



**HAL**  
open science

# Épidémiologie de l'enroulement viral sur une parcelle modèle : analyse spatiotemporelle avec le logiciel R

Jean-Jacques Fuchs

► **To cite this version:**

Jean-Jacques Fuchs. Épidémiologie de l'enroulement viral sur une parcelle modèle : analyse spatiotemporelle avec le logiciel R. *Méthodologie [stat.ME]*. 2010. dumas-00520121

**HAL Id: dumas-00520121**

**<https://dumas.ccsd.cnrs.fr/dumas-00520121v1>**

Submitted on 22 Sep 2010

**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.

Épidémiologie de l'enroulement viral sur une  
parcelle modèle : analyse spatiotemporelle  
avec le logiciel R

INRA  
28 rue de Herrlisheim  
68021 COLMAR  
Maître de stage : Olivier LEMAIRE

**Jean Jacques FUCHS**  
Master Statistique 1<sup>ère</sup> année  
Université de Strasbourg  
31 mai 2010- 13 août 2010



# Sommaire

|   |      |
|---|------|
| Remerciements .....   | 4    |
| INTRODUCTION .....  | 5    |
| I – Présentation de l'établissement d'accueil .....   | 6    |
| I-1- Le centre INRA de Colmar.....  | 6    |
| I-2- L'UMR « Santé de la vigne et qualité du vin ».....   | 6    |
| I-3- L'équipe Virologie et Vecton.....  | 7    |
| II - Contexte de l'étude : éléments de biologie .....   | 7    |
| II-1- L'enroulement viral de la vigne.....  | 7    |
| II-1-1- Généralités .....   | 7    |
| II-1-2- Symptômes de la maladie .....   | 8    |
| II-1-3- Conséquences sur le rendement et la qualité des moûts .....   | 8    |
| II-1-4- Transmission du virus .....   | 8    |
| II-2- Les cochenilles sur la vigne .....  | 9    |
| III – Présentation de la problématique.....   | 9    |
| III-1- La parcelle modèle de Bonzon .....   | 9    |
| III-1-1- Descriptif .....   | 9    |
| III-1-2- Épidémiologie de la parcelle .....   | 10   |
| III-2- Démarche de l'analyse.....   | 11   |
| III-2-1- Jeu de données à analyser.....   | 11   |
| III-2-2- Raisonnement .....   | 11   |
| IV – Etude pratique.....  | 12   |
| IV-1- Analyse de l'agrégation des symptômes et des cochenilles dans la parcelle.....  | 12   |
| IV-1-1- Outils et méthodes.....   | 12   |
| IV-1-2- Utilisation du variogramme et des tests de permutations.....  | 13   |
| IV-1-2-1- Test d'indépendance totale .....  | 13   |
| IV-1-2-2- Test d'indépendance entre les lignes.....   | 14   |
| IV-1-3- Étude de l'agrégation des symptômes dans la parcelle modèle.....  | 15   |
| IV-1-3-1- Test d'indépendance par permutation totale des ceps.....  | 16   |
| IV-1-3-2- Test d'indépendance inter-rang .....  | 16   |
| IV-1-4- Étude de l'agrégation des cochenilles dans la parcelle modèle.....  | 17   |
| IV-1-4-1- Test d'indépendance totale des ceps des données « cochenilles 08 » et<br>« P.aceris08 ».....                                  | 17   |
| IV-1-4-2- Test d'indépendance spatiale entre les rangs.....   | 18   |
| IV-2- Corrélation entre la présence de <i>P. aceris</i> et de l'enroulement viral dans la parcelle modèle.....                          | 19   |
| IV-2-1- Test d'indépendance dans le cadre non-spatial.....  | 19   |
| IV-2-2- Test d'indépendance dans le cadre spatial .....   | 21   |
| IV-2-2-1- Test d'indépendance spatiale de deux grilles par translations.....  | 21   |
| IV-2-2-2- Test d'indépendance spatiale entre la présence de symptômes et de<br><i>P. aceris</i> .....                                   | 22   |
| IV-2-3- Modélisation.....   | 23   |
| IV-2-3-1- Régression logistique .....   | 23   |
| IV-2-3-2- Test d'indépendance spatiale des symptômes conditionnellement au nombre de cochenilles<br>( <i>P. aceris</i> ) observées..... | 25   |
| IV-3- L'enroulement viral provient-il des parcelles contiguës ? .....   | 27   |
| IV-3-1- Test d'indépendance de la structure des foyers de symptômes .....   | 27   |
| IV-3-2- Test d'indépendance spatiale de la structure des symptômes avec les données des parcelles<br>voisines.....                      | 29   |
| IV-3-3- Étude de la population virale : piste moléculaire.....  | 30   |
| CONCLUSION DE L'ETUDE .....   | 31   |
| BIBLIOGRAPHIE.....  | 32   |
| ANNEXE 1 .....  | I    |
| ANNEXE 2 .....  | VIII |
| ANNEXE 3 .....  | X    |



