
GENEXP, un logiciel simulateur de paysages agricoles pour l'étude de la diffusion de transgènes

**Katarzyna Adamczyk¹, Frédérique Angevin², Nathalie Colbach³,
Claire Lavigne⁴, Florence Le Ber^{5,6}, Jean-François Mari⁶**

¹ Unité MIA, INRA, F-78350 Jouy-en-Josas

Katarzyna.Adamczyk@jouy.inra.fr

² UAR 1240 Eco-innov, INRA, F-78850 Thiverval Grignon

Frederique.Angevin@jouy.inra.fr

³ UMR 1210 Biologie et Gestion des Adventices, INRA, F-21000 Dijon

colbach@dijon.inra.fr

⁴ UR 1115 Plantes et systèmes de culture horticoles, INRA, F-84000 Avignon

claire.lavigne@avignon.inra.fr

⁵ UMR MA 101 CEVH, ENGEES, F-67000 Strasbourg

florence.leber@engees.u-strasbg.fr

⁶ UMR 7503 LORIA, Université Nancy 2, F-54500 Vandœuvre-lès-Nancy

Jean-Francois.Mari@loria.fr

RÉSUMÉ. GENEXP est un simulateur de paysages agricoles qui engendre des découpages parcellaires en utilisant une géométrie algorithmique classique. GENEXP fournit, sur la base de parcelles réels ou réalistes, des cartes pluriannuelles de paysages agricoles utilisables par des logiciels qui simulent la dispersion des pollens et des graines d'OGM à différentes échelles.

ABSTRACT. The software GENEXP allows to simulate 2-dimensional agricultural landscapes by using a traditional algorithmic geometry. Based on real or realistic field-patterns, GENEXP provides multiannual maps of agricultural landscapes, which are used by softwares simulating the dispersal of GM pollen grains and seeds at various scales.

MOTS-CLÉS : paysage agricole, parcellaire, simulation, diagrammes de Voronoï, distribution de germes, processus ponctuel markovien, flux de gènes.

KEYWORDS: agricultural landscape, field-pattern, simulation, Voronoï tessellation, germs distribution, Markov point process, gene flow.

1. Introduction

Comment la structuration d'un paysage agricole joue-t-elle sur la diffusion par le vent, par les repousses ou les transports mécaniques, des transgènes issus des cultures génétiquement modifiées (OGM) ? Quels sont les paysages où le risque est limité ? Comment ces structures paysagères jouent-elles sur d'autres problèmes, tels que la diffusion de spores de maladies, le développement d'espèces invasives ou l'érosion ?

Pour répondre à ces questions, il est nécessaire d'élaborer des outils qui permettent de simuler ces phénomènes, sur la base d'expertises multiples : biologistes, agronomes, informaticiens, statisticiens. Ainsi, le travail présenté ici a été effectué dans le cadre de l'appel d'offre « Impact des OGM » du Ministère de la recherche (2002). Nous avons développé un simulateur de paysages agricoles bidimensionnels, GENEXP, qui engendre des découpages parcellaires à partir des paramètres statistiques, en utilisant une géométrie algorithmique classique. Les parcelles sont ensuite affectées d'une culture (colza, maïs, OGM ou non, blé, etc.) selon une distribution extraite de données réelles ou simulées. L'objectif est de simuler des paysages actuels (mais non forcément réels) ou possibles, dotés des caractéristiques utiles pour étudier un phénomène particulier : pour la diffusion des transgènes, les caractéristiques retenues sont par exemple la surface et la compacité des parcelles, le périmètre, la distance entre certains types de cultures ... La limitation à deux dimensions n'est *a priori* pas pénalisante, au vu des paramètres décrits dans les modèles actuels de flux de gènes. Soulignons encore qu'il ne s'agit pas ici d'utiliser les éventuelles données d'un système d'information géographique, ni d'envisager une plateforme générique comme décrite dans (Gaucherel *et al.*, 2006).

Cependant, GENEXP a été construit pour autoriser des couplages avec divers logiciels et intégrer différentes méthodes de pavage. En particulier, GENEXP est couplé avec le logiciel de statistiques **R** – ce qui permet de paramétrer et de décrire les paysages générés. Un couplage avec le logiciel de fouille de données CARROTAGE (Mari *et al.*, 2002) est également réalisé, pour pouvoir simuler des successions de culture dans le temps et l'espace. Finalement, GENEXP fournit sous différents formats des cartes pluriannuelles de paysages agricoles qui sont utilisées par les logiciels MAPOD-maïs (Angevin *et al.*, 2006) et GENESYS-colza (Colbach *et al.*, 2001) pour étudier le rôle du parcellaire agricole dans la dispersion des pollens et des graines d'OGM.

Dans cet article nous présentons les premiers résultats obtenus par les logiciels MAPOD-maïs et GENESYS-colza sur des parcellaires générés par GENEXP, et nous nous focalisons sur le lien possible entre des parcellaires réels numérisés et les parcellaires construits au moyen des diagrammes de Voronoï. Ce lien est exploré à partir d'une étude de la distribution des barycentres des parcelles sur six parcellaires réels différenciés dont nous disposons.

L'article est structuré en quatre parties. La première présente le contexte politico-agronomique de notre projet, qui est celui de l'introduction des cultures OGM. Le logiciel GENEXP, sa base géométrique et ses fonctionnalités, sont détaillés dans une deuxième partie. La troisième partie s'intéresse à la simulation de paysage sur la base

de parcelles réels, grâce aux outils d'estimation et de simulation fournis par le logiciel **R**. La quatrième partie présente des résultats obtenus sur les parcelles générés par GENEXP par les logiciels de simulation des flux de transgènes (couplage avec MAPOD-maïs et GENESYS-colza). Enfin nous concluons et présentons quelques perspectives de ce travail.

