

Modélisation et statistiques de la dynamique des populations de colza hors champ à l'aide de processus de branchements multitypes

Olivier David* & Catherine Larédo*, Aurélie Garnier** & Jane Lecomte**

*INRA, Département MIA, Jouy-en-Josas

** Laboratoire d'Ecologie, Systématique et Evolution, Université Paris-Sud

1 Introduction

Cette étude s'inscrit dans la problématique de suivi de flux de gènes à l'échelle du paysage agricole. Le colza (*Brassica Napus*) est une espèce dont la culture est bien développée en Europe. Elle a la particularité de présenter de nombreuses populations échappées des cultures, observées sur les bords de routes, sur les chemins...

Une question centrale est d'évaluer le rôle que jouent ces populations hors champ pour les flux de gènes dans les agrosystèmes.

La culture du colza génétiquement modifié présente des risques pour l'environnement. Les variétés de colza résistant à un herbicide peuvent conférer le gène de résistance aux variétés conventionnelles cultivées au voisinage ou bien aux mauvaises herbes apparentées. Le risque de flux de gènes inter et intraspécifique est amplifié par la présence de colza en bordures des routes, issu de graines disséminées pendant le transport. Les plantes colonisant les bordures, appelées *populations spontanées* peuvent constituer un relais pour le transgène.

Afin d'étudier l'origine et la dynamique des populations spontanées de colza, deux suivis de terrain ont été menés : un premier suivi a été effectué tous deux fois par an pendant quatre années à l'échelle d'un bassin de production de colza situé dans le centre de la France (Adamczyk et al, 2005). Un deuxième suivi a été mené dans cette région à une échelle plus fine : durant trois années consécutives, les populations de colza spontanées situées sur trois routes et trois chemins ont été suivies quasi mensuellement. Elles sont décrites population par population avec leurs effectifs observés dans les différents stades de leur développement. La description de l'évolution de ces populations spontanées de colza hors champ est ainsi aussi fine que possible. Cependant, il est difficile sinon impossible pour ce type d'études de faire un suivi individuel des plantes. Ceci exclut de fait les approches classiques en dynamique des populations. Nous modélisons par des processus de branchement multitypes la dynamique de ces populations structurées en stades. D'abord, l'analyse statistique de ces données est faite en utilisant des méthodes classiques de statistique de processus de branchements multitypes. Ceci nous permet d'estimer certains paramètres du modèle et de tester des hypothèses les concernant: y-a-t-il un effet saisonnier, une différence de comportement entre les populations de bord de routes ou de chemins Ceci nous conduit à construire un modèle annuel plus complexe, où les

saisons sont prises en compte. Nous proposons de nouvelles méthodes d'estimation pour analyser ces données qui, dans le cadre écologique détaillé ici sont non standard. L'objectif de ces analyses, est d'acquérir des connaissances fines sur la dynamique des populations spontanées de colza, et, à terme, d'enrichir l'étude statistique du suivi à grande échelle (Adamczyk et al).

2 Description des données

Les données ont été recueillies en 2001, 2002 et 2003 sur l'ensemble des populations situées au bord de trois routes et trois chemins de la région de Selommès. Le nombre de populations suivies était important (186 en 2001, 215 en 2002, 155 en 2003). Dans chaque population, nous avons relevé les effectifs des plantes à 6 stades de développement entre le stade rosette et le stade plante mature (stades R, D, E, F, G, G5). Ces stades correspondent à des descriptions classiques de biologie du développement des plantes.

Les relevés ont eu lieu quasiment tous les mois entre janvier et juillet, ainsi qu'en octobre en 2001 et 2002. Les nombres de graines présentes dans le sol et produites par les plantes n'ont pas pu être mesurés. Il n'y a pas eu de suivi individuel des plantes.

Plusieurs covariables ont également été mesurées, comme l'état sanitaire de la population, la localisation GPS de la population, le type de route, la proportion de plantes fauchées ou ayant reçu un herbicide, les données climatiques.

3 Modèles probabilistes et analyses statistiques

3.1 Un premier modèle

Une première étape consiste à supposer que les données sont régies par un processus de Galton-Watson multitype. C'est un processus de branchements à trois types associé à une description annuelle au début de l'été de ces populations. Le cadre probabiliste est le suivant. A l'instant initial ($n=0$), on suppose qu'il y a S_0 graines, R_0 rosettes, et F_0 fleurs.

