

Modélisation par EDP de la dynamique saisonnnière de populations de carabes au sein d'une mosaïque agricole

S. Poggi N. Parisey Y. Bourhis

INRA - UMR IGEPP
Équipe Écologie et Génétique des Insectes



Atelier ModStatSP
5 et 6 décembre 2013 - Avignon

CONTEXTE

Objectifs :

- ▶ prédire la dynamique de *P. cupreus* durant sa saison d'activité

CONTEXTE

Objectifs :

- ▶ prédire la dynamique de *P. cupreus* durant sa saison d'activité
- ▶ étudier l'impact des occupations du sol sur ces effectifs

CONTEXTE

Objectifs :

- ▶ prédire la dynamique de *P. cupreus* durant sa saison d'activité
- ▶ étudier l'impact des occupations du sol sur ces effectifs

Nous disposons de données de comptage de *Carabidae*¹ :

- ▶ échantillonnées dans la Zone Atelier Armorique (ZAA)

1. projet ANR Landscaphid

CONTEXTE

Objectifs :

- ▶ prédire la dynamique de *P. cupreus* durant sa saison d'activité
- ▶ étudier l'impact des occupations du sol sur ces effectifs

Nous disposons de données de comptage de *Carabidae*¹ :

- ▶ échantillonnées dans la Zone Atelier Armorique (ZAA)
- ▶ collectées via 75 pots Barber entre mai et juillet 2010

1. projet ANR Landscaphid

CONTEXTE

Objectifs :

- ▶ prédire la dynamique de *P. cupreus* durant sa saison d'activité
- ▶ étudier l'impact des occupations du sol sur ces effectifs

Nous disposons de données de comptage de *Carabidae*¹ :

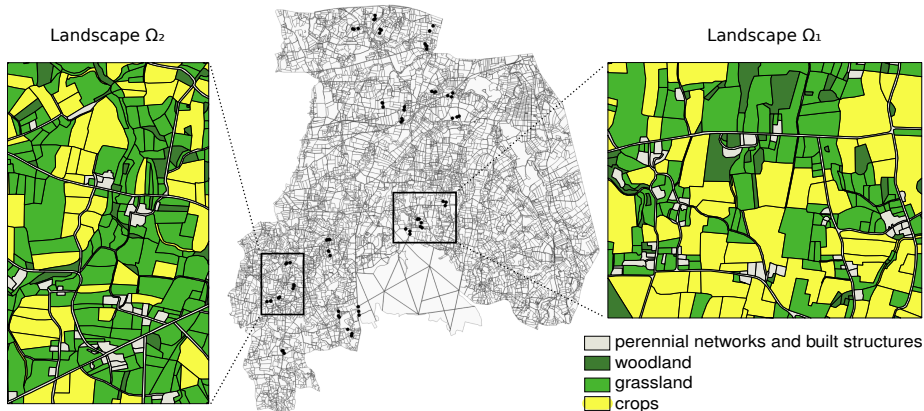
- ▶ échantillonnées dans la Zone Atelier Armorique (ZAA)
- ▶ collectées via 75 pots Barber entre mai et juillet 2010
- ▶ géolocalisées sur la ZAA