## Remerciements

Je tiens tout d'abord à remercier Monsieur Olivier Lemaire, mon maître de stage, qui m'a accueilli dans son équipe.

De plus, je tiens à remercier tout particulièrement Jean Le Maguet de m'avoir accordé sa confiance pour participer à ces travaux, ainsi que son aide et son accompagnement.

Je tiens aussi à remercier Joël Chadoeuf, biométricien de l'INRA d'Avignon, pour les conseils qu'il m'a donné.

Je remercie également toute l'équipe ViVe et plus généralement l'ensemble du personnel de l'INRA pour son accueil chaleureux.

## INTRODUCTION

En France, la vigne et le vin font partie intégrante du patrimoine. Les climats ainsi que les sols sont favorables à la culture de la vigne, c'est pour cela qu'on l'exploite dans diverses régions telles que la Bourgogne, la Champagne, le Bordelais, la Provence ou l'Alsace. Avec environ une cinquantaine de cépages différents cultivés en France, les vins français sont appréciés partout à travers le monde. Pour conserver cette renommée, il faut comprendre et protéger la vigne car cette culture peut être sujette à de nombreux problèmes sanitaires. Les différentes maladies touchant la vigne doivent être maîtrisées car elles entraînent des dégâts sur les plantes et par conséquent sur le raisin.

L'Institut National de la Recherche Agronomique (INRA) de Colmar se situe au cœur du vignoble alsacien et a ainsi choisi d'axer ses recherches autour de la vigne. L'un de ses rôles est de comprendre et d'essayer de combattre certaines de ces maladies.

Durant mon stage, je me suis intéressé plus particulièrement à l'une de ces maladies, provoquée par un complexe de virus. Mes travaux se sont portés sur l'enroulement viral de la vigne et sa vection par des cochenilles. Mon stage a ainsi consisté à étudier la dynamique épidémiologique de cette maladie sur une parcelle modèle afin d'apporter la preuve que la progression de la maladie sur ce vignoble est due à l'infestation des ceps par les cochenilles.

Ainsi, j'ai effectué une analyse de données spatio-temporelle à l'aide de différentes méthodes statistiques. Dans un premier temps, j'ai évalué l'agrégation de la maladie et des cochenilles dans la parcelle. Puis dans une seconde étape, j'ai étudié la corrélation entre la présence de la maladie et la présence de cochenilles. Enfin j'ai poursuivi l'étude pour tenter de prouver que l'épidémie d'enroulement viral provient des parcelles contiguës.

# I – Présentation de l'établissement d'accueil

## I-1- Le centre INRA de Colmar

L'Institut National de la Recherche Agronomique (INRA) regroupe une centaine d'unités dans différents centres répartis dans toute la France. Mon stage s'est déroulé au centre INRA de Colmar.

L'INRA de Colmar emploie une centaine d'agents de différentes qualifications : chercheurs, enseignants-chercheurs, ingénieurs, assistants ingénieur, techniciens, et doctorants. Tous ces agents associent leurs compétences multiples afin de mener à bien leurs projets de recherche : agronomie, biochimie, biologie moléculaire, entomologie, génétique, œnologie, physiologie végétale et virologie.

Outre les laboratoires, les équipes de recherche disposent de 3000 m<sup>2</sup> de serres dans lesquelles les plantes nécessaires aux expérimentations sont produites. L'INRA de Colmar possède également 55 ha de grandes cultures et de 15 ha de vignoble.

À l'échelle nationale, la plupart des équipes de l'INRA sont regroupées en Unités Mixtes de Recherches (UMR).

