

Modélisation de la dynamique des adventices

Sebastian Le Coz

12/03/2019

Co-directeurs : Nathalie Peyrard, Pierre-Olivier Cheptou
Financement: Région Occitanie + ANR AGROBIOSE

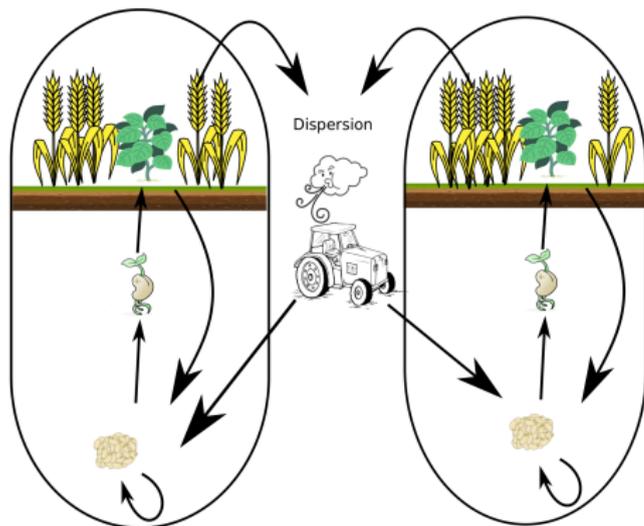


Une adventice ? C'est quoi ?



C'est une plante qui pousse dans un champ agricole sans y avoir été intentionnellement introduite, souvent appelée mauvaise herbe.

Cycle de vie d'une adventice



Processus

Colonisation
Dormance
Germination

Pertes de récoltes dues aux adventices

Les adventices sont en compétition avec la culture semée.

D'après Oerke et al (2012), la perte de récolte s'élève à

- 9.4 % pour le blé et l'orge,
- 9.3 % pour le maïs,
- 7 % pour la pomme de terre.

Pertes de récoltes dues aux adventices

Les adventices sont en compétition avec la culture semée.

D'après Oerke et al (2012), la perte de récolte s'élève à

- 9.4 % pour le blé et l'orge,
- 9.3 % pour le maïs,
- 7 % pour la pomme de terre.

Intérêts pour l'étude de la dynamique des adventices :

- + de rendement
- – de produits chimiques
- sauvegarde des abeilles sauvages

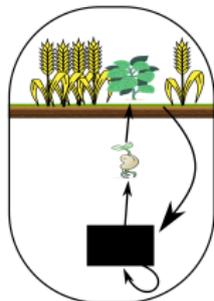
Modéliser la dynamique des adventices pourrait permettre de répondre à plusieurs questions :

- Quel est le processus le plus influent entre colonisation et dormance?
- Es-ce que la dynamique d'une adventice dépend de la culture mise en place ?
- Es-ce que l'influence des tracteurs est négligeable dans la dispersion des graines adventices ?

Modéliser la dynamique des adventices pourrait permettre de répondre à plusieurs questions :

- Quel est le processus le plus influent entre colonisation et dormance?
- Es-ce que la dynamique d'une adventice dépend de la culture mise en place ?
- Es-ce que l'influence des tracteurs est négligeable dans la dispersion des graines adventices ?

Difficulté de modélisation



La banque de graines n'est pas visible !

Etat de l'art

Modèles de métapopulation (Levins 1969)

Dynamique locale

- Extinction

Dynamique régionale

- Colonisation

Etat de l'art

Modèles de métapopulation (Levins 1969)

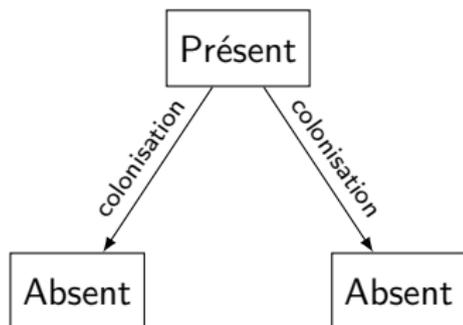
Dynamique locale

- Extinction

Dynamique régionale

- Colonisation

Exemple :



Etat de l'art

Modèles de métapopulation (Levins 1969)

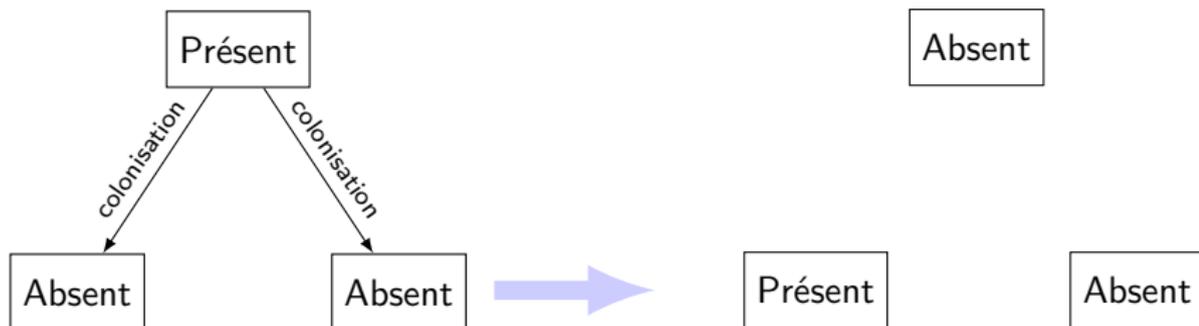
Dynamique locale

- Extinction

Dynamique régionale

- Colonisation

Exemple :



Problème des modèles de métapopulation appliqués aux adventices

Ils ne modélisent pas

La dormance

La dormance des graines peut conduire à déclarer de manière erronée l'extinction locale de l'espèce (Bullock et al 2006, Freckleton et al 2002).

Problème des modèles de métapopulation appliqués aux adventices

Ils ne modélisent pas

La dormance

La dormance des graines peut conduire à déclarer de manière erronée l'extinction locale de l'espèce (Bullock et al 2006, Freckleton et al 2002).

La colonisation

✓ Colonisation spatiale

× Colonisation extérieure

Problème des modèles de métapopulation appliqués aux adventices

Ils ne modélisent pas

La dormance

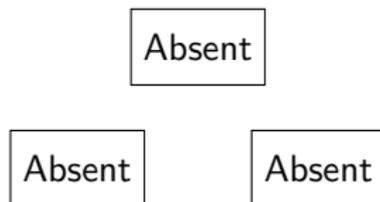
La dormance des graines peut conduire à déclarer de manière erronée l'extinction locale de l'espèce (Bullock et al 2006, Freckleton et al 2002).