1. projet ANR Landscaphid



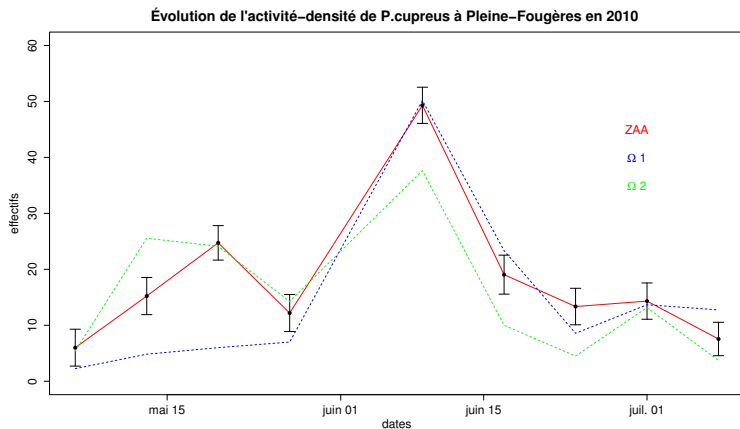
STRUCTURATION DES DONNÉES

- Dans l'espace : 2 domaines comprenant 12 pièges chacun



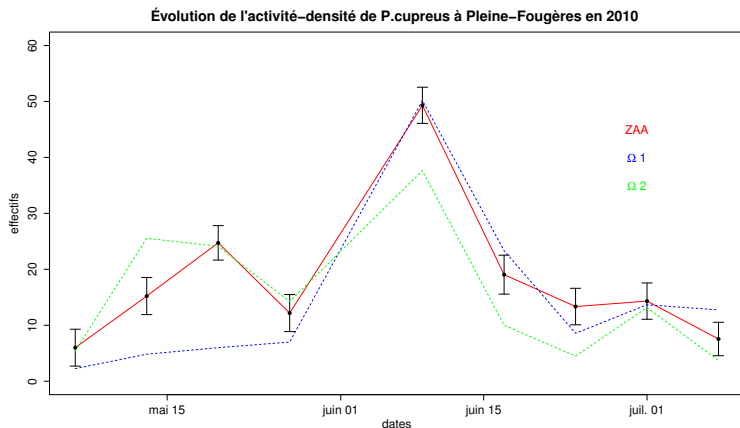
STRUCTURATION DES DONNÉES

- Dans le temps



STRUCTURATION DES DONNÉES

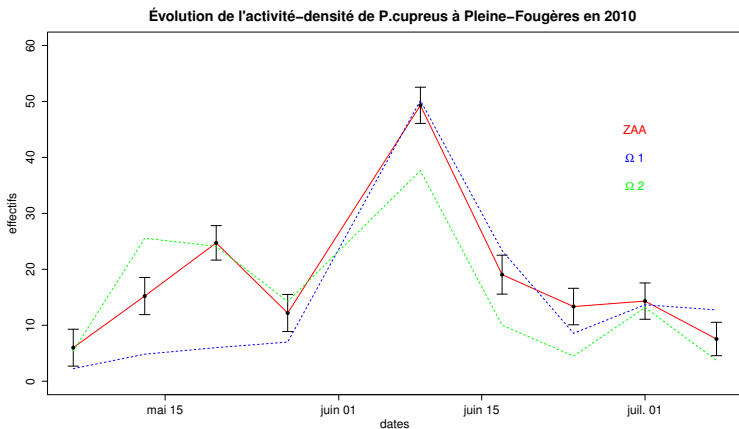
- ▶ Dans le temps
 - 9 dates d'échantillonnage (6 mai - 8 juillet)





STRUCTURATION DES DONNÉES

- ▶ Dans le temps
 - 9 dates d'échantillonnage (6 mai - 8 juillet)
 - Pièges ouverts pendant 5 jours avant chaque échantillonnage



MODÈLE DE RÉACTION-DIFFUSION

Équation de Fisher-KPP :

$$\frac{\partial P_{x,y,t}}{\partial t} = D\nabla^2 P_{x,y,t} + P_{x,y,t} \left(\frac{4\alpha_{x,y}}{K} \left(1 - \frac{P_{x,y,t}}{K} \right) - \mu t \right)$$

$$P_{x,y,0} = P_0 \mathbb{1}_{xy}$$

MODÈLE DE RÉACTION-DIFFUSION

Équation de Fisher-KPP :

$$\frac{\partial P_{x,y,t}}{\partial t} = D \nabla^2 P_{x,y,t} + P_{x,y,t} \left(\frac{4\alpha_{x,y}}{K} \left(1 - \frac{P_{x,y,t}}{K} \right) - \mu t \right)$$

$$P_{x,y,0} = P_0 \mathbb{1}_{xy}$$

Avec 4 paramètres :

- ▶ une constante de **diffusion**
- ▶ un **taux de croissance** spatialisé
- ▶ une **capacité d'accueil**
- ▶ un **taux de mortalité**

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
- ▶ les prairies

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
- ▶ les prairies
- ▶ les linéaires semi-naturels