## I-2- L'UMR « Santé de la vigne et qualité du vin »

L'Unité Mixte de Recherche Santé de la Vigne et Qualité du Vin (SVQV) associe des chercheurs de l'INRA et des enseignants-chercheurs de l'Université de Strasbourg (UDS) autour de trois thématiques de recherche majeures :

- Les virus de la vigne et leur transmission par des vecteurs invertébrés pour définir des stratégies de protection contre ces virus au vignoble.
- Les sources de résistance au mildiou et à l'oïdium chez les Vitacées pour construire des variétés durablement résistantes et produisant des vins de qualité.
- L'étude du métabolisme secondaire de la vigne pour comprendre son rôle dans les mécanismes de défense contre les pathogènes et son importance dans la synthèse des arômes des vins.

Le personnel de l'UMR comprend :

- 18 chercheurs et ingénieurs, dont 15 docteurs et 2 enseignants-chercheurs docteurs,
- 25 adjoints techniques, techniciens et assistants ingénieurs,
- 7 thèses en cours dans l'unité.

L'UMR est constituée de trois équipes :

- Génétique et Amélioration de la Vigne,
- Virologie et Vection,
- Métabolisme Secondaire de la Vigne

Les recherches conduites dans l'UMR ont pour objectif d'obtenir des données scientifiques afin de pouvoir proposer à plus long terme des solutions alternatives à la lutte chimique. Ces projets s'inscrivent pleinement dans le cadre des recommandations du Grenelle de l'Environnement pour faire évoluer les productions conventionnelles vers des agricultures plus durables.

## I-3- L'équipe Virologie et Vection

J'ai effectué mon stage au sein de l'équipe Virologie et Vection (ViVe). Elle comprend principalement des virologues et des entomologistes. Ses travaux portent sur l'étude des virus de la vigne et la transmission de virus par des vecteurs invertébrés.

Deux axes de recherche sont développés dans cette équipe :

- le premier, vection des phytovirus par vecteurs invertébrés, se décline en trois parties : la biologie de la transmission des virus par cochenilles, les mécanismes moléculaires de la transmission des virus par nématodes et de ceux de la transmission des polérovirus par puceron;
- le deuxième, portant sur l'étude des interactions virus-vigne, se décline en (i) l'identification des facteurs viraux et végétaux responsables du pouvoir pathogène des virus de la vigne et (ii) en l'obtention de vignes résistantes aux virus avec l'étude des mécanismes sous-jacents.

Au sein de cette équipe, j'ai été en charge de l'analyse d'une partie des données de la thèse en cours de Jean Le Maguet intitulée : « Etude de la transmission des virus associés à l'enroulement de la vigne par les cochenilles ».

## II - Contexte de l'étude : éléments de biologie

### II-1- L'enroulement viral de la vigne

#### II-1-1- Généralités

L'enroulement viral de la vigne est présent dans toutes les régions viticoles du monde. Les virus responsables de cette maladie sont appelés « Grapevine leafroll associated virus » ou GLRaV. Ces virus filamenteux de la famille des *Closteroviridae* (Fig. 1) sont inféodés aux vaisseaux du phloème qu'ils obstruent, empêchant ainsi les produits de la photosynthèse de parvenir des feuilles jusqu'aux baies. À ce jour, 8 espèces d'enroulement ont été identifiées, les GLRaV-1, -2, -3, -4, -5, -6, -7, et -9.

Dans les vignobles septentrionaux (Alsace, Champagne et Bourgogne) seules les espèces d'enroulement 1 et 3 sont présentes.

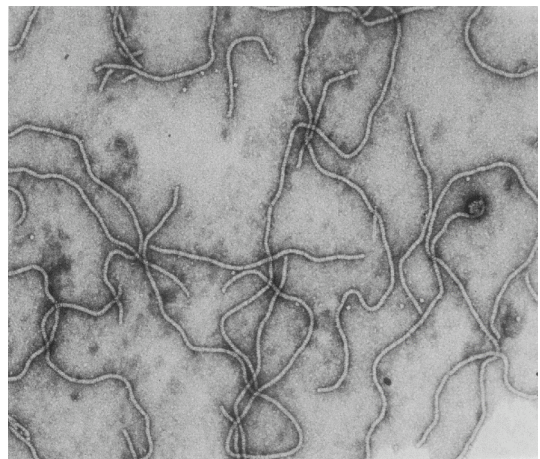


Fig. 1 : photographie d'un *Closteroviridae* (John Antoniwick-ICTV)