2. Contexte

Dans les années quatre-vingt dix dominait l'idée d'une évaluation au cas par cas des plantes transgéniques à mettre sur le marché, évaluation basée sur des essais en parcelles. Rapidement, le besoin est apparu d'intégrer ces nouvelles variétés dans des systèmes de culture de manière à mieux appréhender leur gestion par un agriculteur. Des essais comportant des variétés transgéniques de trois espèces de grande culture ont ainsi été mis en place par les instituts techniques, dans des conditions agricoles proches de la réalité (Champolivier *et al.*, 1997). Cependant, depuis quelques années, l'échelle de quelques parcelles ne paraît plus suffisante pour étudier les questions liées à la coexistence des cultures OGM, conventionnelles et biologiques et l'idée prévaut d'une gestion globale des cultures OGM à l'échelle d'un paysage agricole (Angevin *et al.*, 2002). Plusieurs causes peuvent expliquer ce changement d'approche. D'un point de vue biologique, la nécessité de ne pas considérer un champ transgénique comme une unité isolée tient, d'une part, à la mise en évidence de la capacité de certaines plantes cultivées à se maintenir pendant plusieurs années dans les champs et leurs bordures (voir par exemple (Crawley *et al.*, 1995, Pessel *et al.*, 2001)) et, d'autre part, à une évolution dans la vision de la communauté scientifique quant à l'appréhension des phénomènes de dispersion aussi bien du pollen (Klein *et al.*, 2006) que des graines (Colbach *et al.*, 2001, Von der Lippe, 2004). Il apparaît ainsi que la dispersion du pollen à longue distance a probablement longtemps été sous-estimée car elle était uniquement mesurée à partir d'expérimentations spatialement limitées (par exemple sur le colza : (Scheffler *et al.*, 1993) ou plus récemment (Funk *et al.*, 2006)) et non par des piègeages à l'échelle d'un paysage (voir par exemple (Squire *et al.*, 1999, Rieger *et al.*, 2002, Devaux *et al.*, 2005) sur colza). Par conséquent, au moins pour les espèces allogames, des échanges de gènes peuvent se produire entre champs cultivés distants de plusieurs centaines de mètres.

D'un point de vue agronomique, deux situations sont à considérer. La première concerne la mise en culture de variétés transgéniques et non transgénique d'une même espèce. La Commission Européenne a émis des recommandations (2003/556/EC) concernant la mise en place de règles de coexistence entre cultures OGM et non-OGM dans les pays membres et établi des règles d'étiquetage et de traçabilité des produits contenant des OGM (Réglementations EC 1829/2003 et 1830/2003). Le respect d'un seuil de 0,9 % d'OGM par ingrédient au dessus duquel les produits doivent être étiquetés rend nécessaire la maîtrise des risques de présence fortuite d'OGM par les agriculteurs et les collecteurs produisant du non-OGM ainsi que la mise en place de procédure de traçabilité. Par ailleurs, certains transgènes insérés dans des espèces de

grande culture (colza, maïs) ont pour vocation la production de substituts à la pétrochimie, de médicaments ; ce qui pose alors le problème de la coexistence de filières alimentaires et non-alimentaires dans les paysages agricoles.

La seconde situation concerne la mise en culture d'espèces différentes exprimant la même catégorie de transgène (par exemple, résistances à de mêmes populations d'insectes) qui nécessite de prendre en compte les interactions entre cultures pour comprendre l'éventuelle gestion des repousses entre cultures successives et les conditions dans lesquelles un transgène pourrait être sélectionné ou contourné hors de son champ d'origine. Ainsi, une même catégorie de transgène utilisée dans une rotation peut conduire à l'apparition de populations de ravageurs ou d'adventices résistants. Des croisements intra-espèce (colza) entre variétés tolérantes à différents herbicides peuvent produire des repousses multi-résistantes par la suite. Des croisements inter-espèces (colza-ravenelle) peuvent générer des adventices tolérantes à l'herbicide visé.

3. Le logiciel GENEXP

3.1. Diagrammes de Voronoï

Dans le cadre de l'étude de la diffusion de transgènes, notre objectif n'est pas de construire des paysages « réels », mais des paysages dotés des caractéristiques nécessaires à la simulation des phénomènes de dispersion de graines ou pollens. En particulier, nous nous limitons à deux dimensions. De plus, nous avons décidé de schématiser un paysage agricole bidimensionnel par un pavage polygonal. C'est sur cette base que nous avons réalisé la première version de GENEXP (Guerreiro, 2003), en construisant les paysages à l'aide d'un pavage bidimensionnel polygonal, obtenu à l'aide d'un diagramme de Voronoï (Chassery *et al.*, 1991). D'autres méthodes ont également été implantées ou étudiées (pavage à base de polygones simples quelconques, pavage rectangulaire aléatoire, ...) (Le Ber *et al.*, 2006). Nous ne traitons ici que de la méthode fondée sur les diagrammes de Voronoï, méthode que nous avons utilisée dans l'étude présentée ci-après.

Un diagramme de Voronoï est une partition du plan euclidien \mathbb{R}^2 , générée à partir d'un ensemble E de points, appelés *sites* ou *germes*. À chaque germe g est associé un élément p de cette partition, définie par le sous-ensemble des points du plan qui sont plus proches de g que de tous les autres germes de E . La partition ainsi construite est constituée d'autant de polygones que de germes. Ces polygones, appelés *polygones de Voronoï* sont de forme convexe. Leurs arêtes sont donc constituées des points à égale distance de deux germes (figure 1). Concrètement, GENEXP utilise la triangulation de Delaunay (Chassery *et al.*, 1991) pour déterminer le diagramme de Voronoï. Chaque polygone est ensuite identifié comme une parcelle du paysage agricole simulé.

Les inconvénients de cette approche sont de deux ordres : d'une part tous les polygones obtenus sont convexes, d'autre part le nombre de sommets est souvent élevé. Néanmoins elle fournit une bonne base de paysage, qui peut être améliorée avec des post-traitements, comme proposé dans GENEXP (voir ci-dessous). En particulier, on

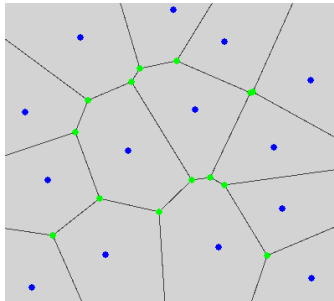


Figure 1. Un diagramme de Voronoï, avec ses germes et ses sommets affichés

peut jouer sur la taille et l'orientation – voire l'alignement – des polygones en jouant sur le processus de dissémination des germes. De plus, comme nous le verrons ci-dessous, cette méthode permet de relier assez naturellement les caractéristiques de paysages simulés aux caractéristiques de paysages réels.

3.2. Fonctionnalités de GENEXP

Les fonctionnalités principales de GENEXP se répartissent en trois temps : génération du parcellaire, post-traitements, description (Delaître, 2004). Nous décrivons ici les fonctionnalités qui concernent les « paysages Voronoï », mais qui sont également utilisables pour d'autres types de générations (pavage à partir de polygones ou rectangles aléatoires). Toutes ces fonctionnalités ont été implantées en Java, ainsi que l'interface de saisie et de visualisation.