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
 - ▶ les prairies
 - ▶ les linéaires semi-naturels
- } valeurs à estimer

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
 - ▶ les prairies
 - ▶ les linéaires semi-naturels
 - ▶ les bois, routes, voies d'eaux, bâtis où la croissance du carabe est supposée nulle → valeur fixée à zéro
- } valeurs à estimer

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
 - ▶ les prairies
 - ▶ les linéaires semi-naturels
 - ▶ les bois, routes, voies d'eaux, bâtis où la croissance du carabe est supposée nulle → valeur fixée à zéro
- } valeurs à estimer

L'occupation du sol est :

- ▶ définie sur des données géomatiques continues (vectorielles)

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
 - ▶ les prairies
 - ▶ les linéaires semi-naturels
 - ▶ les bois, routes, voies d'eaux, bâtis où la croissance du carabe est supposée nulle → valeur fixée à zéro
- } valeurs à estimer

L'occupation du sol est :

- ▶ définie sur des données géomatiques continues (vectorielles)
- ▶ approximées pour les besoins de la simulation par un paysage matriciel (raster de cellules $10 \times 10 \text{ m}^2$).

SPATIALISATION DU MODÈLE

Le taux de croissance α est dépendant de l'occupation du sol et catégorisé selon 4 grandes classes :

- ▶ les cultures (maïs, blé)
 - ▶ les prairies
 - ▶ les linéaires semi-naturels
 - ▶ les bois, routes, voies d'eaux, bâtis où la croissance du carabe est supposée nulle → valeur fixée à zéro
- } valeurs à estimer

L'occupation du sol est :

- ▶ définie sur des données géomatiques continues (vectorielles)
- ▶ approximées pour les besoins de la simulation par un paysage matriciel (raster de cellules $10 \times 10 \text{ m}^2$).

N.B. : Comptages effectués uniquement au sein des cultures.

SCHÉMA NUMÉRIQUE

- ▶ Modèle résolu en différence finies
- ▶ Par *Operator Splitting*
 - réaction → RK4
 - diffusion → *Alternate Direction Implicite method (ADI)*²
- ▶ Conditions initiales interpolées (IDW) à partir des données de la première semaine

2. Peaceman & Rachford, 1955.

SCHÉMA NUMÉRIQUE

- ▶ Modèle résolu en différence finies
- ▶ Par *Operator Splitting*
 - réaction \rightarrow RK4
 - diffusion \rightarrow *Alternate Direction Implicite method (ADI)*²
- ▶ Conditions initiales interpolées (IDW) à partir des données de la première semaine

Temps de calcul :

- ▶ d'une saison d'activité ($1000 \Delta t$)
- ▶ sur Ω_1 et Ω_2 (resp. 159×189 et 205×112 cellules)

2. Peaceman & Rachford, 1955.

SCHÉMA NUMÉRIQUE

- ▶ Modèle résolu en différence finies
- ▶ Par *Operator Splitting*
 - réaction \rightarrow RK4
 - diffusion \rightarrow *Alternate Direction Implicite method (ADI)*²
- ▶ Conditions initiales interpolées (IDW) à partir des données de la première semaine

Temps de calcul :

- ▶ d'une saison d'activité ($1000 \Delta t$)
- ▶ sur Ω_1 et Ω_2 (resp. 159×189 et 205×112 cellules) } $t_c \simeq 45$ s.

2. Peaceman & Rachford, 1955.

SCHÉMA NUMÉRIQUE

- ▶ Modèle résolu en différence finies
- ▶ Par *Operator Splitting*
 - réaction → RK4
 - diffusion → *Alternate Direction Implicit method (ADI)*²
- ▶ Conditions initiales interpolées (IDW) à partir des données de la première semaine

Temps de calcul :

- ▶ d'une saison d'activité ($1000 \Delta t$)
- ▶ sur Ω_1 et Ω_2 (resp. 159×189 et 205×112 cellules) } $t_c \simeq 45$ s.

N.B. : Une version ReacTran³ (méthode des lignes) est disponible ($t_c \simeq 55$ s.)

