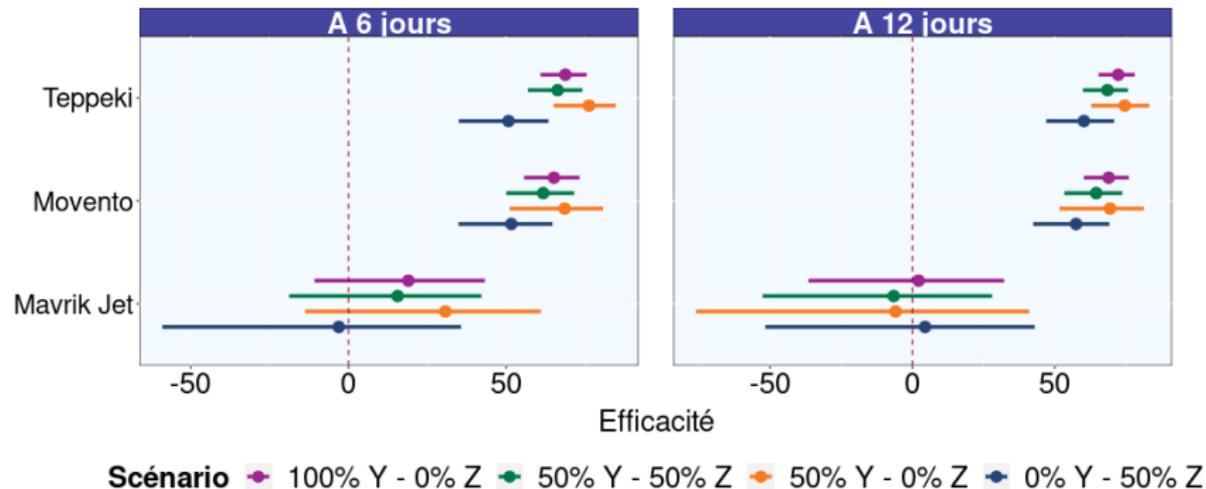
Vérification de la convergence - Diagnostic de Gelman–Rubin



Last iteration in chain

name ● 97.5% — median

Résultats sur les données réelles



- Résultats cohérents entre les différents scénarios
- Tailles des intervalles de confiance :

$$100\% Y - 0\% Z < 50\% Y - 50\% Z < 50\% Y - 0\% Z < 0\% Y - 50\% Z$$

Simulations selon les paramètres inférés sur les données réelles

974 jeux de données simulés, pour différents nombre d'essais (20, 40 et 80)

4 nouveaux scénarios :

100% W : W disponible dans tous les essais

100% Y — 100% Z : intensité et prévalence disponible dans tous les essais

0% Y — 100% Z : intensité non disponible et prévalence disponible dans tous les essais

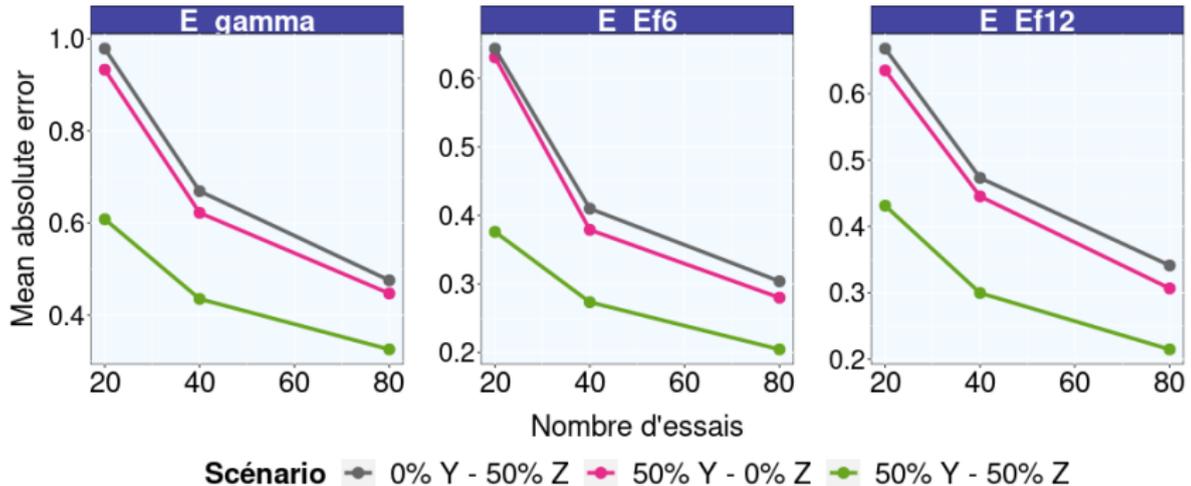
50% Y — 100% Z : intensité disponible dans la moitié des essais et prévalence disponible dans tous les essais

Métriques d'évaluations :

$$E_{\gamma} = \frac{1}{2 \times 3} \sum_{j=2}^4 \left(\frac{|\gamma_{0j} - \hat{\gamma}_{0j}|}{|\gamma_{0j}|} + \frac{|\gamma_{1j} - \hat{\gamma}_{1j}|}{|\gamma_{1j}|} \right)$$