La colonisation

✓ Colonisation spatiale

✗ Colonisation extérieure

Exemple :



Modèles intégrant le stade de dormance

Modèle	Col spatiale	Col extérieure	Survie de BG illimitée	BG cachée
Levin 1984	oui	non	oui	non
Venable 1988	oui	non	oui	non
Jarry 1995	non	non	oui	non
Maffei 2004	non	non	oui	oui
Mistro 2005	non	non	non	non
Garnier 2006	oui	oui	oui	non
David 2010	non	variable!	oui	oui
Lamy 2013	non	oui	oui	oui
Fréville 2013	non	oui	non	oui
Han 2014	non	non	oui	non
Borgy 2015	non	non	oui	oui
Manna 2017	oui 2 patches	oui	non	oui

Objectif : Modéliser la dynamique locale et régionale des adventices

Caractéristiques statistiques :

- Les populations en stade de dormance ne sont pas observées
- La flore levée et la banque de graines sont en classes d'abondance

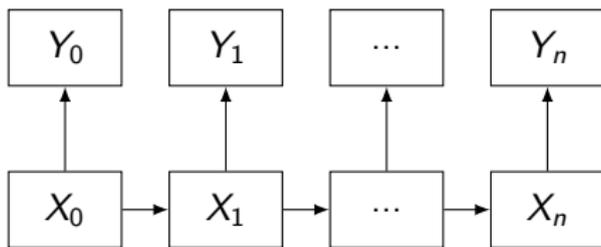
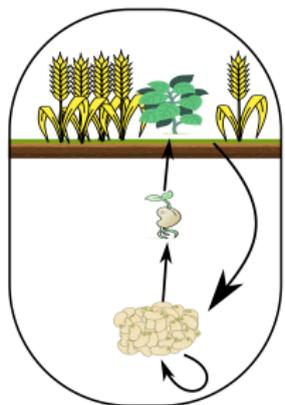
Caractéristiques biologiques :

- Processus de dormance
- Processus de colonisation spatiale
- Processus de colonisation extérieure
- Processus de germination
- Temps de dormance illimité.

- 1 Construction du graphe associé à la dynamique d'espèces avec stade caché (MHMM-DF).
- 2 Paramétrisation des lois du modèle.
- 3 Construction de l'algorithme d'estimation à l'aide du EM pour le MHMM-DF.
- 4 Résultats sur données simulées.
- 5 Analyse des données d'Epoisses.
- 6 Conclusion.
- 7 Perspectives.

Construction du graphe Multi Hidden Markov Model with data feedback (MHMM-DF)

Les chaînes de Markov cachées (HMM) appliquées aux adventices.



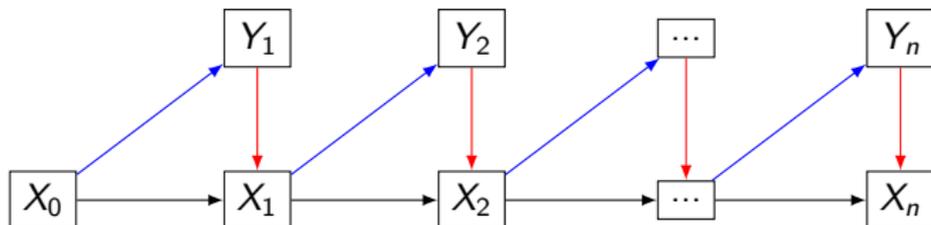
$(X_n)_{n \in \{0, \dots, N\}}$ correspondent à l'état de la banque de graines.

$(Y_n)_{n \in \{0, \dots, N\}}$ correspondent à l'état de la flore levée.

Problème ! Le HMM ne modélise pas la production de nouvelles graines entrant dans la banque de graines.

Le modèle de Borgy et al (2015)

Inclut l'influence de la population observable sur la population cachée (data feedback).



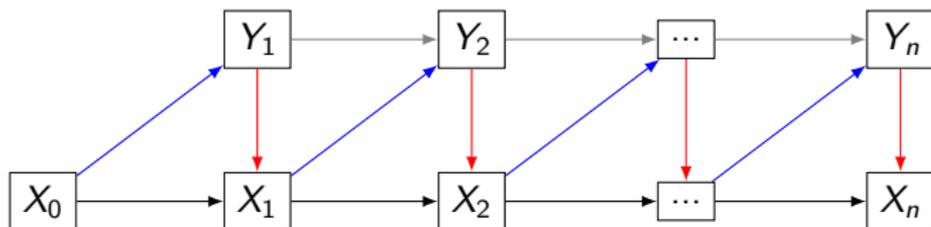
Le modèle ne modélise pas

- la colonisation spatiale,
- la survie des variables observées.

Cadre HMM-DF

La survie des variables observées

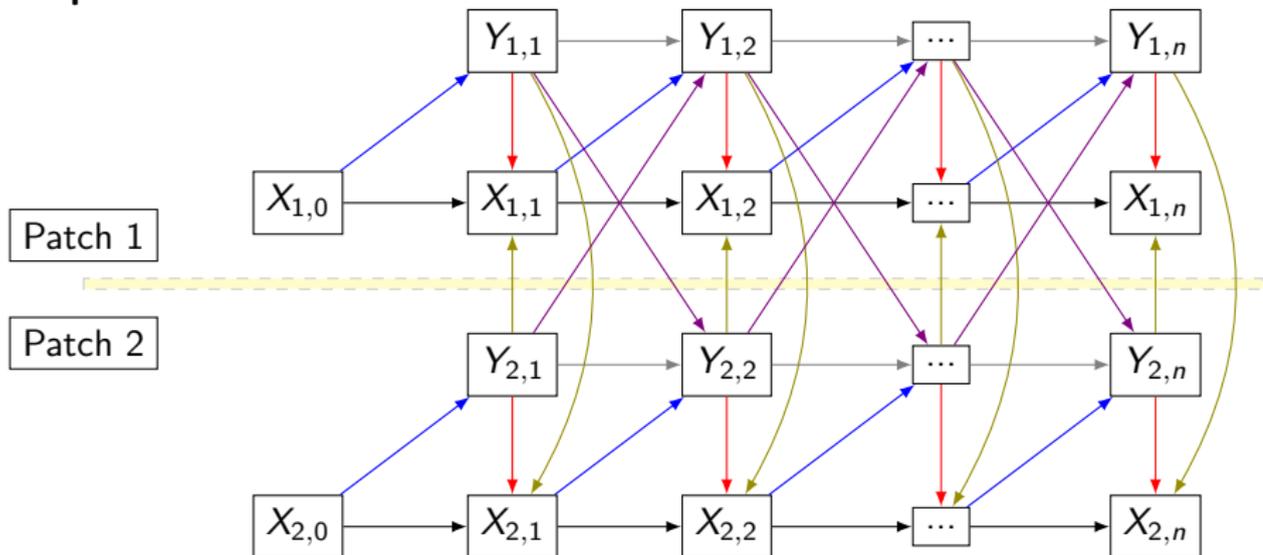
Inclut l'influence entre deux populations observables successives.



La colonisation spatiale

Inclut l'influence de populations observables vers des populations voisines.