2. Peaceman & Rachford, 1955.

3. Soetaert *et al.*, 2012.

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations
- ▶ 2 domaines \times 12 pièges \times 9 dates = 216 observations

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations
- ▶ 2 domaines \times 12 pièges \times 9 dates = 216 observations
- ▶ Utilisation des moindres carrés ordinaires

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations
- ▶ 2 domaines \times 12 pièges \times 9 dates = 216 observations
- ▶ Utilisation des moindres carrés ordinaires

Réalisé via *nlopt*⁴

- ▶ Algorithme multi-local : *Multi Level Single Linkage*

4. S.G. Johnson, The NLOpt nonlinear-optimization package.

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations
- ▶ 2 domaines \times 12 pièges \times 9 dates = 216 observations
- ▶ Utilisation des moindres carrés ordinaires

Réalisé via *nlopt*⁴

- ▶ Algorithme multi-local : *Multi Level Single Linkage*
- ▶ Algorithme local : *Nelder-Mead Simplex*

4. S.G. Johnson, The NLOpt nonlinear-optimization package.

ESTIMATION DES PARAMÈTRES

- ▶ Objectif : minimiser les écarts entre prédictions et observations
- ▶ 2 domaines \times 12 pièges \times 9 dates = 216 observations
- ▶ Utilisation des moindres carrés ordinaires

Réalisé via *nlopt*⁴

- ▶ Algorithme multi-local : *Multi Level Single Linkage*
- ▶ Algorithme local : *Nelder-Mead Simplex*

Valeurs estimées des paramètres :

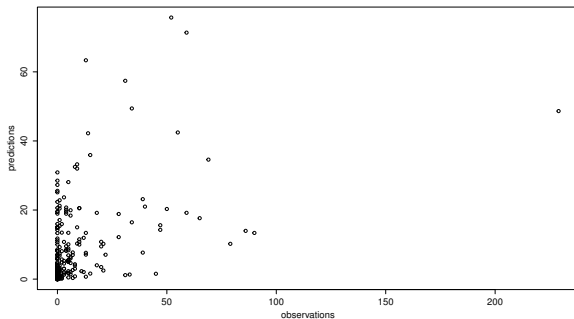
Paramètres	D	K	μ	$\alpha_{culture}$	$\alpha_{prairies}$	$\alpha_{seminat.}$
Valeurs	$2,07 \cdot 10^{-4}$	871,4	0,255	305,4	45,6	64,9
Unité	$m^2 \cdot s^{-1}$	-	s^{-1}	s^{-1}	s^{-1}	s^{-1}

4. S.G. Johnson, The NLOpt nonlinear-optimization package.

SIMULATION AVEC LE MODÈLE PARAMÉTRÉ

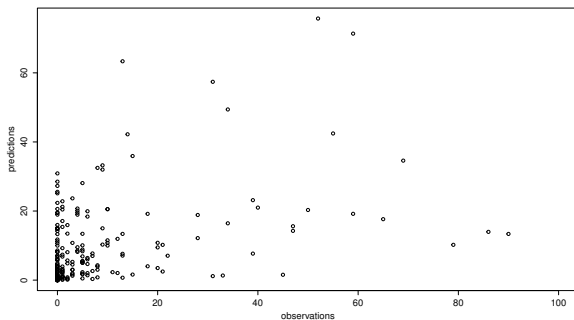
CRITIQUES

Analyse des résidus :



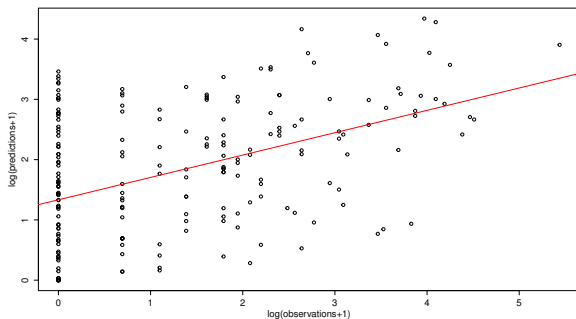
CRITIQUES

Analyse des résidus :