3.2.1. Génération du parcellaire

Pour créer un nouveau paysage, l'utilisateur commence par définir ses dimensions, puis a la possibilité de partitionner cet espace à l'aide de deux outils : un outil de création de zones fermées et un outil de création de routes. Cette fonctionnalité permet à l'utilisateur, d'une part de définir un réseau de routes ou de zones non agricoles de formes quelconques, et d'autre part, de générer des parcellaires différents dans un paysage. Il est ainsi possible de simuler une zone composée de nombreux petits champs à côté d'une zone composée de quelques champs de dimensions importantes, orientés de manière différente.

Chaque zone est ensuite paramétrable indépendamment (figure 2). L'utilisateur décide d'un processus ponctuel pour générer les germes du diagramme de Voronoï, au moyen de fonctions disponibles sous \mathbf{R} (cf. section 4.3). Le choix des occupations du sol, assolement annuel ou assolement pluriannuel (Le Ber *et al.*, 2006), est également effectuée à ce niveau. Le parcellaire est alors calculé et l'utilisateur visualise le résultat

sous différentes options (figure 3). Les paysages sont exportés au format vecteur sous XML et au format maille.



Figure 2. Sélection d'une zone (en rouge) : le menu permet successivement de paramétrer la zone, puis de calculer le diagramme de Voronoï et enfin de visualiser différents éléments du paysage généré

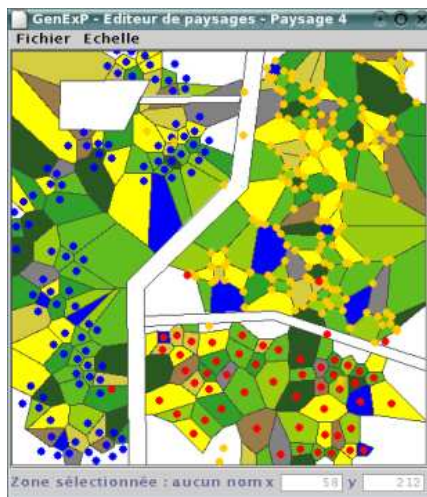


Figure 3. Paysage contenant des zones agricoles et non agricoles. Chaque parcellaire est affiché avec l'occupation du sol mais dispose de paramètres de visualisation différents :

- La région ouest est affichée avec ses germes (gris foncé) ;
- La région nord-est avec les sommets des parcelles (gris clair) ;
- La région sud-est avec les centres de gravité des parcelles (gris moyen).

3.2.2. Post-traitements et descripteurs

Dans la version actuelle, GENEXP permet d'effectuer quelques post-traitements sur le paysage généré :

- suppression des parcelles qui dépassent les limites des paysages ;
- élimination des côtés de parcelle de taille inférieure à une valeur donnée ;
- déplacement ou suppression de sommets (manuel).

Finalement, GENEXP dispose d'une bibliothèque pour calculer des descripteurs des parcelles (surface, périmètre, nombre de sommets, barycentre, rapport entre surface et périmètre) afin de caractériser les parcellaires obtenus, en référence aux caractéristiques de parcellaires réels. Les résultats à l'échelle du parcellaire sont affichés sous forme d'histogrammes. Cette fonctionnalité graphique permet de plus d'afficher simultanément les descripteurs de plusieurs paysages afin de les comparer.

4. Simulations de parcellaires à partir de germes de paysages réels

4.1. Démarche

Les modèles simulant les flux de gènes à l'échelle d'un paysage utilisent à cette fin des fragments de paysages existants. La question se pose alors de savoir dans quelle mesure les prédictions fournies par ces modèles dépendent des paysages en entrée. Une manière de répondre à cette question serait de simuler des paysages qui reproduisent certains descripteurs d'un paysage existant et d'étudier comment varient les résultats des modèles en fonction de ces descripteurs. Plusieurs descripteurs ont été définis et étudiés à cet effet : surface, périmètre, élongation, orientation, nombre de voisins, angle minimal d'une parcelle, ainsi que la localisation des barycentres (Chevauchet, 2004). Ces descripteurs visent à synthétiser l'ensemble des connaissances sur les paramètres importants de formes de parcelles, connaissances établies d'après les résultats d'expérimentations de dispersion de pollen entre deux parcelles (voir par exemple (Klein *et al.*, 2006)).

Le lien entre le paysage existant et le parcellaire simulé à partir d'un diagramme de Voronoï peut s'établir au travers de la distribution des germes. Une fois les positions des germes connues, la partition polygonale qui s'en suit est entièrement déterminée par l'algorithme (*cf.* section 3.1). L'idée est donc de trouver dans le paysage existant des points qui correspondent aux germes du diagramme de Voronoï, de caractériser leur distribution et de simuler les germes selon cette distribution. Ici, nous nous limitons à approximer les positions de ces points par les barycentres des polygones.

Nous proposons un modèle qui permet de contrôler 1) le nombre de barycentres, 2) la distribution de la distance entre des paires de barycentres. Nous estimons les paramètres de ce modèle en considérant les barycentres de six parcellaires. Pour vérifier si le modèle estimé conserve bien les deux caractéristiques ci-dessus nous procédons par simulations.

4.2. Descripteurs des barycentres des parcelles

Nous avons considéré les fragments de six paysages contrastés provenant de régions différentes¹, restreints à des régions de dimension 1,5 km × 1,5 km (figure 4). À chaque paysage est associée une configuration des barycentres des parcelles. Afin de décrire et de modéliser la répartition des barycentres, nous nous sommes placés dans le cadre des processus ponctuels spatiaux bidimensionnels. Ainsi, une configuration x des barycentres observés dans une région finie $W \subset \mathbb{R}^2$ est considérée comme une réalisation d'un processus ponctuel X .

Le premier descripteur du processus est le nombre de points. Selon ce descripteur les paysages peuvent être classés en trois catégories (voir tableau 1) : deux paysages

1. Ces parcellaires proviennent de différentes zones de production des cultures étudiées (colza, maïs) et une première étude a montré leur variabilité vis-à-vis d'un ensemble de descripteurs.

à 200 barycentres environ (*pern* et *tren*), trois paysages à 100 barycentres environ (*origmap*, *alsace* et *selom1*) et un paysage à 60 barycentres environ (*selom2*).

Le deuxième descripteur caractérise la distance entre des points du processus : la fonction $G(r)$ est définie comme la probabilité que la distance d'un point $x \in X$ à son plus proche voisin soit inférieure ou égale à r . Cet indicateur est utilisé pour comparer le processus étudié à un processus de Poisson homogène de même intensité, modélisant une répartition spatiale « au hasard ». Si la probabilité d'avoir un voisin à une distance donnée est plus forte que la même probabilité pour le processus de Poisson, on parle de processus *agregé*, si elle est plus faible, on parle de processus *régulier*.