Graphe du MHMM-DF



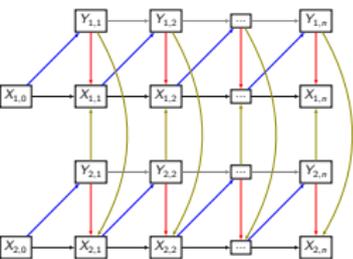
Les variables cachées :

$$X^{C,N} = (X_{1,1}, X_{2,1}, \dots, X_{C,1}, \dots, X_{C,1}, X_{1,2}, \dots, X_{1,n}, \dots, X_{1,N}, \dots, X_{C,N})$$

Les variables observées :

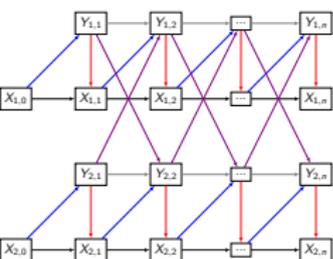
$$Y^{C,N} = (Y_{1,1}, Y_{2,1}, \dots, Y_{C,1}, \dots, Y_{C,1}, Y_{1,2}, \dots, Y_{1,n}, \dots, Y_{1,N}, \dots, Y_{C,N})$$

Applications du MHMM-DF



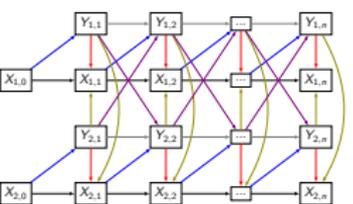
- Dynamique des plantes

✓ $Y_{c,n}$ vers $X_{c',n}$ × $Y_{c,n-1}$ vers $Y_{c',n}$



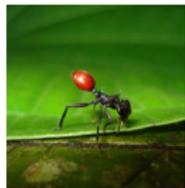
- Dynamique des puces

× $Y_{c,n}$ vers $X_{c',n}$ ✓ $Y_{c,n-1}$ vers $Y_{c',n}$

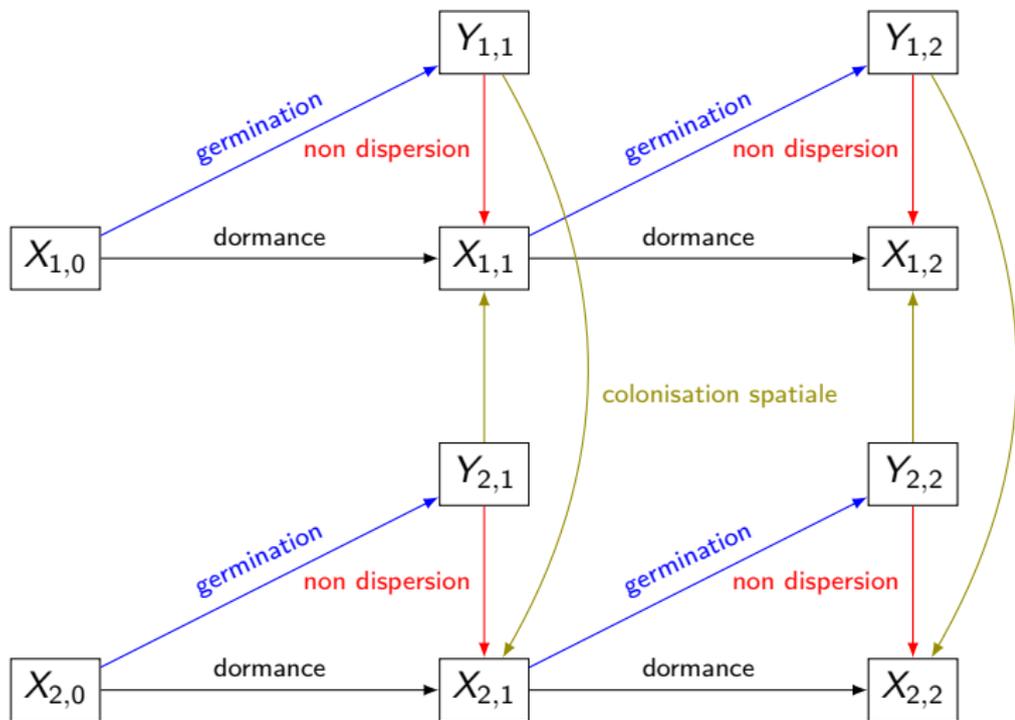


- Dynamique du parasite
Myrmecconema neotropicum

✓ $Y_{c,n}$ vers $X_{c',n}$ ✓ $Y_{c,n-1}$ vers $Y_{c',n}$



Dynamique des adventices annuelles

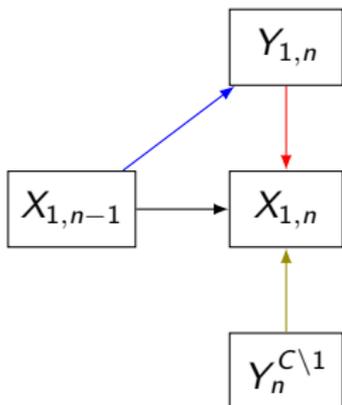


Que retenir de la construction du graphe ?

- La colonisation spatiale ne s'effectue que de populations observables vers populations voisines. Autrement dit, Les populations cachées n'influencent pas les populations voisines.
- Le MHMM-DF est applicable à plusieurs types d'espèces.

Paramétrisation des lois du MHMM-DF

Le MHMM-DF est défini par



- Probabilité initiale

$$\pi(x_{c,0}) = \mathbb{P}(X_{c,0} = x_{c,0}) \text{ ou } \mathbb{P}(x_0^C) = \prod_{c=1}^C \pi(x_{c,0}) .$$

- Probabilité d'émission

$$\phi(x_{c,n-1}, y_{c,n}) = \mathbb{P}(Y_{c,n} = y_{c,n} | X_{c,n-1} = x_{c,n-1})$$

- Probabilité de transition

$$A(x_{c,n-1}, x_{c,n}, y_n^C) =$$

$$\mathbb{P}(x_{c,n} | X_{c,n-1} = x_{c,n-1}, Y_n^{C \setminus c} = y_n^{C \setminus c}, Y_{c,n} = y_{c,n})$$

Modélisation Binomiale Logistique (BL)

Les paramètres (π, ϕ, A) sont décrits par une loi binomiale de paramètres respectifs $(|\Omega_X| - 1, p_\pi)$, $(|\Omega_Y| - 1, p_\phi)$, $(|\Omega_X| - 1, p_A)$.

Exemple: Loi d'émission

$$\phi(x_{c,n-1}, y) = \binom{|\Omega_Y| - 1}{y} [p_\phi]^y [1 - p_\phi]^{|\Omega_Y| - y - 1}$$

où

$$p_\phi = \frac{1}{1 + e^{-(\mu_1 \frac{x_{c,n-1}}{|\Omega_X|} + \mu_0)}}$$

Modélisation Binomiale Logistique (BL)

Les paramètres (π, ϕ, A) sont décrits par une loi binomiale de paramètres respectifs $(|\Omega_X| - 1, p_\pi)$, $(|\Omega_Y| - 1, p_\phi)$, $(|\Omega_X| - 1, p_A)$.

Exemple: Loi d'émission

$$\phi(x_{c,n-1}, y) = \binom{|\Omega_Y| - 1}{y} [p_\phi]^y [1 - p_\phi]^{|\Omega_Y| - y - 1}$$

où

$$p_\phi = \frac{1}{1 + e^{-(\mu_1 \frac{x_{c,n-1}}{|\Omega_X|} + \mu_0)}}$$

Avantage

La colonisation extérieure est modélisée à travers μ_0 .