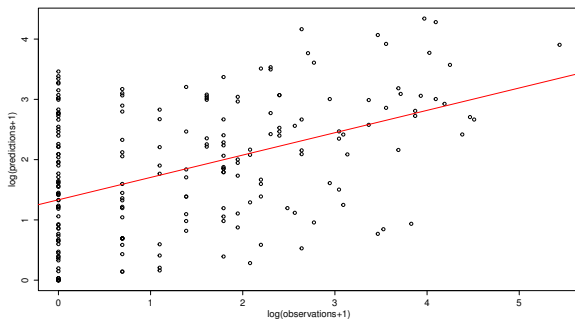
CRITIQUES

Analyse des résidus :



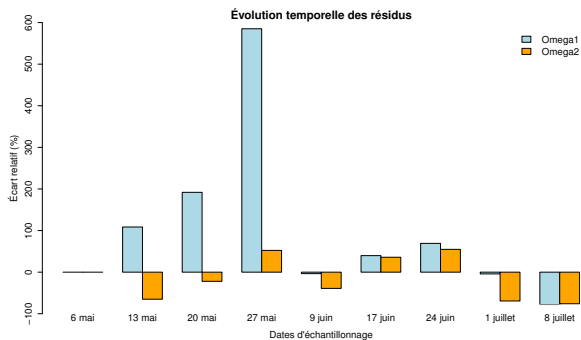
CRITIQUES

Analyse des résidus :

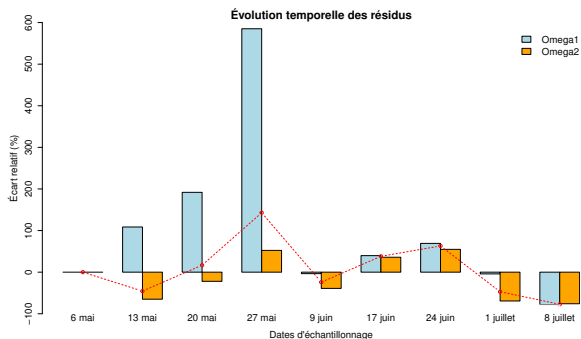


- ▶ Mauvaises prédictions en général ($R^2 < 0.2$), surtout aux niveaux des observations de pièges vides.

CRITIQUES

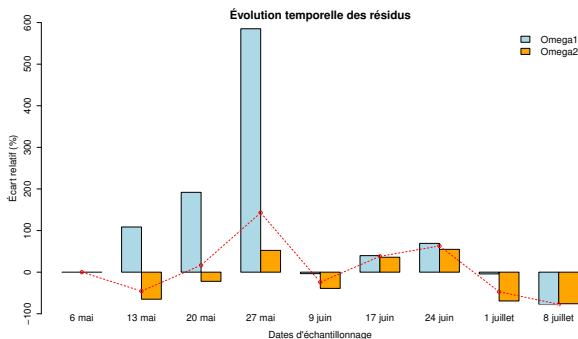


CRITIQUES



- ▶ sur-prédictions sur Ω_1
- ▶ sous-prédictions sur Ω_2

CRITIQUES



- ▶ sur-prédictions sur Ω_1
- ▶ sous-prédictions sur Ω_2
- ▶ autocorrélation spatiale⁵
- ▶ pas d'autocol. temporelle⁵

5. Moran's I et Box-Ljung test

QUE SONT NOS DONNÉES ?

- ▶ Les observations sont ici considérées comme des abondances immédiates.

QUE SONT NOS DONNÉES ?

- ▶ ~~Les observations sont ici considérées comme des abondances immédiates.~~
- ▶ Elles sont des intégrales en temps d'abondances (→ temps d'ouverture du piège $\tau = 5 \text{ jours}$)

$$y_{x,t} = \int_{t-\tau}^t u_x$$

avec y l'observation et u l'abondance immédiate.

QUE SONT NOS DONNÉES ?

- ▶ ~~Les observations sont ici considérées comme des abondances immédiates.~~
- ▶ Elles sont des intégrales en temps d'abondances (→ temps d'ouverture du piège $\tau = 5\text{jours}$)

$$y_{x,t} = \int_{t-\tau}^t u_x$$

avec y l'observation et u l'abondance immédiate.

