

Estimation de la dynamique spatio-temporelle d'une métapopulation sur réseau. Application au colza

Florence Carpentier, Fabien Laroche, Catherine Laredo, Sylvie Huet

► **To cite this version:**

Florence Carpentier, Fabien Laroche, Catherine Laredo, Sylvie Huet. Estimation de la dynamique spatio-temporelle d'une métapopulation sur réseau. Application au colza. Colloque Ecologie 2010, Sep 2010, Montpellier, France. 28 diapos. hal-02816409

HAL Id: hal-02816409

<https://hal.inrae.fr/hal-02816409>

Submitted on 6 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.

Estimation de la dynamique spatio-temporelle d'une métapopulation sur réseau. Application au colza

Florence Carpentier, Fabien Laroche, Catherine Larédo, Sylvie Huet

MIA INRA JOUY

3 septembre 2010

- Agro-ecosystème

- ▶ Contrôler les flux de gènes (pollen, graines)
- ▶ Risques de contaminations :
 - ★ sauvages ⇒ cultivées : qualité agronomique
 - ★ cultivées ⇒ sauvages : ex OGM (transgènes)

- Colza

- ▶ très répandu en Europe
- ▶ grande capacité de dispersion (pollen)
- ▶ repousse dans les champs (banques de graines, repousses)
- ▶ habitats (semi-)naturels : bordures champs et routes

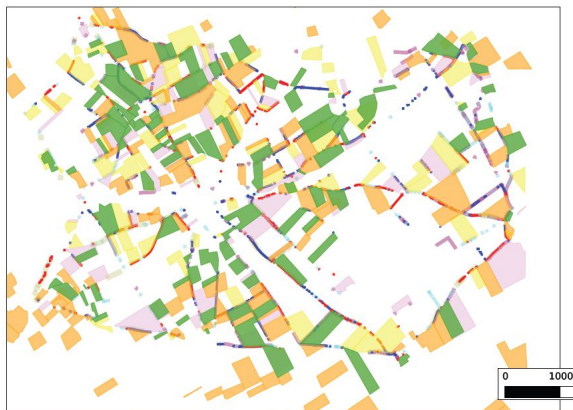
⇒ Risques de contamination élevés

Populations férales :
populations bordures de routes et chemins

Dynamique des populations férales :

- Déterminer les sources :
 - ▶ Immigration : champs, camions
 - ▶ Auto-recrutement : populations, banques de graines
- Capacité de persistance
- Evaluer rôle de ces populations
relais ou réservoir de gènes cultivés (OGM)

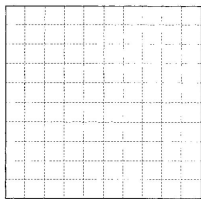
Particularités des populations férales



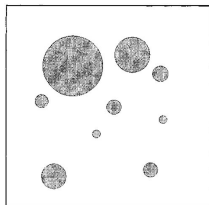
- Délimitation difficile
 - ▶ densité très variable
 - ▶ variable dans le temps
- Réseau routier
Habitat fragmenté
- Champs
 - ▶ source importante
 - ▶ varie (espace et temps)

Métapopulation sur réseau

Modèles spatialement explicites en écologie (Hanski et al. 98)

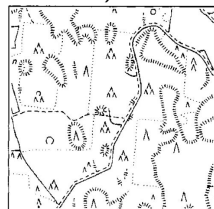


Théorique



Métapopulation

ensemble de patch favorables
différents degrés d'isolement



Paysage

Métapopulation sur Réseau :

Chaque route décomposée en segments

- régulier (3m)
- orienté (2 bords de routes)



Modéliser les "populations" de **chaque segment**.

Les données initiales

- Aire étudiée :
 - ▶ 42 km² centré sur Selommes (Loir et Cher)
 - ▶ 90 routes ~ 70 000 segments de 3 m
- 7 années consécutives d'observations (2000-2006)

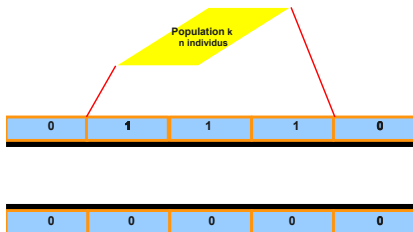
Chaque année, sur l'ensemble du réseau routier sont observés :