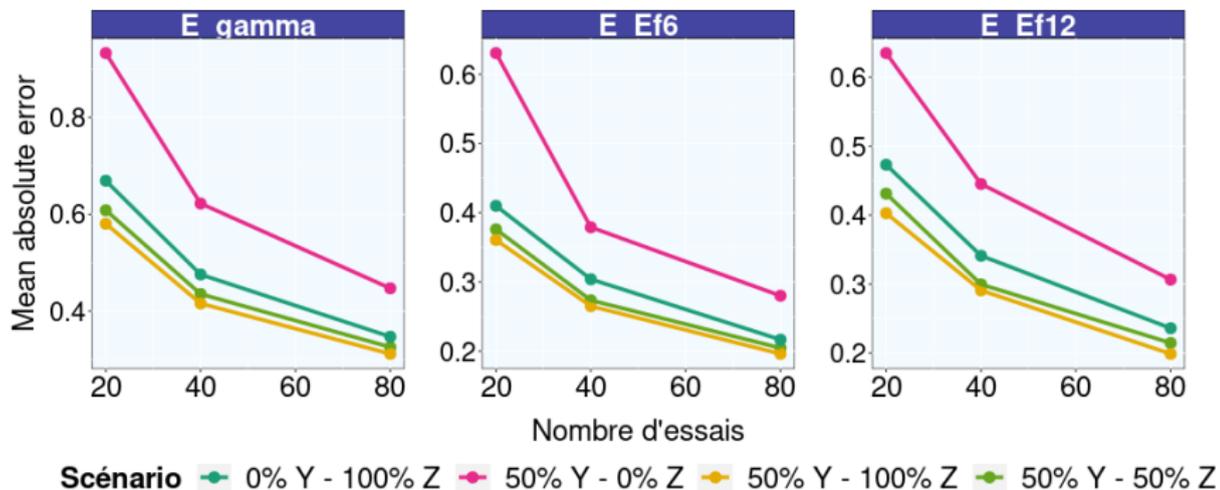
$$E_{E_{f_t}} = \frac{1}{3} \sum_{j=2}^4 \frac{|E_{f_{jt}} - \hat{E}_{f_{jt}}|}{|E_{f_{jt}}|}$$

Intérêt de combiner les essais avec intensité et les essais avec prévalence



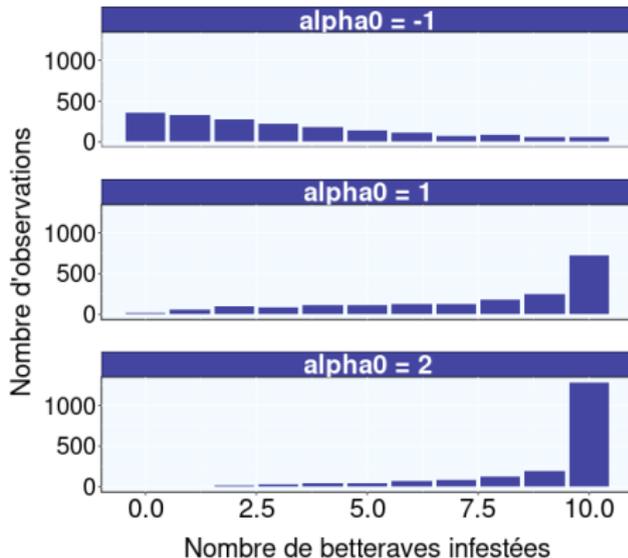
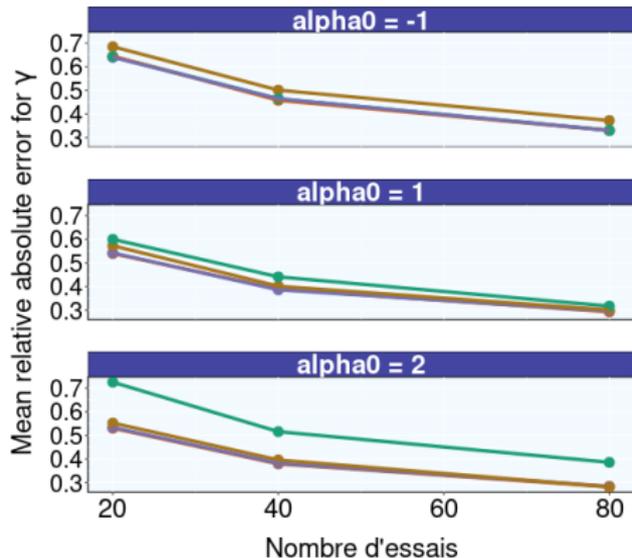
⇒ Gain significatif en combinant les essais avec intensité et les essais avec prévalence

Intérêt d'ajouter l'intensité lorsque la prévalence est mesurée



⇒ Ajouter l'intensité à la prévalence permet d'améliorer la précision de l'estimation

Résultats



Scénario 0% Y - 100% Z 100% W 100% Y - 0% Z 100% Y - 100% Z

⇒ Prévalence non informative lorsque la pression en pucerons est très forte

- 1 Contexte
- 2 Description des données
- 3 Modélisation
- 4 Conclusion**

- modèle qui permet de combiner l'intensité et la prévalence
- intensité plus informative que la prévalence
- applications pratiques



The screenshot shows a webpage header for ITB (Institut Technique de la Betterave) with the text 'Recherche et expertise au service de la filière betteravière'. Below the header, the article title is 'Expérimentation : est-il préférable de compter les pucerons ou les plantes infestées ?' with a date of '29/03/2023'. There is a small image of a leaf with aphids. To the right, there is a sidebar with the text 'Ça pourrait aussi vous intéresser' and two dates: '09/01/2024 Comité technique ITB Somme / Oise' and '17/01/2024 SDV au comité technique du Nord - Pas-de-Calais'. At the bottom of the sidebar, it says 'LE CAHIER TECHNIQUE' and 'Article - 12/12/2023 Le Cahier Technique n°1174'.