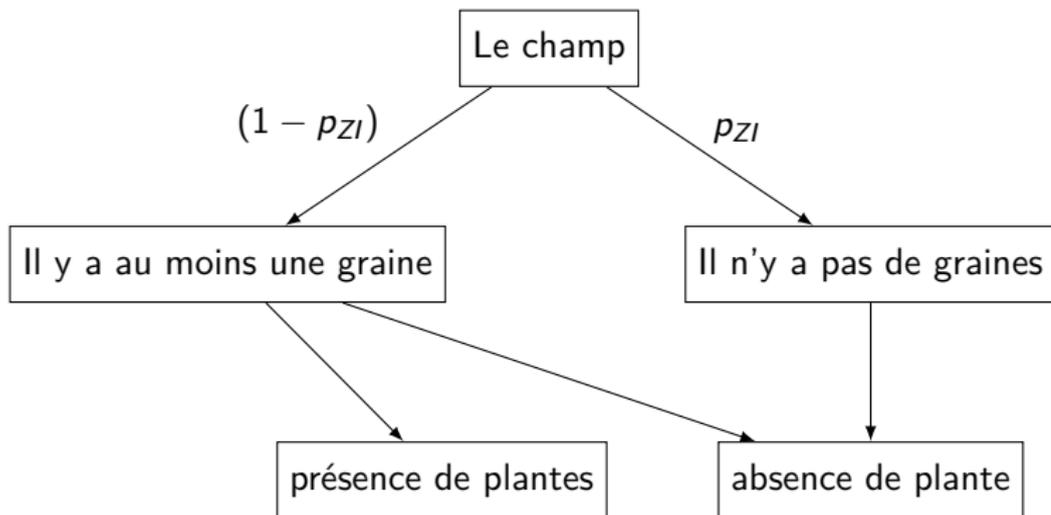
Inconvénient

ϕ ne doit pas inclure de colonisation extérieure dans le cadre des adventices.

Modèles Zero Inflated

Ils permettent de modéliser une variable aléatoire qui est fréquemment nulle.

Exemple : Soit p_{ZI} la probabilité d'absence de graine dans le champ.



Loi Zéro Inflated Binomiale Logistique (ZIBL)

$Y_{c,n+1} | X_{c,n}, Y_{c,n}, Y_n^{c \setminus c}$ suit une *ZIBL* avec paramètres $(|\Omega_Y| - 1, p_\phi, p_{ZI})$
si pour tout $y \in \Omega_Y$,

$$\begin{aligned} \phi(x_{c,n-1}, y) = & \mathbb{1}_{\{y=0\}} p_{ZI} \\ & + (1 - p_{ZI}) \binom{|\Omega_Y| - 1}{y} [p_\phi]^y [1 - p_\phi]^{|\Omega_Y| - y - 1} \end{aligned}$$

Dans la suite on pose $p_{ZI} = \mathbb{1}_{\{x_{c,n-1}=0\}}$.

Récapitulatif des hyper-paramètres associés à la dynamique d'une adventice annuelle

Hyper-paramètre	Population de départ	Population d'arrivée	Processus écologique
τ (π)	-	Banque de graines	-
μ_0 (ϕ)	-	Plantes adultes	-
μ_1 (ϕ)	Banque de graines	Plantes adultes	germination
ν_0 (A)	-	Banque de graines	colonisation extérieure
ν_1 (A)	Banque de graines	Banque de graines	dormance
ν_2 (A)	Plantes adultes	Banque de graines	non dispersion
ν_3 (A)	Plantes voisines	Banque de graines	colonisation spatiale

Récapitulatif des hyper-paramètres associés à la dynamique d'une adventice annuelle

Hyper-paramètre	Population de départ	Population d'arrivée	Processus écologique
τ (π)	-	Banque de graines	-
μ_0 (ϕ)	-	Plantes adultes	-
μ_1 (ϕ)	Banque de graines	Plantes adultes	germination
ν_0 (A)	-	Banque de graines	colonisation extérieure
ν_1 (A)	Banque de graines	Banque de graines	dormance
ν_2 (A)	Plantes adultes	Banque de graines	non dispersion
ν_3 (A)	Plantes voisines	Banque de graines	colonisation spatiale

Que retenir des paramétrisations ?

- Peu de paramètres.
- Chaque processus écologique est associé à un hyper-paramètre.
- Identifiabilité générique.
- Applicable pour le MHMM-DF complet.

Estimation à l'aide de l'algorithme EM

Le EM est un algorithme qui cherche les paramètres qui maximisent la vraisemblance d'un modèle avec données manquantes à l'aide de la fonction Q .

$$Q(\lambda|\lambda_{it}) = E_{\mathbb{P}}[\ln(\mathbb{P}(X^{C,N}, Y^{C,N}|\lambda)) | Y^{C,N} = y^{C,N}, \lambda_{it}]$$

où $\lambda = (\tau, \mu_0, \mu_1, \nu_0, \nu_1, \nu_2, \nu_3)$

L'algorithme EM:

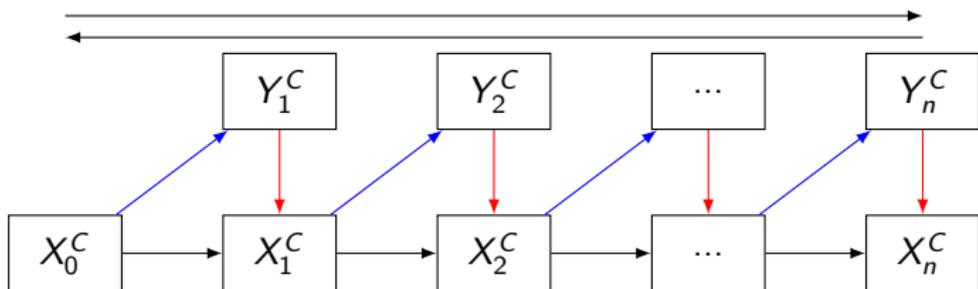
Initialisation des paramètres $\lambda_0 = (\tau_{ini}, \mu_{0,ini}, \mu_{1,ini}, \nu_{0,ini}, \nu_{1,ini}, \nu_{2,ini}, \nu_{3,ini})$.

Itérer sur les étapes E et M jusqu'à convergence de λ_{it} .

- Etape E: calcul des probabilités conditionnelles de $Q(\lambda|\lambda_{it})$.
- Etape M: mise à jour du paramètre λ_{it} par maximisation de $Q(\lambda|\lambda_{it})$.

Étape E : Forward-Backward

Forward Backward sur **toutes** les chaînes



Calculer

- $\rho_n(x_n^C) = \mathbb{P}(X_n^C = x_n^C | Y^{C,N}, \lambda_{it})$
- $\xi_n(x_n^C, x_{n-1}^C) = \mathbb{P}(X_n^C = x_n^C, X_{n-1}^C = x_{n-1}^C | Y^{C,N}, \lambda_{it})$

Nécessaires pour évaluer $Q(\lambda | \lambda_{it})$

Comment calculer ρ et ξ ?

ρ et ξ peuvent être calculés à l'aide des probabilités Forward (α) et Backward (β).

$$\alpha_n(x_n^C) = \mathbb{P}(y^{C,n}, X_n^C = x_n^C | \lambda_{it})$$

$$\beta_n(x_n^C) = \mathbb{P}(y_{n+1}^C, \dots, y_N^C | X_n^C = x_n^C, \lambda_{it})$$

Les α et les β sont calculés récursivement.