Soient $G_1(i, j, k)$, $(i, j, k) \in N^3$ la loi de reproduction d'une graine (resp. $G_2(i, j, k)$ d'une rosette), (resp $G_3(i, j, k)$ d'une fleur). Par exemple, la quantité $G_3(i, j, k)$ représente la probabilité qu'une fleur (type 3 ici) donne naissance à i graines, j rosettes et k fleurs au bout d'un an.

On suppose que toutes les plantes (i.e. S, R, F) vivent une durée de vie unité, et, à la fin de leur vie sont remplacées par leur descendance. De plus, toutes les plantes se reproduisent indépendamment les unes des autres, selon les mêmes lois G_i

Soulignons que ces lois de reproduction annuelles sont reliées aux probabilités de survie. Par exemple, $G_1(i, j, k)$ est la probabilité qu'une graine pendant l'année germe, fleurisse

et que la fleur obtenue produise i graines, j rosettes, et k fleurs. Par conséquent, on a $G_1(i, j, k) = 0$ dès que $i + j + k > 1$. Il en va de même pour G_2 . Seul $G_3(i, j, k)$ est non nul pour des valeurs de (i, j, k) telles que $i + j + k > 1$. C'est une vraie loi de reproduction.

L'étude de la taille des différentes populations à un instant n est régie par la matrice $M = m_{i,j}, 1 \leq i, j \leq 3$ contenant les nombres moyens de descendants de type j d'un individu de type i

$$m_{i,1} = EG_i(k, \cdot, \cdot) = \sum_{k,l,m} kG_i(k, l, m)$$

$$m_{i,2} = EG_i(\cdot, l, \cdot) = \sum_{k,l,m} lG_i(k, l, m)$$

La croissance de la population est telle que si $Z_0 = (S_0, R_0, F_0)$, la taille de la population Z_n vérifie: $E(Z_n/Z_0) = Z_0M^n$. La probabilité d'extinction est 1 si la plus grande valeur propre ρ de M est inférieure ou égale à 1, comprise strictement entre 0 et 1 sinon (c.f Athreya-Ney). Ces résultats montrent l'importance d'estimer "in situ" les paramètres de ce modèle.

La vraisemblance à partir des observations est difficile à manipuler, car elle requiert de calculer les produits de convolutions successifs des lois (G_1, G_2, G_3) . Une autre approche consiste à estimer les paramètres par la méthode des moindres carrés conditionnels. Ils sont ici faciles à calculer. Ils vérifient de plus la consistance et convergent à la bonne vitesse vers une loi "mixed normal" (cf Hall et Heyde)

3.2 Un modèle plus réaliste

Le modèle précédent, s'il est synthétique, ne permet pas d'estimer tous les paramètres du modèle. Une démarche naturelle est de revenir à une description plus réaliste de la dynamique du colza. C'est encore un modèle simplifié de la croissance des plantes de colza spontané, où nous conservons le cycle de vie des plantes.

La croissance du colza est ici structuré en trois stades : graines (S), rosettes (R), et fleurs (F). Notons que les passages de $S \rightarrow R$ et de $R \rightarrow F$ sont des probabilités de transition. Il y a reproduction uniquement dans la transition $F \rightarrow S$.

Nous gardons ainsi dans ce modèle simplifié les difficultés du modèle à 6 stades correspondant aux données recueillies, car ajouter des stades revient à rajouter des probabilités de transition qui s'estiment de façon analogue.

Contrairement aux plantes cultivées, les populations spontanées ont des périodes de germination et de floraison plus étendues dans le temps. Certaines transitions, a priori peu probables pour des populations cultivées sont ici observées.