La répartition des barycentres est soumise à des contraintes liées à la structure du paysage. La distance minimale entre deux barycentres est supérieure à un seuil qui dépend des dimensions des plus petites parcelles (10 à 90 m pour les paysages considérés). On s'attend aussi à ce que la probabilité de trouver une paire de barycentres dont la distance dépasse légèrement le seuil reste relativement faible (en référence au processus de Poisson), même si elle n'est pas nulle. La distribution des barycentres serait donc *a priori* régulière.

La figure 6 présente les distributions empiriques de la distance au plus proche voisin pour les six processus des barycentres. À titre de comparaison, les graphiques de la fonction G pour le processus de Poisson ont également été tracés. On remarque que les valeurs de G pour les processus des barycentres sont inférieures à celles calculées pour le processus de Poisson pour tous les paysages hormis *origmap*. La distribution des barycentres est donc en effet régulière dans la majorité des cas. Le graphique de la fonction G renseigne également sur l'étendue des distances entre des barycentres. La distance minimale séparant deux points varie de 10 m environ pour les paysages *origmap* et *tren*, à 90 m pour le paysage *selom2*. La distance maximale est de l'ordre de 180 m pour tous les paysages sauf *selom1*, où elle atteint 225 m environ.

4.3. Modèle des barycentres

L'analyse de la distance au plus proche voisin a mis en évidence la régularité du processus des barycentres dans la majorité des paysages. Afin de tenir compte des dépendances observées entre les points, nous avons choisi un modèle markovien pour le processus des barycentres. Nous avons supposé que X est un processus à interaction de paires dont la densité s'écrit :

$$f(x; \alpha, \beta, \theta) = \alpha \beta^n \prod_{i < j} \gamma(\|x_i - x_j\|; \theta) \quad [1]$$

où x est un vecteur de points de longueur n , β est un paramètre d'intensité supposée constante, γ est une fonction d'interaction dépendant de la distance entre les points et d'un vecteur de paramètres θ ; α est une constante de normalisation. Pour $\gamma \equiv \mathbf{1}$, l'équation (1) décrit la densité du processus de Poisson homogène d'intensité β . Les

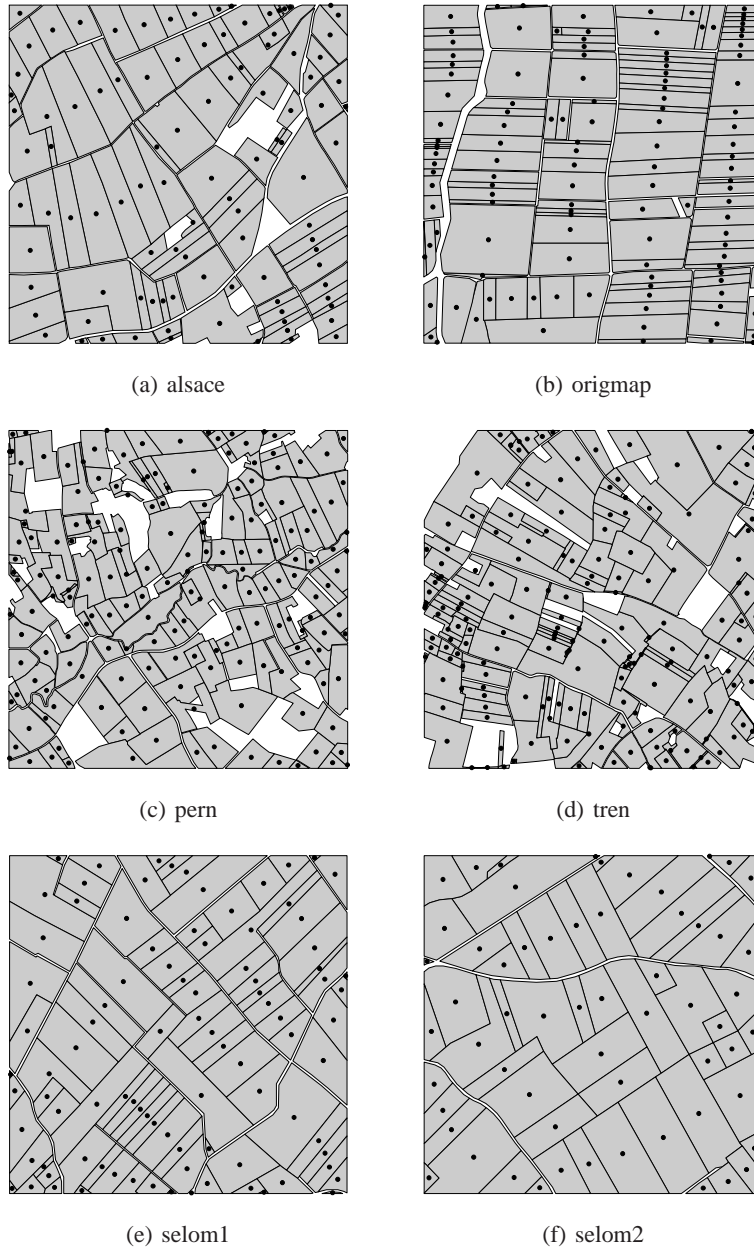


Figure 4. Six parcelaires utilisés pour estimer les paramètres du modèle des barycentres – données originales fournies par l'AUP-ONIGC (ex ONIC, Office National Interprofessionnel des Céréales) et l'IPSC, Institut pour la Protection et la Sécurité des Citoyens (Centre Commun de Recherche)

valeurs de $\gamma(r, \theta)$ inférieures (resp. supérieures) à 1 favorisent les configurations de points avec moins (resp. plus) de paires de points séparés par la distance r comparativement au processus de Poisson. Comme il est difficile d'intercepter la forme d'interaction, qui peut varier selon le paysage, nous avons choisi de la modéliser par une fonction constante par morceaux :

$$\gamma(r; \theta) = \sum_{k=1}^K \theta_k \cdot \mathbf{1}_{[r_{k-1}, r_k]}(r) \quad [2]$$

où $r_0 = 0 < r_1, \dots < r_K$ sont les distances-seuils et $\theta_1, \dots, \theta_K$ sont les valeurs prises par $\gamma(r, \theta)$ sur les intervalles associés.

Les paramètres du modèle (1) ont été estimés pour chaque paysage par la méthode du maximum de pseudovraisemblance (Baddeley *et al.*, 2000). À titre d'exemple, nous présentons ici l'estimation de la fonction d'interaction pour le paysage *selom1* (figure 5). Les calculs des estimateurs ont été réalisés dans la librairie *spatstat* du logiciel **R** (Baddeley *et al.*, 2005).