- Populations en été
 - ▶ ensemble d'individus séparés par <10m
 - ▶ coordonnées GPS
 - ▶ effectifs
- Champs
 - ▶ Position
 - ▶ Entre 80 et 100 champs (11 à 14% de l'aire étudiée)

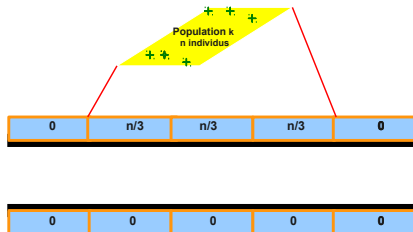
Création du jeu de données projection sur les segments

Populations

Présence-Absence

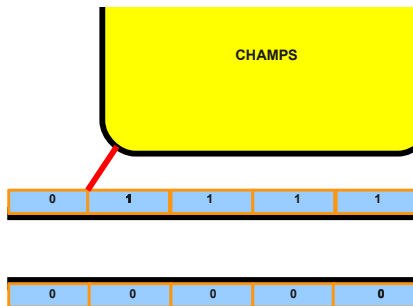


Densité



Création du jeu de données projection sur les segments

Champs codés en présence-absence :



● Notations

- ▶ $Z_{i,n}$, la présence ou l'absence de population sur le segment i à l'année n
- ▶ Z_n , l'ensemble des observations sur l'année n

Le modèle Présence-Absence

- Notations

- ▶ $Z_{i,n}$, la présence ou l'absence de population sur le segment i à l'année n
- ▶ Z_n , l'ensemble des observations sur l'année n

- Modélisation de la dépendance temporelle

- ▶ modèle auto-régressif

$$\Pr(Z_1 = z_1, \dots, Z_N = z_N | Z_0 = z_0) = \prod_{n=1}^N \Pr(Z_n = z_n | Z_{n-1} = z_{n-1})$$

- Notations

- ▶ $Z_{i,n}$, la présence ou l'absence de population sur le segment i à l'année n
- ▶ Z_n , l'ensemble des observations sur l'année n

- Modélisation de la dépendance temporelle

- ▶ modèle auto-régressif

$$\Pr(Z_1 = z_1, \dots, Z_N = z_N | Z_0 = z_0) = \prod_{n=1}^N \Pr(Z_n = z_n | Z_{n-1} = z_{n-1})$$

- Modélisation de la dépendance spatiale :

- ▶ Indépendance spatiale des réalisations

$$\Pr(Z_n = z_n | Z_{n-1} = z_{n-1}) = \prod_i \Pr(Z_{n,i} = z_{n,i} | Z_{n-1} = z_{n-1})$$

- Notations

- ▶ $Z_{i,n}$, la présence ou l'absence de population sur le segment i à l'année n
- ▶ Z_n , l'ensemble des observations sur l'année n

- Modélisation de la dépendance temporelle

- ▶ modèle auto-régressif

$$\Pr(Z_1 = z_1, \dots, Z_N = z_N | Z_0 = z_0) = \prod_{n=1}^N \Pr(Z_n = z_n | Z_{n-1} = z_{n-1})$$

- Modélisation de la dépendance spatiale :

- ▶ Indépendance spatiale des réalisations

$$\Pr(Z_n = z_n | Z_{n-1} = z_{n-1}) = \prod_i \Pr(Z_{n,i} = z_{n,i} | Z_{n-1} = z_{n-1})$$

- ▶ Mais spatialement liées aux travers des covariables

$$(Z_{n,i} | Z_{n-1}) \sim \text{Bern}(p_{i,n})$$

$$\text{logit}(p_{i,n}) = \delta + \gamma/4 \sum_{k \in V(i)} Z_{n-1,k} + \epsilon/2 \sum_{k \in V'(i)} C_{n-1,k} + \zeta/2 \sum_{k \in V'(i)} C_{n,k}$$

Le modèle Présence-Absence

Plus précisément,
la probabilité présence/absence est
modélisée **linéairement** selon :

- **Population** année $n - 1$ dans
dans le voisinage

$$\gamma/4 \sum_{k \in V(i)} Z_{n-1,k}$$



Le modèle Présence-Absence

Plus précisément,
la probabilité présence/absence est
modélisée **linéairement** selon :