Le cycle de vie annuel de ces populations spontanées se résume ainsi, saison par saison (cf Caswell) :

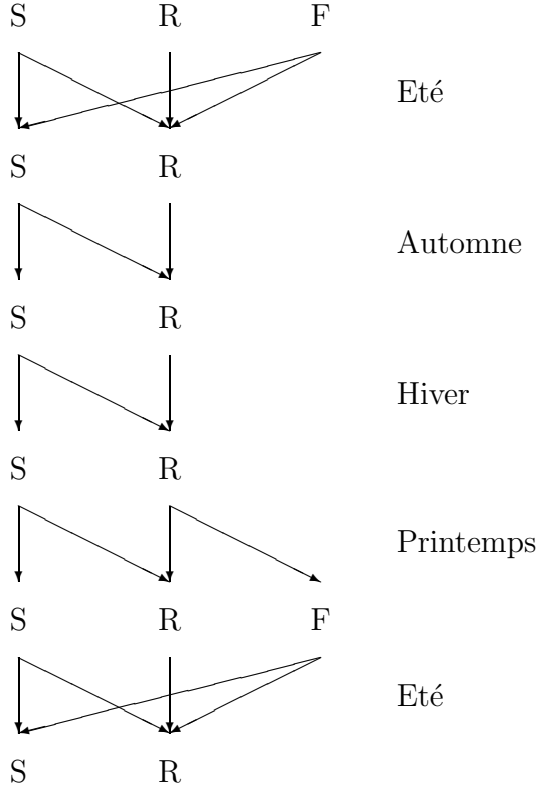


Figure 1. Cycle de vie annuel du colza. S : graine, R : rosette, F : plante en fleur.

Remarque: Dans ce schéma ne figure pas l'immigration. Elle intervient durant l'été, et alimente à la fin de l'été les stades S (graines) et rosettes (R). Nous avons choisi dans un premier temps de ne pas en tenir compte, pour simplifier.

Nous disposons des effectifs des types 2 (rosettes) et 3 (Fleurs) pratiquement tous les mois pour chaque population durant trois années.

Prenons l'exemple de la transition $R \rightarrow F$ qui arrive durant le printemps et désignons par t et $t+1$ les effectifs pour des mois successifs. Un stade intermédiaire E de description des plantes intervient alors dans les analyses.

L'effectif à un stade donné au mois $t+1$ est modélisé à l'aide des effectifs des stades précédents à t . Par exemple pour le stade F, nous considérons des modèles de la forme :

$$\begin{aligned}
 E(F_{t+1} | E_t, F_t) &= p_{EF} E_t + p_{FF} F_t, \\
 \text{var}(F_{t+1} | E_t, F_t) &= p_{EF} (1 - p_{EF}) (1 + (E_t - 1) \tau_{EF}) E_t \\
 &\quad + p_{FF} (1 - p_{FF}) (1 + (F_t - 1) \tau_{FF}) F_t,
 \end{aligned}$$

où E_t, F_t sont les effectifs dans les stades E et F au mois t , p_{EF} est la probabilité qu'une plante passe du stade E au stade F en un mois, p_{FF} est la probabilité qu'une plante au

stade F reste au stade F en un mois, et τ_{EF} , τ_{FF} sont des paramètres de surdispersion. Les paramètres du modèle (par exemple p_{EF} , p_{FF} , τ_{EF} , τ_{FF}) sont estimés par quasi-vraisemblance à l'aide du logiciel nls2 de R (Huet et al).

Il est possible d'étudier l'influence de facteurs sur ces paramètres, car nous disposons de répétitions : elles sont contenues dans les différentes populations (qui évoluent indépendamment), répétitions des mois dans les années. Nous pouvons aussi étudier l'effet des mois ou des saisons sur ces paramètres.

En regroupant les saisons par année, il est alors possible d'intégrer ces connaissances au modèle annuel décrit en 3.1 et ainsi d'étudier le devenir de ces populations de colza.

Bibliographie

- [1] Adamczyk, K., Pivard, S., Bouvier, A., Lecomte, J., Gouyon, P.-H. et Huet, S. (2005) Statistical analysis of oilseed rape dispersion data along a road network, XXXVIIèmes Journées de statistique à Pau.
- [2] Athreya, K. B. et Ney, P. E. (1972) *Branching processes*, Springer-Verlag.
- [3] Caswell, H. (2001) *Matrix population models : construction, analysis, and interpretation*, Sinauer, second edition.
- [4] Guttorp, P. (1991) *Statistical inference for branching processes*, John Wiley.
- [5] Hall, P. et Heyde, C. C. (1980) *Martingale limit theory and its application*, Probability and Mathematical Statistics, Academic Press.
- [6] Huet, S., Bouvier, A., Poursat, M.-A. et Jolivet, E. (2004) *Statistical tools for nonlinear regression : a practical guide with S-PLUS and R examples*, Springer Series in Statistics, Springer-Verlag, second edition.