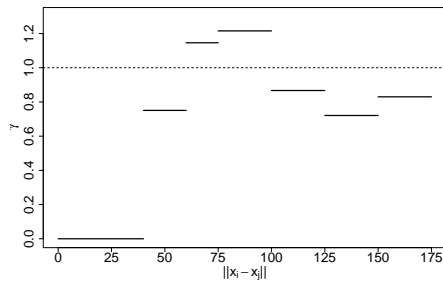


Figure 5. Estimation de la fonction d'interaction (2) pour le paysage *selom1*

4.4. Validation du modèle

Afin de vérifier si le modèle (1) estimé pour chaque paysage conserve bien les deux propriétés des barycentres, nous avons eu recours aux simulations. Nous illustrons notre démarche sur l'exemple du paysage *selom1*.

Nous avons simulé 100 réalisations du processus à interactions de paires (1) avec les paramètres estimés pour le paysage *selom1*. Les simulations ont été réalisées en utilisant l'algorithme de Metropolis-Hastings implémenté dans la librairie *spatstat*. Nous avons calculé le nombre de points et la fonction G pour chaque simulation du processus. Les graphiques des fonctions G pour 100 simulations ont été tracés ensemble avec la fonction G pour les barycentres de *selom1* (figure 6(e)). On note que

les valeurs de G pour les barycentres sont comprises entre les valeurs de G pour des processus simulés. De même, le nombre de barycentres – 93 points – appartient à un intervalle de confiance, déterminé à partir du nombre de points des processus simulés : $[80, 110]$. En d’autres termes, la configuration des barycentres de *selom1* peut être considérée comme une réalisation du processus (1). La même vérification a été effectuée pour tous les paysages (voir figure 6 et tableau 1), confortant le choix du modèle (1).

Paysage	nombre de barycentres	quantile 2.5%	quantile 97.5%
alsace	100	70	102
origmap	124	104	133
pern	175	143	179
tren	180	128	181
selom1	93	80	110
selom2	63	49	70

Tableau 1. Nombre des barycentres avec des intervalles de confiance calculés à partir de 100 simulations du modèle (1) pour chaque paysage

5. Simulation des flux de transgènes sur les parcellaires simulés

5.1. Caractéristiques des modèles MAPOD-maïs et GENESYS-colza

Le logiciel GENEXP fournit des cartes parcellaires à deux modèles de flux de gènes traitant de cultures différentes et déjà utilisés lors d’études européennes sur la coexistence des cultures transgéniques, conventionnelles et biologiques (Angevin *et al.*, 2002, Messéan *et al.*, 2006). GENESYS-colza (Colbach *et al.*, 2001) est un modèle pluriannuel qui permet de prédire des taux de présence fortuite dans les récoltes de colza (par exemple, taux d’OGM dans des récoltes non-OGM ou teneur en acide érucique dans des variétés 00) pour un parcellaire composé de champs cultivés et d’espaces hors-champs tels que les bordures de route, une succession de culture sur chaque parcelle et des techniques culturales associées à chaque culture, l’ensemble étant défini par l’utilisateur en entrée du modèle. MAPOD-maïs (Angevin *et al.*, 2006), quant à lui, simule la dissémination de gènes de maïs dans un espace spatialement hétérogène à l’échelle d’une campagne agricole, sans prendre en compte les successions de culture et les transferts de graines, celles-ci ne persistant pas d’une campagne à l’autre sous nos climats. Plus précisément, MAPOD-maïs prend en compte la forme, la taille et la localisation relative des parcelles de maïs-OGM et non-OGM. Un premier module calcule les dates de floraison des différentes variétés en fonction de données climatiques et des dates de semis, puis estime les quantités de pollen produites. Un second module calcule la dispersion du pollen comme une fonction de la distance entre les parcelles émettrice et receptrice, de la direction et de la force du vent dominant et des hauteurs des maïs dans les parcelles.

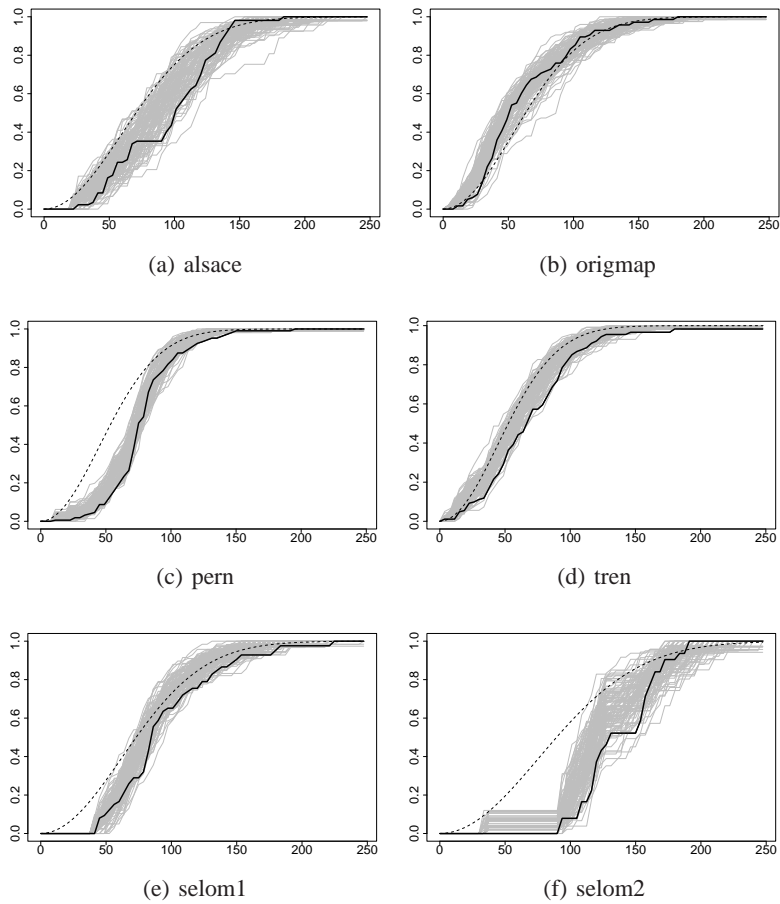


Figure 6. Distribution de la distance au plus proche voisin pour : le processus des barycentres (trait noir continu), le processus de Poisson (trait noir discontinu) et les 100 processus simulés selon le modèle (1) (traits gris) – en abscisse : les distances r (en m), en ordonnée : les valeurs de $G(r)$

Ces deux modèles spatialement explicites considèrent des parcellaires donnés en entrée et leurs conclusions sont au moins partiellement dépendantes du parcellaire choisi². Afin de rechercher les caractéristiques des parcellaires qui influencent le plus les taux d'impureté des récoltes entre parcelles, il est apparu nécessaire de disposer de modèles probabilistes de parcellaires agricoles, décrivant la répartition spatiale des

2. Voir (Colbach *et al.*, 2005, Klein *et al.*, 2006) pour une analyse de l'impact des formes et tailles des parcelles sur un parcellaire composé de deux parcelles et (Colbach *et al.*, 2005) pour l'analyse de situations contrastées.