- **Population** année $n - 1$ dans
dans le voisinage

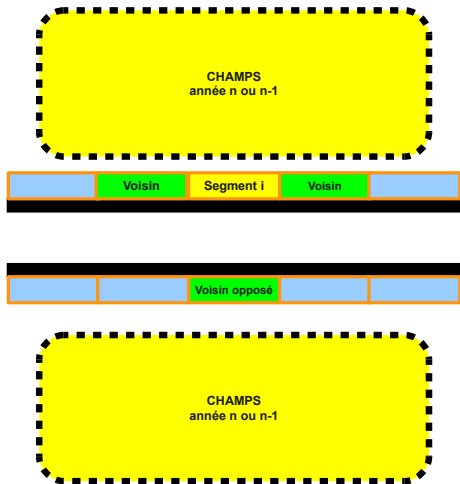
$$\gamma/4 \sum_{k \in V(i)} Z_{n-1,k}$$

- **Champ** année $n - 1$ dans le
voisinage

$$\epsilon/2 \sum_{k \in V'(i)} C_{n-1,k}$$

- Champ année n dans le
voisinage

$$\zeta/2 \sum_{k \in V'(i)} C_{n,k}$$



● Notations

- ▶ Y_n^* , ensemble des **observations non nulles**
- ▶ $Y_{i,n}^*$, densité de population sur le segment i à l'année n
- ▶ Y_n , ensemble des observations sur l'année n

● Hypothèses identiques

- ▶ Dépendance temporelle

$$\Pr(Y_1^* = y_1, \dots, Y_N^* = y_N | Y_0 = y_0) = \prod_{n=1}^N \Pr(Y_n^* = y_n | Y_{n-1}^* = y_{n-1})$$

- ▶ Indépendance spatiale

$$\Pr(Y_n^* = y_n | Y_{n-1}^* = y_{n-1}) = \prod_i \Pr(Y_{n,i}^* = y_{n,i} | Y_{n-1}^* = y_{n-1})$$

- Modèle

$$(Y_{n,i}^* | Y_{n-1}^*) \sim \log \mathcal{N}(\mu_{i,n}, \tau)$$
$$E(Y_{n,i}^* | Y_{n-1}^*) = \delta + \frac{\gamma}{4} \sum_{k \in V(i)} Y_{n-1,k} + \frac{\epsilon}{2} \sum_{k \in V'(i)} C_{n-1,k} + \frac{\zeta}{2} \sum_{k \in V'(i)} C_{n,k}$$

- Remarques :

- ▶ Y_{n-1} et non pas seulement Y_{n-1}^*
- ▶ Modèle linéaire **sans fonction de lien** ($\mathcal{U}(0, 50)$)
⇒ interprétation directe des paramètres
 - ★ γ taux de reproduction des populations férales
 - ★ ϵ contribution des champs n (perte lors des semis)
 - ★ ζ contribution des champs $n - 1$ (dispersion)
 - ★ δ apport extérieur
- ▶ Paramètres moyens et "efficaces"

Premiers Résultats

Données réduites :

- 10 routes parmi les 90 du site
- 5 000 segments sur les 70 000

Résultats obtenus par MCMC (jags)

Présence-Absence
 $\Pr(Z_{i,n} = 1|E)$

Apport Extérieur

$Z_{v(i),n-1} = 0, C_{v(i),n} = 0, C_{v(i),n-1} = 0$ 1.1% [1.0,1.2]

Populations n dans $V(i)$

$Z_{v(i),n-1} = 1, C_{v(i),n} = 0, C_{v(i),n-1} = 0$ 1.3% [1.0,1.6]

$Z_{v(i),n-1} = 4, C_{v(i),n} = 0, C_{v(i),n-1} = 0$ 2.5 % [1.17,4.9]

Champs n

$Z_{v(i),n-1} = 0, C_{v(i),n} = 1, C_{v(i),n-1} = 0$ 0.8% [0.6,1.0]

Champs $n - 1$

$Z_{v(i),n-1} = 0, C_{v(i),n} = 0, C_{v(i),n-1} = 1$ 3.6% [3.2,4.0]

Premiers Résultats

Données réduites :

- 10 routes parmi les 90 du site
- 5 000 segments sur les 70 000

Résultats obtenus par MCMC (jags)