Étape M :

Mise à jour du paramètre λ avec :

$$\lambda_{it+1} = \arg \max_{\lambda} E_{\mathbb{P}}[\ln(\mathbb{P}(X^{C,N}, Y^{C,N}|\lambda)) | Y^{C,N} = y^{C,N}, \lambda_{it}]$$

Problème 1) La propriété cruciale pour calculer les ρ et ξ n'est pas vérifiée dans le MHMM-DF.

$$\sum_{x_n^C \in \Omega_X^C} [\alpha_n(x_n^C) \times \beta_n(x_n^C)] = \mathbb{P}(Y^{C,N} = y^{C,N} | \lambda_{it})$$

Problème 2) Le Forward-Backward sur toutes les chaînes à la fois prend beaucoup de mémoire et a un temps de calcul important !

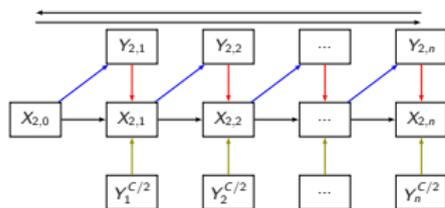
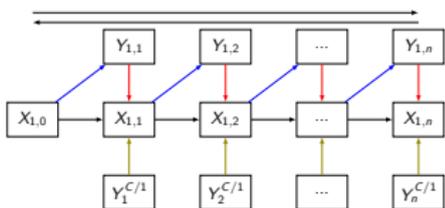
Formules du FB	Complexité mémoire	Complexité en temps
$\rho_n(x_n^C)$	$ \Omega_X ^C \times N$	$O(\Omega_X ^C \times N)$
$\xi_n(x_n^C, x_{n-1}^C)$	$ \Omega_X ^{2C} \times N$	$O(\Omega_X ^{2C} \times N \times C)$

Comment résoudre ces problèmes ?

Algorithme EM chaîne par chaîne

Étape E :

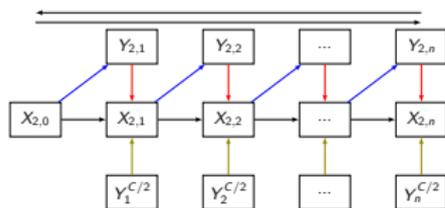
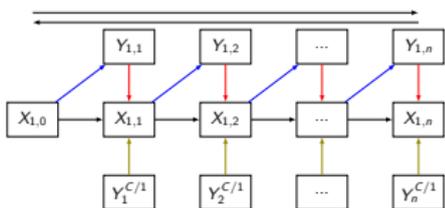
Forward Backward sur **chaque** chaîne indépendamment



Algorithme EM chaîne par chaîne

Étape E :

Forward Backward sur **chaque** chaîne indépendamment

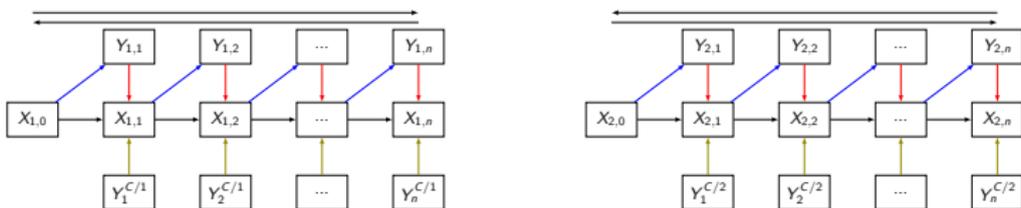


Pourquoi peut-on faire une FB sur chaque chaîne ?

Algorithme EM chaîne par chaîne

Étape E :

Forward Backward sur **chaque** chaîne indépendamment



Pourquoi peut-on faire une FB sur chaque chaîne ?

$$X_c^N \perp\!\!\!\perp X^{C \setminus c, N} \text{ sachant } Y^{C, N}$$

Calcul des ρ et ξ

Les ρ et ξ sont définis au niveau d'une chaîne.

- $\rho_{\mathbf{c},n}(x_{\mathbf{c},n}) = \mathbb{P}(X_{\mathbf{c},n} = x_{\mathbf{c},n} | Y^{C,N}, \lambda_{it})$
- $\xi_{\mathbf{c},n}(x_{\mathbf{c},n}, x_{\mathbf{c},n-1}) = \mathbb{P}(X_{\mathbf{c},n} = x_{\mathbf{c},n}, X_{\mathbf{c},n-1} = x_{\mathbf{c},n-1} | Y^{C,N}, \lambda_{it})$

Pour calculer les ξ et ρ , les α et β sont définis au niveau d'une chaîne.

$$\alpha_{\mathbf{c},n}(x_{\mathbf{c},n}) = \mathbb{P}(y^{C,n}, X_{\mathbf{c},n} = x_{\mathbf{c},n} | \lambda_{it})$$

$$\beta_{\mathbf{c},n}(x_{\mathbf{c},n}) = \mathbb{P}(y_{n+1}^C, \dots, y_N^C | X_{\mathbf{c},n} = x_{\mathbf{c},n}, \lambda_{it})$$

Problème : La propriété suivante n'est toujours pas vérifiée.

$$\sum_{x_{c,n} \in \Omega_X} [\alpha_{c,n}(x_{c,n}) \times \beta_{c,n}(x_{c,n})] = \mathbb{P}(Y^{C,N} = y^{C,N} | \lambda_{it})$$

Problème : La propriété suivante n'est toujours pas vérifiée.

$$\sum_{x_{c,n} \in \Omega_X} [\alpha_{c,n}(x_{c,n}) \times \beta_{c,n}(x_{c,n})] = \mathbb{P}(Y^{C,N} = y^{C,N} | \lambda_{it})$$

On définit les β différemment !

$$\alpha_{c,n}(x_{c,n}) = \mathbb{P}(y^{C,n}, X_{c,n} = x_{c,n} | \lambda_{it})$$

$$\begin{aligned} \beta_{c,n}(x_{c,n}) &= \mathbb{P}(y_{n+1}^C, \dots, y_N^C | X_{c,n} = x_{c,n}, \lambda_{it}) \\ &= \mathbb{P}(y_{n+1}^C, \dots, y_N^C | y^{C,n}, X_{c,n} = x_{c,n}, \lambda_{it}) \end{aligned}$$

Avec les nouveaux β , on a

$$\sum_{x_{c,n} \in \Omega_X} [\alpha_{c,n}(x_{c,n}) \times \beta_{c,n}(x_{c,n})] = \mathbb{P}(Y^{C,N} = y^{C,N} | \lambda_{it})$$

On peut ensuite calculer ρ et ξ .

Étape M :

Mise à jour du paramètre λ avec :

$$\lambda_{it+1} = \arg \max_{\lambda} E_{\mathbb{P}}[\ln(\mathbb{P}(X^{C,N}, Y^{C,N}|\lambda)) | Y^{C,N} = y^{C,N}, \lambda_{it}]$$

Que retenir de l'estimation?

Méthode d'estimation	Complexité mémoire	Complexité en temps
FB Toutes chaînes	$ \Omega_X ^C \times N(\Omega_X ^2 + 3)$	$O(\Omega_X ^{2C} \times N)$
FB chaque chaînes	$ \Omega_X \times C \times N(\Omega_X + 3)$	$O(\Omega_X ^2 \times N \times C)$

- L'algorithme EM est de complexité linéaire en terme du nombre de chaînes !
- L'étape E est exacte.
- L'algorithme EM est applicable au MHMM-DF.

Que retenir de l'estimation?