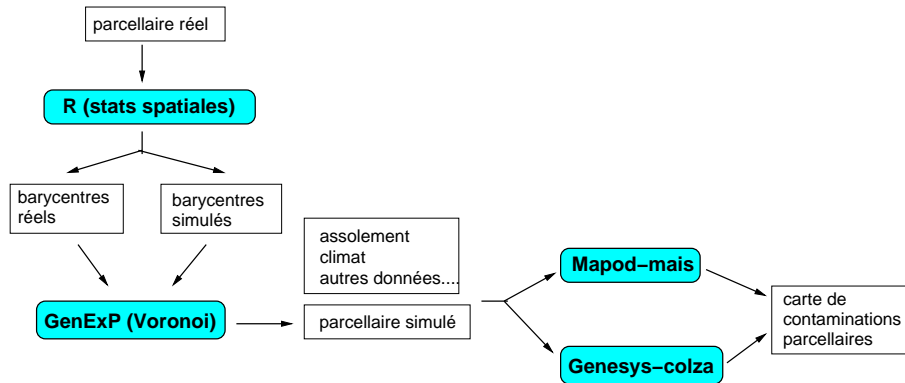


Figure 7. Enchaînement des différents modèles et données pour la simulation de flux de gènes dans un paysage agricole

champs (surface totale couverte, agrégation des champs d'une même espèce ou variété, distribution des surfaces des champs, distance minimale entre champs, distance minimale entre groupes de champs de variétés différentes). Cette relation a été étudiée, dans un premier temps, sur des parcellaires existants.

Cependant, il est nécessaire d'étudier plusieurs réalisations d'un même lot de caractéristiques parcellaires. C'est ici qu'intervient le couplage avec GENEXP qui permet de produire des parcellaires ayant des caractéristiques déterminées, et ainsi permet *via* GENESYS-colza ou MAPOD-maïs de déterminer la distribution des probabilités de respect du seuil d'impureté pour des paysages globalement différents mais similaires pour certaines caractéristiques géométriques.

5.2. Utilisation de MAPOD-maïs et GENESYS-colza sur des parcellaires GENEXP

Une première étude est en cours pour comparer les caractéristiques utiles de paysages réels et des paysages simulés par GENEXP en utilisant les diagrammes de Voronoï. Six parcellaires fournis par l'AUP-ONIGC (ex ONIC) et l'IPSC ont été utilisés pour réaliser des simulations avec GENESYS-colza et MAPOD-maïs (cf. figure 4). Nous ne détaillerons que les résultats obtenus avec un parcellaire alsacien (noté « original » par la suite) à partir duquel ont été générés deux parcellaires avec GENEXP (figure 8) : pour le premier, les barycentres du paysage original ont été conservés comme germes pour les diagrammes de Voronoï (parcellaire noté « b-réel » par la suite). Pour le second (noté « b-simul » par la suite), ces barycentres-germes ont été simulés grâce au modèle de processus ponctuel estimé sur le parcellaire original (cf. section 4.3).

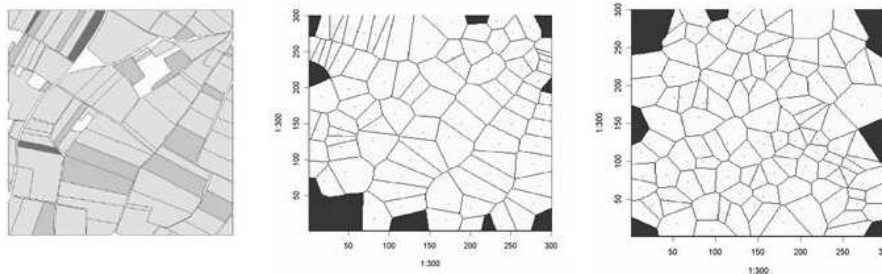


Figure 8. Parcelles construites à partir des barycentres réels (au milieu) et des barycentres simulés par le modèle de processus ponctuel (à droite), à comparer au parcellaire original (à gauche)

Pour MAPOD-maïs, ont été testées plusieurs allocations de cultures résultant de combinaisons entre le pourcentage de la surface totale en maïs et le pourcentage de variétés OGM. Par ailleurs, pour le paysage réel, plusieurs attributions de culture ont été testées : affectation aléatoire des parcelles de maïs OGM et non-OGM sur le paysage, création d'îlots de maïs tels qu'ils peuvent se rencontrer dans les zones irriguées. Ces derniers pouvaient être mixtes (variétés OGM et non-OGM) ou non mixtes (variétés OGM ou non-OGM). Ceci représente les deux stratégies d'adoption possible des variétés OGM (Messéan *et al.*, 2006). Afin de tester de façon privilégiée, les effets des différents parcellaires, l'itinéraire technique ne variait pas entre variétés OGM et non-OGM.

Pour GENESYS-colza, dans tous les cas, la région entière était cultivée avec une rotation colza/blé d'hiver/orge de printemps ; la moitié des variétés de colza était OGM, l'autre non-OGM. La moitié des champs portaient une jachère avant le colza (à la place de l'orge), tous les 6 ans. Au début de la simulation, les quatre cultures étaient distribuées aléatoirement dans le parcellaire en respectant les proportions de 1/3, 1/3, 1/6 et 1/6 pour le colza, le blé, l'orge et la jachère respectivement. De même, pour les colzas, l'utilisation d'une variété OGM ou non-OGM était choisie aléatoirement. Pour chacun des trois parcellaires, deux assolements aléatoires étaient réalisés, aboutissant ainsi à six simulations distinctes. Les années suivantes, la culture suivante de la rotation était affectée à chaque parcelle : par exemple, une parcelle cultivée en blé la première année était cultivée en orge la deuxième, puis en colza la troisième et de nouveau en blé la quatrième. La conduite des cultures était la même dans tous les champs et toutes les simulations : semis aux dates habituelles pour les cultures (fin août pour le colza, fin septembre pour le blé, mi-février pour l'orge), travail du sol effectué au chisel puis à la herse, herbicides dans toutes les cultures sauf le colza non-OGM.