	Présence-Absence $\Pr(Z_{i,n} = 1 E)$	Densité > 0 Contribution
Apport Extérieur		1.4 [1,2]
$Z_{v(i),n} = 0, C_{v(i),n} = 0, C_{v(i),n-1} = 0$	1.1% [1.0,1.2]	
Populations n dans $V(i)$		0.6 [0.02,3.7]
$Z_{v(i),n-1} = 1, C_{v(i),n} = 0, C_{v(i),n-1} = 0$	1.3% [1.0,1.6]	
$Z_{v(i),n-1} = 4, C_{v(i),n} = 0, C_{v(i),n-1} = 0$	2.5 % [1.17,4.9]	
Champs n		5.1 [2.3,9.1]
$Z_{v(i),n-1} = 0, C_{v(i),n} = 1, C_{v(i),n-1} = 0$	0.8% [0.6,1.0]	
Champs $n - 1$		0.27 [0.01,1.13]
$Z_{v(i),n-1} = 0, C_{v(i),n} = 0, C_{v(i),n-1} = 1$	3.6% [3.2,4.0]	

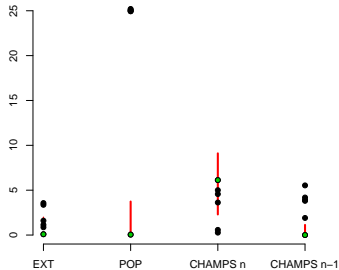
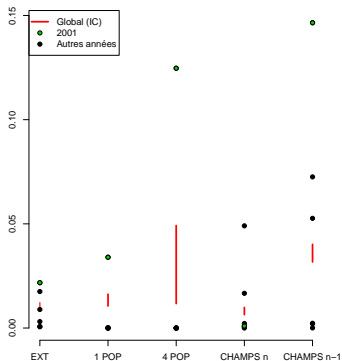
Introduction de variabilité temporelle

Nouveau modèle :

$$\delta_n + \frac{\gamma_n}{4} \sum_{k \in V(i)} Y_{n-1,k} + \frac{\epsilon_n}{2} \sum_{k \in V'(i)} C_{n-1,k} + \frac{\zeta_n}{2} \sum_{k \in V'(i)} C_{n,k}$$

Présence-Absence

Densité non nulle



- Prendre en compte les différentes routes

- ▶ $\delta_n + R_{r(i)} + \frac{\gamma_n}{4} \sum_{k \in V(i)} Y_{n-1,k} + \frac{\epsilon_n}{2} \sum_{k \in V'(i)} C_{n-1,k} + \frac{\zeta_n}{2} \sum_{k \in V'(i)} C_{n,k}$
- ▶ Présence-Absence : effet entre -2.8 et 1.8
- ▶ Densité non nulle : contribution entre 0.01 et 5.8

- Décomposer l'apport du voisinage

- ▶
 - ★ a0 : apport du segment i de l'année $n - 1$
 - ★ a1 : apport d'un segment adjacent $n - 1$
 - ★ aOp : apport du segment opposé
- ▶ Présence-Absence :
 - ★ a0 : 1.8 [-17,11]
 - ★ a1 : 1.1 [-9,11]
 - ★ aOp : 22 [-81,79] (problème d'estimation)
- ▶ Densité : pas suffisamment de données

Nouveau modèle : Densité sur tous les segments

Objectifs :

- Utiliser toutes les données simultanément
 - ▶ $Y_{n,i}$ et non plus sous-ensemble $Y_{n,i}^*$
 - ▶ modéliser $Y_{n,i}$ mais aussi covariables
- Prendre en compte les zéros issus :
 - ▶ environnements défavorables
 - ▶ absence de source

Nouveau modèle : Densité sur tous les segments

Objectifs :

- Utiliser toutes les données simultanément
 - ▶ $Y_{n,i}$ et non plus sous-ensemble $Y_{n,i}^*$
 - ▶ modéliser $Y_{n,i}$ mais aussi covariables
- Prendre en compte les zéros issus :
 - ▶ environnements défavorables
 - ▶ absence de source

Modèle proposé :

- Même hypothèses (dépendance temporelle, indépendance spatiale)
- $Y_{n,i}$ suivant une loi "zero-inflated"

Nouveau modèle : Densité sur tous les segments

Objectifs :

- Utiliser toutes les données simultanément
 - ▶ $Y_{n,i}$ et non plus sous-ensemble $Y_{n,i}^*$
 - ▶ modéliser $Y_{n,i}$ mais aussi covariables
- Prendre en compte les zéros issus :
 - ▶ environnements défavorables
 - ▶ absence de source