- ▶ Les observations sont les réalisations d'un processus de piégeage que l'on doit identifier afin de mieux les prendre en compte.

QUE SONT NOS DONNÉES ?

- ▶ ~~Les observations sont ici considérées comme des abondances immédiates.~~
- ▶ Elles sont des intégrales en temps d'abondances (→ temps d'ouverture du piège $\tau = 5\text{jours}$)

$$y_{x,t} = \int_{t-\tau}^t u_x$$

avec y l'observation et u l'abondance immédiate.

- ▶ Les observations sont les réalisations d'un processus de piégeage que l'on doit identifier afin de mieux les prendre en compte.

Besoin d'un modèle d'observation !

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁶.

- ▶ On considère le modèle non linéaire généralisé :

$$Y|X \sim \mathcal{L}\{\eta(X, \beta), \sigma\}$$

6. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁶.

- ▶ On considère le modèle non linéaire généralisé :

$$Y|X \sim \mathcal{L}\{\eta(X, \beta), \sigma\}$$

Il relie la var. réponse Y à un vecteur de var. explicatives X par une loi \mathcal{L} centrée sur les réalisations de $\eta(X, \beta)$ et de paramètre de forme σ (facultatif).

6. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁶.

- ▶ On considère le modèle non linéaire généralisé :

$$Y|X \sim \mathcal{L}\{\eta(X, \beta), \sigma\}$$

Il relie la var. réponse Y à un vecteur de var. explicatives X par une loi \mathcal{L} centrée sur les réalisations de $\eta(X, \beta)$ et de paramètre de forme σ (facultatif).

- ▶ L'objectif est d'estimer $\theta = (\beta, \sigma)$, les paramètres du modèle et le paramètre de forme de la loi.

6. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁶.

- ▶ On considère le modèle non linéaire généralisé :

$$Y|X \sim \mathcal{L}\{\eta(X, \beta), \sigma\}$$

Il relie la var. réponse Y à un vecteur de var. explicatives X par une loi \mathcal{L} centrée sur les réalisations de $\eta(X, \beta)$ et de paramètre de forme σ (facultatif).

- ▶ L'objectif est d'estimer $\theta = (\beta, \sigma)$, les paramètres du modèle et le paramètre de forme de la loi.
- ▶ On veut donc inférer $\theta|x, \mathbf{y}$, réalisations de X et de Y :

$$p(\theta|x, \mathbf{y}) = \frac{p(\mathbf{y}|\theta, x)\pi(\theta)}{p(\mathbf{y}|x)}$$

avec $\pi(\theta)$ la distribution *a priori* des paramètres.

6. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁷.

Comment calculer $p(\mathbf{y}|\theta, \mathbf{x})$?

- ▶ Modèle d'observation, cas d'observations poissonniennes :

7. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁷.

Comment calculer $p(\mathbf{y}|\theta, \mathbf{x})$?

- ▶ Modèle d'observation, cas d'observations poissonniennes :

Var. réponses : y_i

Var. explicatives : t_i, \mathbf{x}_i

Les mesures y_i sont des réalisations de Y :

$$Y_i|t_i, \mathbf{x}_i, u \sim \text{Poisson} \left(\alpha \int_{\tau_i} u(t_i, \mathbf{x}_i) d\mathbf{x}_i \right)$$

7. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁷.

Comment calculer $p(\mathbf{y}|\theta, \mathbf{x})$?

- Modèle d'observation, cas d'observations poissonniennes :

Var. réponses : y_i

Var. explicatives : t_i, \mathbf{x}_i

Les mesures y_i sont des réalisations de Y :

$$Y_i | t_i, \mathbf{x}_i, u \sim \text{Poisson} \left(\alpha \int_{\tau_i} u(t_i, \mathbf{x}_i) d\mathbf{x}_i \right)$$

Rappel :

$$P(Z = z) = e^{-\lambda} \frac{\lambda^z}{z!}$$

7. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁷.

Comment calculer $p(\mathbf{y}|\theta, \mathbf{x})$?