5.3. Résultats

L'objectif de ces simulations est de pouvoir analyser si les paysages simulés peuvent rendre compte des contraintes de coexistence observées sur les paysages réels. Pour cela, les taux moyens d'impuretés GM dans les parcelles non-OGM ont été étudiés. Par la suite, des études similaires ont été effectuées pour comparer les surfaces de non-OGM devant être déclassées car ne respectant pas le seuil légal d'étiquetage (fixé à 0,9%).

Dans le cas de GENESYS-colza, la méthode d'obtention du parcellaire n'a pas d'effet significatif sur les résultats (voir ³ tableau 2). La distribution des cultures dans la région a en revanche un effet très important puisque les taux d'impuretés au niveau régional varient du simple au double l'année de l'introduction des OGM. Cet effet semble s'estomper au cours du temps puisque 6 ans plus tard, la différence entre simulations s'est considérablement réduite. La raison en est la progression des repousses OGM sur l'ensemble de la région qui fait disparaître l'effet de la proximité des colzas OGM avec les champs non-OGM.

parcellaire assolé	% OGM dans récolte non-OGM (moyenne régionale)		% champs dépassant			
	Année 0	Année 6	0,30%		0,90%	
			An0	An6	An0	An6
original 1	0,14 (0,45)	1,11 (3,52)	22	78	9	65
original 2	0,24 (0,75)	0,57 (1,81)	43	100	14	57
b-réel 1	0,15 (0,46)	1,08 (3,41)	53	95	11	79
b-réel 2	0,28 (0,90)	1,10 (3,46)	64	82	18	73
b-simul 1	0,19 (0,60)	1,01 (3,19)	65	94	0	88
b-simul 2	0,28 (0,88)	0,97 (3,06)	61	100	33	94

Tableau 2. Résultats GENESYS-colza pour le parcellaire alsace (deux répétitions par parcellaire avec des allocations de cultures différentes) : simulation avec 30% de colza dans l'assolement régional, dont 50% de variétés transgéniques, système de culture à risque intermédiaire

Pour MAPOD-maïs, où, rappelons-le, la simulation s'effectue sur une année, le rôle de la méthode d'obtention du parcellaire est plus marqué. Les résultats, très variables, présentés sur le tableau 3, ne permettent pas de tirer des conclusions dès à présent. Une carte de diffusion des transgènes issus des maïs OGM sur les parcelles non-OGM est présentée sur la figure 9 (parcellaire original).

3. Les valeurs de taux entre parenthèses ont été corrigées et les seuils affichés sont en fait de 0,1% et 0,4%, respectivement, pour prendre en compte la sous-estimation systématique par GENESYS-colza du flux de gènes dans l'espace (Colbach *et al.*, 2001). Pour le calcul de la moyenne régionale, le taux d'un champ n'est pas pondéré par la surface.

parcellaire assolé	Taux moyen d'OGM par parcelle non-OGM	% parcelles qui dépassent 1/1000	% parcelles qui dépassent 1/100
original 1	0,0087	10/12 (83%)	4/12 (33%)
original 2	0,0120	10/12 (83%)	3/12 (25%)
b-réel 1	0,0048	6/15 (40%)	3/15 (20%)
b-réel 2	0,0062	11/16 (69%)	4/16 (25%)
b-simul 1	0,0050	9/14 (64%)	2/14 (14%)
b-simul 2	0,0070	12/18 (66%)	5/18 (28%)

Tableau 3. Résultats MAPOD-maïs pour le parcellaire alsace (deux répétitions par parcellaire avec des allocations de cultures différentes) : simulations avec 30% maïs et 50% OGM dans ce maïs

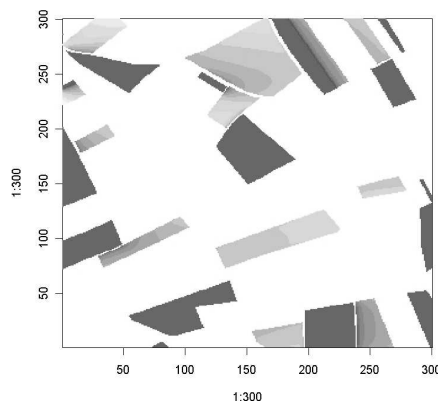


Figure 9. Résultat de MAPOD-maïs pour le parcellaire alsace : les parcelles de maïs OGM sont affichées en gris foncé, les taux de présence fortuite d'OGM sur les parcelles de maïs non-OGM – liés au vent dominant – sont visualisés par les dégradés de gris

6. Conclusion

Lors des études sur la coexistence de filières OGM et non-OGM (Angevin *et al.*, 2002), menées à l'échelle de paysages agricoles européens représentatifs de certains systèmes de production (Saxe, Bourgogne, plaine du Po, Beauce), il est apparu que la structure du parcellaire et l'allocation des cultures jouent un rôle important dans la faisabilité de la coexistence. Deux questions de recherche ont alors été soulevées :

– comment avoir une approche statistique des risques de présence fortuite d'OGM dans les récoltes lorsqu'on ne dispose que d'un paysage caractéristique d'une région ?

– quelle est l'extrapolabilité à d'autres régions des règles de coexistence définies grâce à ce type d'étude ?

Ces questions de recherche s'appliquent de façon plus générale à tous les phénomènes de flux (spore, érosion, espèce invasive, ...) et intéressent donc plusieurs communautés : écologues, agronomes et statisticiens. Ce projet de recherche pluridisciplinaire a donc réuni ces équipes de recherche ainsi que des informaticiens.

Nous avons présenté dans cet article deux aspects du travail entrepris pour répondre à ces questions. Le premier aspect concerne la mise au point d'un logiciel de simulation de paysages agricoles, le logiciel GENEXP. Ce logiciel a été construit de manière ouverte, par briques successives, pour permettre de générer des paysages « à la demande » selon plusieurs approches. Il est couplé avec le logiciel de statistique **R** et produit des cartes de parcelles utilisables par les logiciels MAPOD-maïs et GENESYS-colza.

Le second aspect concerne la mise en relation de paysages simulés et de paysages réels. La modélisation des barycentres des parcelles par un processus à interactions de paires permet de conserver deux caractéristiques de leur distribution : le nombre de points et la distribution des distances entre des paires de points. La simulation du modèle ajusté à partir des paysages existants reproduit correctement les valeurs de ces descripteurs. L'interface entre **R** et GENEXP permet d'appliquer directement ces résultats dans la simulation des germes du diagramme de Voronoï.

Finalement nous avons présenté nos premiers résultats concernant la simulation de la diffusion de transgènes sur les paysages simulés par comparaison aux paysages réels. Une étude plus complète est en cours et devrait permettre à la fois de valider l'approche choisie, d'améliorer le logiciel GENEXP et de mieux cerner l'intérêt des parcelles simulés par rapport aux parcelles réels accessibles aujourd'hui par numérisation.