Méthode d'estimation	Complexité mémoire	Complexité en temps
FB Toutes chaînes	$ \Omega_X ^C \times N(\Omega_X ^2 + 3)$	$O(\Omega_X ^{2C} \times N)$
FB chaque chaînes	$ \Omega_X \times C \times N(\Omega_X + 3)$	$O(\Omega_X ^2 \times N \times C)$

- L'algorithme EM est de complexité linéaire en terme du nombre de chaînes !
- L'étape E est exacte.
- L'algorithme EM est applicable au MHMM-DF.

Remarque

La complexité de l'étape E n'est pas linéaire si les populations cachées influencent les populations voisines.

Que faire avec les paramètres estimés?

- Calcul de la vraisemblance du modèle
- Estimation d'états cachés : algorithme de Viterbi.
- Prédiction de l'état de la variable observée au temps suivant.

Étude de la qualité d'estimation sur données simulées

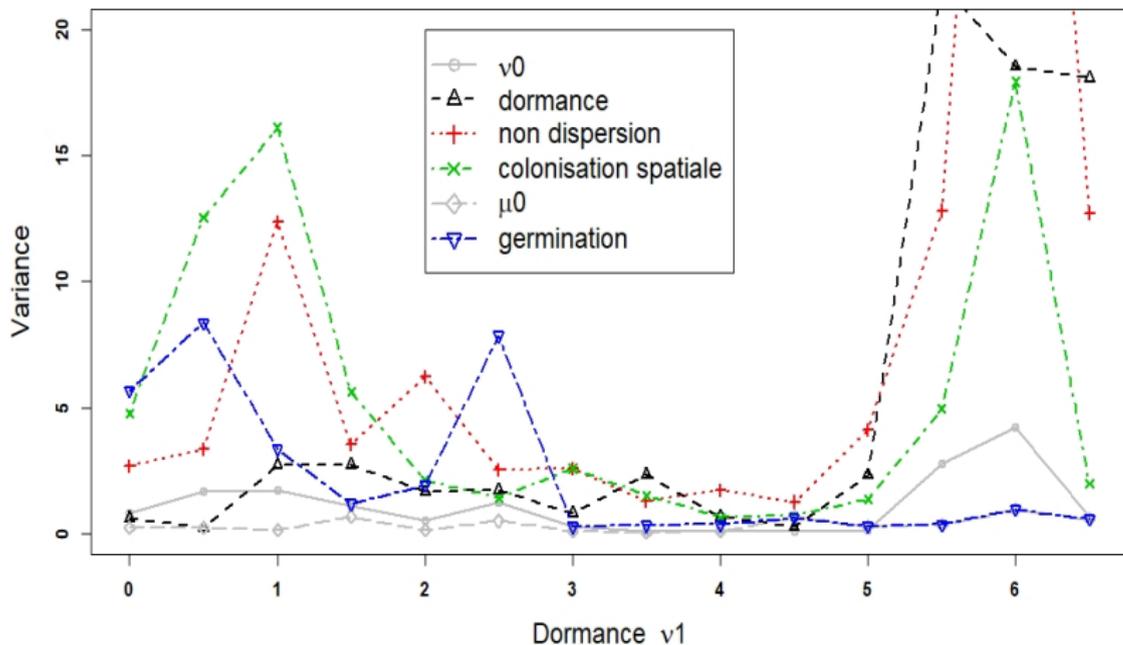
A , ϕ et π suivent des lois BL.

Hyper-paramètre	Processus écologique	Valeur
τ	-	-1
μ_0	colonisation extérieure	-3.7
μ_1	germination	6.5
ν_0	colonisation extérieure	-3
ν_1	dormance	{0, 0.5, ..., 6, 6.5}
ν_2	non dispersion	4
ν_3	colonisation spatiale	2

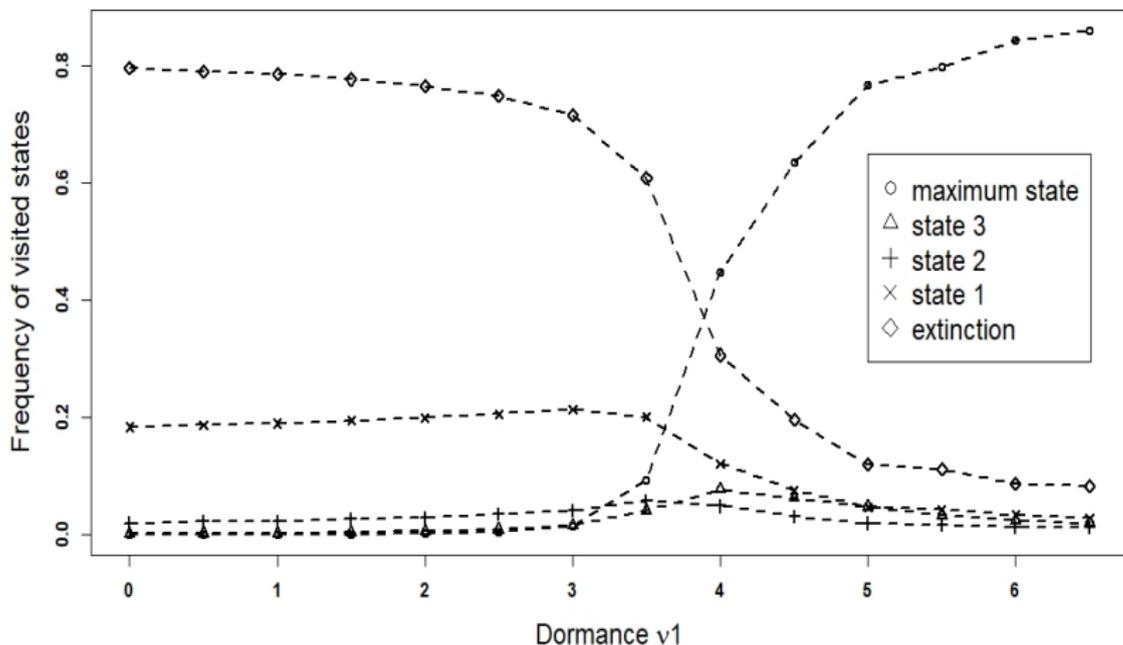
Protocole expérimental

- 10 chaînes
- 5 états pour les variables cachées et observées
- 100 pas de temps
- 14 jeux de paramètres
- 10 jeux de données simulées par jeux de paramètres
- 8 initialisations du EM par jeux de données