- Modèle d'observation, cas d'observations poissonniennes :

Var. réponses : y_i

Var. explicatives : t_i, \mathbf{x}_i

Les mesures y_i sont des réalisations de Y :

$$Y_i | t_i, \mathbf{x}_i, u \sim \text{Poisson} \left(\alpha \int_{\tau_i} u(t_i, \mathbf{x}_i) d\mathbf{x}_i \right)$$

Rappel :

$$P(Z = z) = e^{-\lambda} \frac{\lambda^z}{z!}$$

On en déduit la vraisemblance :

$$p(\mathbf{y}|\theta, \mathbf{t}, \mathbf{x}) = \prod_{i=1}^I e^{-\eta\{(t_i, \mathbf{x}_i), \theta\}} \frac{\eta\{(t_i, \mathbf{x}_i), \theta\}^{y_i}}{y_i!}$$

7. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

ESTIMATION BAYÉSIENNE DE PARAMÈTRES⁷.

Comment calculer $p(\mathbf{y}|\theta, \mathbf{x})$?

- Modèle d'observation, cas d'observations poissonniennes :

Var. réponses : y_i

Var. explicatives : t_i, \mathbf{x}_i

Les mesures y_i sont des réalisations de Y :

$$Y_i | t_i, \mathbf{x}_i, u \sim \text{Poisson} \left(\alpha \int_{\tau_i} u(t_i, \mathbf{x}_i) d\mathbf{x}_i \right)$$

Rappel :

$$P(Z = z) = e^{-\lambda} \frac{\lambda^z}{z!}$$

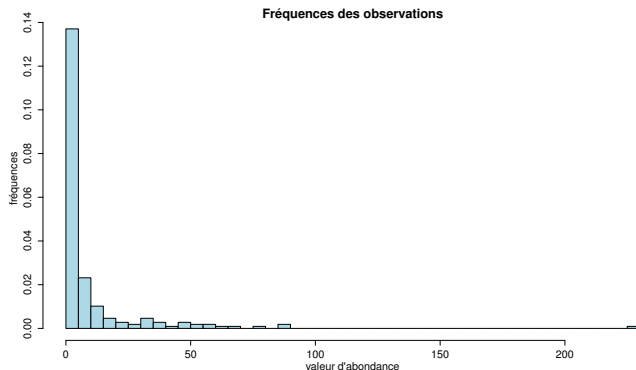
On en déduit la vraisemblance :

$$p(\mathbf{y}|\theta, \mathbf{t}, \mathbf{x}) = \prod_{i=1}^I e^{-\eta\{(t_i, \mathbf{x}_i), \theta\}} \frac{\eta\{(t_i, \mathbf{x}_i), \theta\}^{y_i}}{y_i!}$$

$$\text{où } \eta\{(t, \mathbf{x}), \theta\} = \alpha \int_{t-\tau}^t u_{\mathbf{x}} d\mathbf{x}$$

7. L. Roques, Modèle de réaction-diffusion pour l'écologie spatiale, 2013

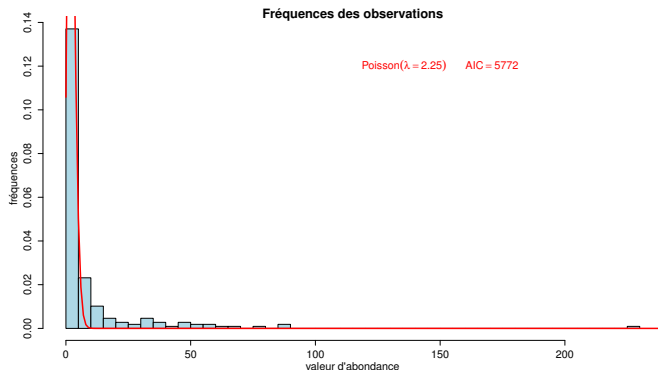
LOIS CANDIDATES⁸



- Données de comptage -> *a priori* Poisson

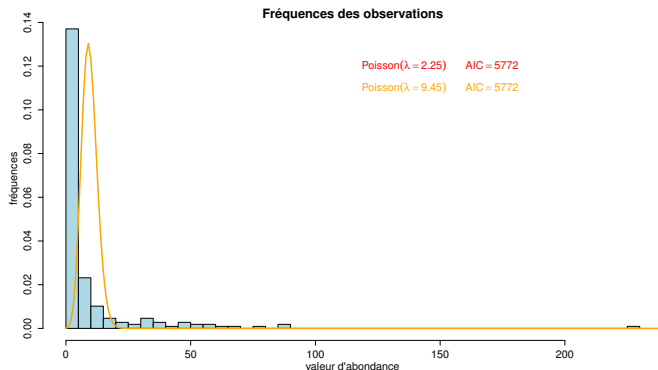
8. ajustées par estimation du maximum de vraisemblance via *bbmle*.