Modèle proposé :

- Même hypothèses (dépendance temporelle, indépendance spatiale)
- $Y_{n,i}$ suivant une loi "zero-inflated"
 - ▶ $\Pr(Y_{n,i} | Y_{n-1}) = (1 - p_i) \mathbb{1}_{(Y_{n,i}=0 | Y_{n-1})} + p_i U_{n,i}$

Nouveau modèle : Densité sur tous les segments

Objectifs :

- Utiliser toutes les données simultanément
 - ▶ $Y_{n,i}$ et non plus sous-ensemble $Y_{n,i}^*$
 - ▶ modéliser $Y_{n,i}$ mais aussi covariables
- Prendre en compte les zéros issus :
 - ▶ environnements défavorables
 - ▶ absence de source

Modèle proposé :

- Même hypothèses (dépendance temporelle, indépendance spatiale)
- $Y_{n,i}$ suivant une loi "zero-inflated"
 - ▶ $\Pr(Y_{n,i} | Y_{n-1}) = (1 - p_i) \mathbb{1}_{(Y_{n,i}=0 | Y_{n-1})} + p_i U_{n,i}$
 - ▶ $\text{logit}(p_i) = \delta + R_T(i)$

Nouveau modèle : Densité sur tous les segments

Objectifs :

- Utiliser toutes les données simultanément
 - ▶ $Y_{n,i}$ et non plus sous-ensemble $Y_{n,i}^*$
 - ▶ modéliser $Y_{n,i}$ mais aussi covariables
- Prendre en compte les zéros issus :
 - ▶ environnements défavorables
 - ▶ absence de source

Modèle proposé :

- Même hypothèses (dépendance temporelle, indépendance spatiale)
- $Y_{n,i}$ suivant une loi "zero-inflated"
 - ▶ $\Pr(Y_{n,i} | Y_{n-1}) = (1 - p_i) \mathbb{1}_{(Y_{n,i}=0 | Y_{n-1})} + p_i U_{n,i}$
 - ▶ $\text{logit}(p_i) = \delta + R_{T(i)}$
 - ▶ $U_{n,i} \sim \mathcal{N}_{>0}(m_{i,n}, \sigma^2)$ avec

$$m_{i,n} = \mu + \frac{\gamma}{4} \sum_{k \in V(i)} Y_{n-1,k} + \frac{\epsilon}{2} \sum_{k \in V'(i)} C_{n-1,k} + \frac{\zeta}{2} \sum_{k \in V'(i)} C_{n,k}$$

Conclusions et Perspectives

- Estimer les modèles sur l'ensemble des segments et des routes
Beaucoup de données nécessaires
 - ▶ très peu de populations sur le réseau
 - ★ entre 0.1% et 4% des segments sous-jeux de données
 - ★ entre 0.4 % et 7.3% des segments du jeux de données total

Conclusions et Perspectives

- Estimer les modèles sur l'ensemble des segments et des routes
Beaucoup de données nécessaires
 - ▶ très peu de populations sur le réseau
 - ★ entre 0.1% et 4% des segments sous-jeux de données
 - ★ entre 0.4 % et 7.3% des segments du jeux de données total
- Introduire des covariables
 - ▶ Modèle linéaire simple complexifiable
 - ▶ Introduire de l'information génétique partielle
 - ★ Pondérer les contributions par indices génétiques de similarité (coefficient d'apparatement)
 - ★ observé : indice estimé
 - ★ inconnu : tirage dans la loi empirique estimée sur l'ensemble des observations

Conclusions et Perspectives

- Estimer les modèles sur l'ensemble des segments et des routes
Beaucoup de données nécessaires
 - ▶ très peu de populations sur le réseau
 - ★ entre 0.1% et 4% des segments sous-jeux de données
 - ★ entre 0.4 % et 7.3% des segments du jeux de données total
- Introduire des covariables
 - ▶ Modèle linéaire simple complexifiable
 - ▶ Introduire de l'information génétique partielle
 - ★ Pondérer les contributions par indices génétiques de similarité (coefficient d'appariement)
 - ★ observé : indice estimé
 - ★ inconnu : tirage dans la loi empirique estimée sur l'ensemble des observations
- Comparer et valider les résultats obtenus avec les autres méthodes
 - ▶ mécaniste
 - ▶ génétique (données exhaustives)
 - ▶ estimés sur région plus fine sur une année