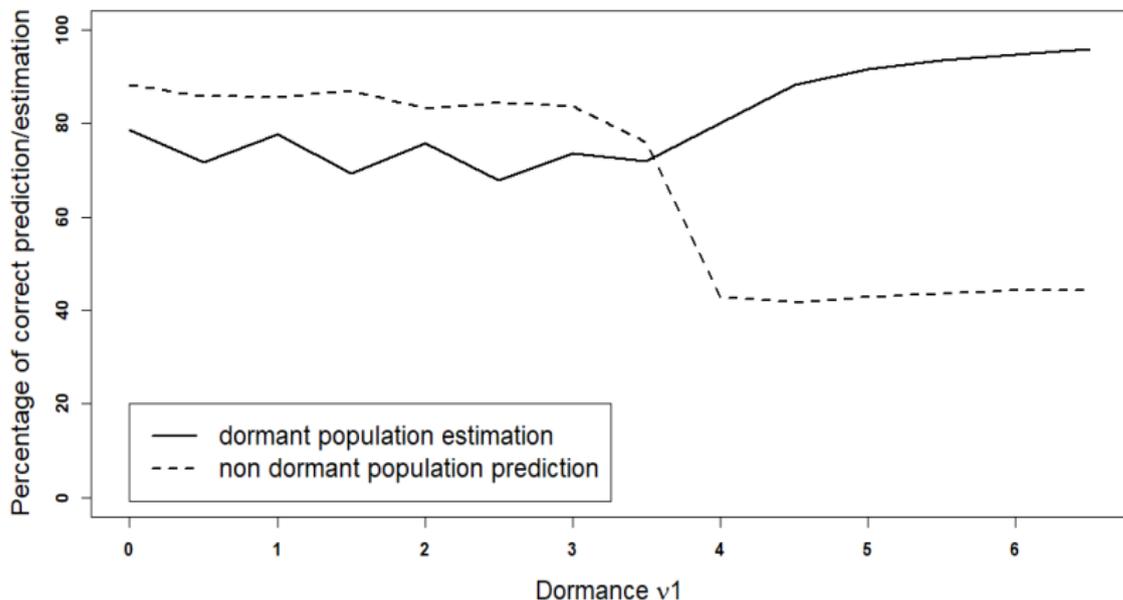
Graphe de variance des estimateurs



Graphe de la fréquence des états cachés visités



Prédiction des états observés et estimation des états cachés



Que retenir des résultats sur données simulées ?

- Dormance entre $[3, 5]$ \Rightarrow variance des estimateurs faible.
- Dormance entre $] - \infty, 3[\cup] 5, +\infty[\Rightarrow$ variance des estimateurs forte. Plusieurs jeux de paramètres peuvent conduire à des dynamiques extrêmes.
- Très bons résultats sur données simulées pour l'estimation d'états cachés et la prédiction.
- Prédiction mauvaise quand la probabilité de succès de la binomiale est proche de 0.5 \Rightarrow forte variance dans la distribution des états.

Article en révision dans le journal : Theoretical Population Biology.

Application à des données adventices d'Epoisses

Objectif mené sur le complexe d'Epoisses

Expérimenter des systèmes de cultures avec différentes stratégies de gestions pour en évaluer les performances agronomiques, économiques, environnementales et sociales.

5 types de stratégies de désherbage ont été utilisées.

- Rien.
- Travail du sol.
- Désherbage mécanique.
- Désherbage mécanique et chimique.
- Désherbage chimique.

Informations complémentaires de l'étude d'Epoisses

- Un champ fait 80m de large sur 250m de long.
- 2 groupes de 5 champs.
- 19 types de culture.
- 70 espèces adventices identifiées.

Les données d'Epoisses

Données sur les 7 espèces adventices les plus présentes dans l'étude d'Epoisses.

Les données sont en densité au mètre carré.

J'ai appliqué une transformation logarithmique puis un découpage uniforme aux données pour les convertir en 5 classes d'abondance.

Jeux de données « quadrats »

32 quadrats
(0.36m²)

Quadrat fixe sur
l'année, déplacé
entre années

Zone fixe

Densité par
comptage
(pl./m²)

D	2 3 1 4	E
C	2 3 1 4	F
B	2 3 1 4	G
A	2 3 1 4	H



Données pour 10 champs sur 17 ans.

Estimateurs du MHMM-DF

A et π suivent des lois BL. ϕ suit une loi ZIBL.

Espèces \ Estimateurs	τ	Probabilité de germination		Probabilité de transition de la banque de graines			
		μ_1	μ_0	ν_1	ν_2	ν_3	ν_0
<i>Alopecurus myosuroides</i>	-0.93	3.58	-3.16	3.91	4.13	1.22	-2.22
<i>Chenopodium album</i>	-2.55	6.16	-4.32	8.07	-6.41	1.98	-2.80
<i>Solanum nigrum</i>	-1.23	9.51	-6.08	4.84	-2.18	1.53	-2.16
<i>Fallopia convolvulus</i>	0.21	3.77	-3.39	6.52	-1.61	0.24	-2.36
<i>Aethusa cynapium</i>	-0.14	5.31	-3.98	6.88	-2.56	1.43	-2.77
<i>Galium aparine</i>	-3.68	7.05	-5.60	5.26	-2.01	2.25	-2.08
<i>Polygonum aviculare</i>	-0.76	4.49	-3.32	6.85	-3.36	2.78	-2.85

- La banque de graines est plus influencée par la banque de graines au temps précédent pour toutes les espèces sauf l'*Alopecurus myosuroides*.
- La banque de graines est plus influencée par la flore levée locale pour l'*Alopecurus myosuroides*.

Période de grenaison et Prédiction

Ici, les estimateurs dépendent de la saison de la culture, hiver ou été.

Espèces	Colonisation spatiale ν_3		Période de grenaison (Arino et al 2012, Taylor, 1999)	Pourcentage de prédiction correcte
	culture d'hiver	culture d'été		
<i>Alopecurus myosuroides</i>	0.89	2.47	été	0.511
<i>Chenopodium album</i>	0.69	2.6	été	0.875
<i>Solanum nigrum</i>	1.32	2.47	été	0.931
<i>Fallopia convolvulus</i>	0.40	7.84	été	0.579
<i>Aethusa cynapium</i>	2.03	0	hiver	0.795
<i>Galium aparine</i>	4.80	0	hiver	0.591
<i>Polygonum aviculare</i>	2.68	4.74	été	0.898

Durée de vie moyenne de la banque de graines

Espèces	Espérance de survie de la banque de graines [MHMM-DF]		Espérance de survie d'une graine (années) Littérature (Arino et al 2012)
	culture d'hiver	culture d'été	
<i>Alopecurus myosuroides</i>	2.04	1.85	3-4
<i>Chenopodium album</i>	1.85	5.55	6-8
<i>Solanum nigrum</i>	1.59	1.35	>10
<i>Fallopia convolvulus</i>	3.45	2.63	6-8
<i>Aethusa cynapium</i>	2.44	3.23	-
<i>Galium aparine</i>	2.44	2.70	3-4
<i>Polygonum aviculare</i>	1.69	4	6-8

Que retenir des résultats sur les données d'Epoisses ?

- Le modèle le plus pertinent pour l'*Alopecurus myosuroides* est un modèle sans la saison de la culture.
- Pour toutes les adventices, la saison de la culture ayant la plus forte colonisation spatiale coïncide avec la période de grenaison de l'adventice.
- La prédiction au hasard dans le modèle aurait un pourcentage de prédiction correcte de 20%. Le MHMM-DF a au minimum une prédiction correcte à 50%.
- La survie de la banque de graines est fortement sous- estimée dans le MHMM-DF !

Conclusion

Cadre MHMM-DF

- MHMM-DF applicable à plusieurs espèces.
- Le Forward-Backward pour le MHMM-DF est exact et a pour complexité $O(|\Omega_X|^2 \times N \times C)$.
- Identifiabilité générique.

Application aux adventices

Le modèle permet :

- d'identifier les processus les plus influents dans la dynamique d'une espèce avec stade caché,
- de savoir s'il faut contrôler l'adventice de façon locale ou régionale,
- de prédire la quantité de flore levée adventice de l'année d'après,

Perspective de recherches

Cadre MHMM-DF

- Les modèles semi-Markoviens cachés permettent une meilleure modélisation du temps de séjour dans les états cachés.
- Meilleure agrégation des variables observées dans la colonisation spatiale.
- Interaction entre chaînes de variables cachées vers des variables voisines.