LOIS CANDIDATES⁸



- Données de comptage -> *a priori* Poisson

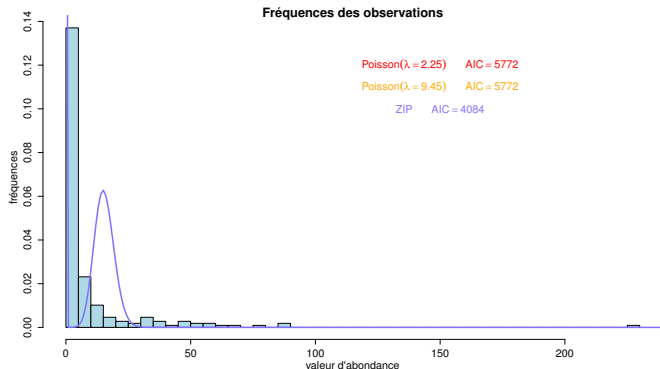
8. ajustées par estimation du maximum de vraisemblance via *bbmle*.

LOIS CANDIDATES⁸

- Données de comptage -> *a priori* Poisson

8. ajustées par estimation du maximum de vraisemblance via *bbmle*.

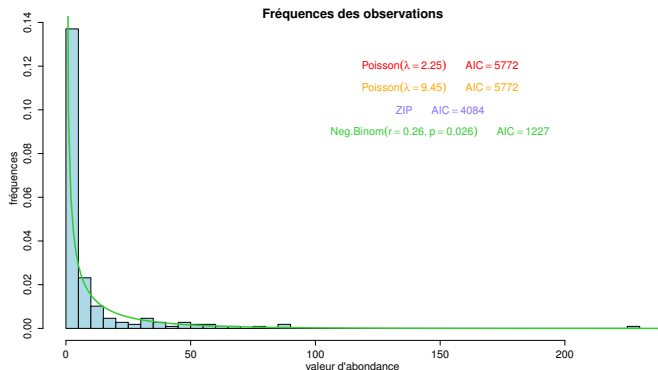
LOIS CANDIDATES⁸



- ▶ Données de comptage -> *a priori* Poisson
- ▶ Sur-dispersée (avec beaucoup de zéros)
 - Zero Inflated Poisson

8. ajustées par estimation du maximum de vraisemblance via *bbmle*.

LOIS CANDIDATES⁸



- ▶ Données de comptage -> *a priori* Poisson
- ▶ Sur-dispersée (avec beaucoup de zéros)
 - Zero Inflated Poisson
 - Binomiale négative

8. ajustées par estimation du maximum de vraisemblance via *bbmle*.

QUESTIONS

- ▶ Comment estimer α ?

$$\eta\{(t, \mathbf{x}), \theta\} = \alpha \int_{t-\tau}^t u_x dx$$

QUESTIONS

- ▶ Comment estimer α ?

$$\eta\{(t, \mathbf{x}), \theta\} = \alpha \int_{t-\tau}^t u_x dx$$

- ▶ $\alpha \rightarrow$ fonctionnement du piège ?

QUESTIONS

- ▶ Comment estimer α ?

$$\eta\{(t, \mathbf{x}), \theta\} = \alpha \int_{t-\tau}^t u_x d_x$$

- ▶ $\alpha \rightarrow$ fonctionnement du piège ?
- ▶ Que penser de ce type de modèle mécaniste :

$$\begin{cases} \frac{\partial P}{\partial t} = f(P) \\ \frac{\partial B}{\partial t} = \alpha P \end{cases}$$

$$\alpha = \mathbb{1}_{x_i, y_i, t_i}$$