Remerciements

Ce projet a été subventionné par le MENRT dans le cadre de l'appel d'offre « Impact des OGM ». Nous remercions les différentes personnes impliquées et tous les stagiaires, en particulier Jérémie Delaître, Antonio Guerreiro et Mathieu Morey, qui ont contribué au développement de GENEXP, ainsi que Kiên Kiêu (INRA MIA) pour ses conseils sur le modèle des barycentres. Nous remercions par ailleurs l'Université Paris XI, l'AUP-ONIGC (ex ONIC, Office National Interprofessionnel des Céréales) et l'IPSC, Institut pour la Protection et la Sécurité des Citoyens (Centre Commun de Recherche), pour les cartes des parcelles utilisées dans cette étude.

7. Bibliographie

Angevin F., Colbach N., Meynard J.-M., Roturier C., « Analysis of necessary adjustments of farming practices », *Scenarios for co-existence of genetically modified, conventional and*

- organic crops in European agriculture*, Technical Report Series, EUR 20394 EN, Joint Research Center of the European Commission, 2002.
- Angevin F., Klein E., Choimet C., Gauffreteau A., Lavigne C., Messéan A., Meynard J.-M., « Modelling impacts of cropping systems and climate on maize cross pollination in agricultural landscapes : The MAPOD model », 2006, Soumis.
- Baddeley A., Turner R., « Practical maximum pseudolikelihood for spatial point patterns », *Australian and New Zealand Journal of Statistics*, vol. 42, n° 3, p. 283-322, 2000.
- Baddeley A., Turner R., « Spatstat : an R package for analyzing spatial point patterns », *Journal of Statistical Software*, vol. 12, n° 6, p. 1-42, 2005.
- Champolivier J., Messéan A., « Impact du colza transgénique dans les systèmes de culture : étude du flux de gènes », *OCL*, vol. 4, p. 111-113, 1997.
- Chassery J.-M., Montanvert A., *Géométrie discrète en analyse d'images*, Hermès, 1991.
- Chevauchet F., « Application de méthodes de statistique spatiale pour décrire et simuler le paysage », Rapport de stage de fin d'études, ENSAI, Rennes, 2004. Unité MIA, INRA, Jouy-en-Josas.
- Colbach N., Clermont-Dauphin C., Meynard J.-M., « GENESYS-Rape. A model of the influence of cropping system on gene escape from herbicide tolerant rapeseed crops to rape volunteers », *Agric. Ecosyst. Environ.*, vol. 83, p. 235-270, 2001.
- Colbach N., Molinari N., Meynard J.-M., Messéan A., « Spatial aspects of gene flow between rapeseed varieties and volunteers », *Agronomy for sustainable development*, vol. 25, n° 3, p. 355-368, 2005.
- Crawley M., Brown S., « Seed limitation and the dynamics of feral oilseed rape on the M25 motorway », *Proc. of the Royal Society of London*, vol. B 259, p. 49-54, 1995.
- Delaître J., « Simulation de paysages agricoles aléatoires bidimensionnels », Mémoire de DUT en informatique, IUT Nancy 2, 2004. LORIA, Nancy.
- Devaux C., Lavigne C., Falentin-Guyomarc'h H., Vautrin S., Lecomte J., Klein E., « High diversity of oilseed rape pollen clouds over an agro-ecosystem indicates long-distance dispersal », *Molecular Ecology*, vol. 14, p. 2269-228, 2005.
- Funk T., Wenzel G., Schwarz G., « Outcrossing frequencies and distribution of transgenic oilseed rape (*Brassica napus* L.) in the nearest neighbourhood », *European journal of agronomy*, vol. 24, p. 26-34, 2006.
- Gaucherel C., Giboire N., Viaud V., Houet T., Baudry J., Burel F., « A domain-specific language for patchy landscape modelling : The Brittany agricultural mosaic as a case study », *Ecological Modelling*, vol. 194, n° 1-3, p. 233-243, 2006.
- Guerreiro A., « Développement d'un générateur expérimental de paysages agricoles aléatoires bidimensionnels », Rapport de stage du DESS méthodes et outils informatiques de la chimie, UHP Nancy 1, 2003. Lab. ESE (CNRS - Université Paris Sud).
- Klein E., Lavigne C., Picault H., Michel R., Gouyon P.-H., « Pollen dispersal of oilseed rape : estimation of the dispersal function and effects of field dimension », *Journal of Applied Ecology*, vol. 43, n° 1, p. 141-151, 2006.
- Le Ber F., Lavigne C., Mari J.-F., Adamczyk K., Angevin F., « GenExP, un logiciel pour simuler des paysages agricoles, en vue de l'étude de la diffusion de transgènes », in C. Weber, P. Gançarski (eds), *Actes du Colloque International de Géomatique et d'Analyse Spatiale (SAGEO 2006)*, Strasbourg, p. 1-12, 2006. Actes sur CD.

- Mari J.-F., Le Ber F., Benoît M., « Segmentation temporelle et spatiale de données agricoles », *Revue Internationale de Géomatique*, vol. 12, n° 4, p. 439-460, 2002.
- Messéan A., Angevin F., Gómez-Barbero M., Menrad K., Rodríguez-Cerezo E., « New case studies on the coexistence of GM and non-GM crops in European agriculture », Technical Report Series of the Joint Research Center of the European Commission, 2006. EUR 22102 En.
- Pessel F. D., Lecomte J., Emeriau V., Krouti M., Messéan A., Gouyon P.-H., « Persistence of oilseed rape (*Brassica napus* L.) outside of cultivated fields », *Theor. appl. Genet.*, vol. 102, p. 841-846, 2001.
- Rieger M. A., Lamond M., Preston C., *et alii*, « Pollen-mediated movement of herbicide resistance between commercial canola fields », *Science*, vol. 296, p. 2386-2388, 2002.
- Scheffler J., Parkinson R., Dale P., « Frequency and distance of pollen dispersal from transgenic oilseed rape (*Brassica napus*) », *Transgenic Research*, vol. 2, p. 356-364, 1993.
- Squire G. D., Crawford J., Ramsay G., Thompson C., « Gene flow at the landscape level », in P. Lutman (ed.), *Gene flow and agriculture : relevance for transgenic crops*, vol. 72 of *British Crop Protection Council Symposium Proceedings*, p. 57-64, 1999.
- Von der Lippe M., « Dispersal of arable crops by vehicles-a potential vector for the spread of genetically modified organisms », *Proceedings of the 34th Annual Conference of the Ecological Society, Giessen, Germany*, p. 276, 2004.