Application aux adventices

- Prise en compte de la proportion de graines qui germent.
- Utilisation des données d'abondance pour les variables observées.

Merci de votre attention !

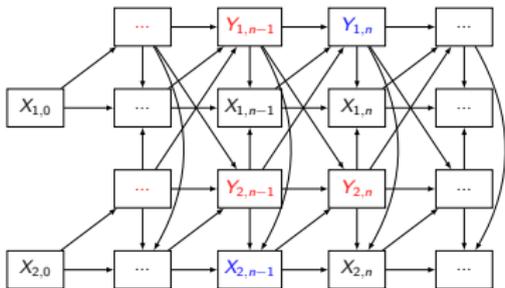
On a une écriture récursive :

$$\alpha_{c,n}(x_{c,n}) = \sum_{x_{c,n-1} \in \Omega_X} \alpha_{c,n-1}(x_{c,n-1}) \mathbb{P}(y_n^{c \setminus c} | y^{c,n-1}, y_{c,n}, x_{c,n-1}) \\ \times A(x_{c,n-1}, x_{c,n}, y_n^c, \lambda_{it}) \phi(x_{c,n-1}, y_{c,n}, \lambda_{it})$$

$$\beta_{c,n}(x_{c,n}) = \sum_{x_{c,n+1} \in \Omega_X} \beta_{c,n+1}(x_{c,n+1}) \mathbb{P}(y_{n+1}^{c \setminus c} | y^{c,n}, y_{c,n+1}, x_{c,n}) \\ \times A(x_{c,n}, x_{c,n+1}, y_{n+1}^c, \lambda_{it}) \phi(x_{c,n}, y_{c,n+1}, \lambda_{it})$$

Les termes en rouges sont des constantes !

On peut voir grâce a la figure que $\mathbb{P}(y_{n+1}^{C \setminus c} | y^{C,n}, y_{C,n+1}, x_{C,n})$ une constante. On suppose que $c = 2$



Ainsi on pose :

$$\tilde{\alpha}_{c,n}(x_{C,n}) = \sum_{x_{C,n-1} \in \Omega_X} \tilde{\alpha}_{c,n-1}(x_{C,n-1}) A(x_{C,n-1}, x_{C,n}, y_n^C, \lambda_{it}) \phi(x_{C,n-1}, y_{C,n}, \lambda_{it})$$

$$\tilde{\beta}_{c,n}(x_{C,n}) = \sum_{x_{C,n+1} \in \Omega_X} \tilde{\beta}_{c,n+1}(x_{C,n+1}) A(x_{C,n}, x_{C,n+1}, y_{n+1}^C, \lambda_{it}) \phi(x_{C,n}, y_{C,n+1}, \lambda_{it})$$

$$\rho_{c,n}(x_{c,n}) = \frac{\tilde{\beta}_{c,n}(x_{c,n})\tilde{\alpha}_{c,n}(x_{c,n})}{\sum_{x \in \Omega_X} \tilde{\beta}_{c,n}(x)\tilde{\alpha}_{c,n}(x)}$$

$$\xi_{c,n}(x_{n-1}, x_n) = \frac{A(x_{n-1}, x_n, y_n^C, \lambda_{it})\tilde{\beta}_{c,n}(x_n)\tilde{\alpha}_{c,n-1}(x_{n-1})\phi(x_{n-1}, y_{c,n}, \lambda_{it})}{\sum_{(x,x') \in \Omega_X^2} A(x', x, y_n^C, \lambda_{it})\tilde{\beta}_{c,n}(x)\tilde{\alpha}_{c,n-1}(x')\phi(x', y_{c,n}, \lambda_{it})}$$

$$\mathbb{P}(y^{C,N}) = \left[\sum_{x_{c,n} \in \Omega_X} \tilde{\alpha}_{c,n}(x_{c,n}) \tilde{\beta}_{c,n}(x_{c,n}) \right] \\ \times \left[\prod_{j=1}^N \prod_{l \in \mathcal{C} \setminus c} \sum_{x_{l,j-1} \in \Omega_X} \phi(x_{l,j-1}, y_{l,j}) \frac{\tilde{\alpha}_{l,j-1}(x_{l,j-1})}{\sum_{x \in \Omega_X} \tilde{\alpha}_{l,j-1}(x)} \right]$$

Param	Hyper-param	Hyper-para exact	Initialisation	Estimateur
A	$\begin{pmatrix} \nu_1 = \nu_{BG} \\ \nu_2 = \nu_{FL} \\ \nu_3 = \nu_{voi} \\ \nu_0 = \nu_{cst} \end{pmatrix}$	$\begin{pmatrix} 7.0 \\ 2.0 \\ 4.0 \\ -8.5 \end{pmatrix}$	$\begin{pmatrix} 0.530 \\ 1.278 \\ 3.390 \\ -2.302 \end{pmatrix}$	$\begin{pmatrix} 7.382 \\ 1.900 \\ 3.709 \\ -8.974 \end{pmatrix}$
ϕ	$\begin{pmatrix} \mu_1 = \mu_{BG} \\ \mu_0 = \mu_{cst} \end{pmatrix}$	$\begin{pmatrix} 6.5 \\ -5 \end{pmatrix}$	$\begin{pmatrix} 0.571 \\ 2.703 \end{pmatrix}$	$\begin{pmatrix} 5.611 \\ -4.403 \end{pmatrix}$
π	τ	1	1.265	1.632
A	$\begin{pmatrix} \nu_1 = \nu_{BG} \\ \nu_2 = \nu_{FL} \\ \nu_3 = \nu_{voi} \\ \nu_0 = \nu_{cst} \end{pmatrix}$	$\begin{pmatrix} 7.0 \\ 2.0 \\ 4.0 \\ -8.5 \end{pmatrix}$	$\begin{pmatrix} 0.558 \\ 3.788 \\ 0.725 \\ -2.814 \end{pmatrix}$	$\begin{pmatrix} 7.382 \\ 1.900 \\ 3.709 \\ -8.974 \end{pmatrix}$
ϕ	$\begin{pmatrix} \mu_1 = \mu_{BG} \\ \mu_0 = \mu_{cst} \end{pmatrix}$	$\begin{pmatrix} 6.5 \\ -5 \end{pmatrix}$	$\begin{pmatrix} 2.272 \\ 1.242 \end{pmatrix}$	$\begin{pmatrix} 5.611 \\ -4.403 \end{pmatrix}$
π	τ	1	2.212	1.632

Intervalle de confiance

David oakes, "Direct calculation of the information matrix via the EM".

$$I(\nu) = -\frac{\partial^2 Q(\nu|\hat{\nu})}{\partial \nu^2} \Big|_{\nu=\hat{\nu}}$$

A l'aide de la matrice de fisher, on peut calculer un intervalle de confiance avec précision $(100(1 - \epsilon))$.

$$[\hat{\nu}_1 - Z_{\epsilon/2} \left(\frac{1}{\sqrt{I(\nu)}} \right)_{\nu_1}, \hat{\nu}_1 + Z_{\epsilon/2} \left(\frac{1}{\sqrt{I(\nu)}} \right)_{\nu_1}]$$