## II-1-2- Symptômes de la maladie

Dans le cas de l'enroulement viral, les principaux symptômes observés sont un enroulement des limbes des feuilles vers l'intérieur ainsi que des modifications pigmentaires des feuilles. Ces symptômes sont décelables dès le mois de juin pour les céps les plus touchés, mais la période la plus favorable pour les observations est l'automne (Le Maguet, 2010). Pour les cépages rouges (pinot meunier, pinot noir) on distingue un rougissement du limbe des feuilles, les nervures restant vertes (Fig. 2). Pour les cépages blancs (riesling, chardonnay) les feuilles s'éclaircissent ou jaunissent, les nervures restant également vertes (Fig. 2bis).



*Fig. 2 : symptômes de l'enroulement sur pinot meunier (JLM)*



*Fig. 2bis : symptômes de l'enroulement sur un cépage blanc (INRA)*

## II-1-3- Conséquences sur le rendement et la qualité des moûts

La présence de l'enroulement viral dans une vigne se traduit par une baisse de la photosynthèse engendrant par la suite une diminution de la quantité de raisins produits. La maladie a des répercussions différentes selon les espèces et les formes de GLRaVs impliqués, et selon les cépages touchés. La baisse de rendement est en moyenne de 15 à 30 % et peut aller jusqu'à 60% pour les parcelles les plus touchées (Boudon-Padieu et al., 2000). La diminution de la production de raisin s'accompagne également d'une moindre qualité des baies. L'enroulement viral n'est pas un virus létal pour la plante, mais une vigne fortement touchée représente une perte sur le long terme.

## II-1-4- Transmission du virus

Il existe deux modes de transmission du virus sur un plant de vigne :

- Transmission par multiplication de matériel végétal virosé : le virus peut être transmis lors de la greffe : pour produire de nouveaux plants de vigne on greffe un greffon sur un porte-greffe. Si le greffon ou le porte-greffe est porteur du virus, le plant obtenu sera virosé.
- Transmission par un organisme vecteur : la seule solution possible pour qu'un plant sain acquière le virus est la transmission par un vecteur. Dans le cas de l'enroulement, ce vecteur correspond à des insectes, les cochenilles.

## II-2- Les cochenilles sur la vigne

Les cochenilles (Fig. 3) sont des insectes faisant partie de l'ordre des homoptères dans lequel on retrouve d'autres ravageurs bien connus: pucerons, psylles et aleurodes. Les cochenilles se nourrissent de la sève des plantes et peuvent ainsi transmettre des particules virales contenues dans leur salive pendant leur prise de nourriture dans les tissus végétaux. On compte 9 espèces de cochenilles vivant sur la vigne en France (Sforza, 2000). Dans les régions septentrionales, les cochenilles n'ont qu'un seul cycle par an. Les espèces vectrices de l'enroulement retrouvées en France sont :

- *Parthenolecanium corni*
- *Pulvinaria vitis*
- *Neopulvinaria innumerabilis*
- *Planococcus citri*
- *Planococcus ficus*
- *Heliococcus bohemicus*
- *Phenacoccus aceris*



Fig. 3 : femelles de *Phenacoccus aceris* (JLM)

## III – Présentation de la problématique

### III-1- La parcelle modèle de Bonzon

#### III-1-1- Descriptif

Un suivi épidémiologique a été conduit sur une parcelle de pieds mères de greffons, située dans le vignoble du Mâconnais (Bonzon, Saône-et-Loire) et plantée en 2001 avec du matériel végétal de base certifié sain (pinot noir clones 667, 828 et 115 sur porte-greffe 3309). Les premiers symptômes d'enroulement sont apparus dès 2003. Les tests de détection de virus entrepris par la Chambre d'Agriculture de Saône-et-Loire de 2003 à 2005 ont permis d'identifier la présence de ceps contaminés par le GLRaV-1.

Ainsi, depuis 2003, les 2800 ceps de la parcelle ont fait l'objet chaque automne d'un suivi annuel des symptômes. Un relevé de symptômes a également été conduit en 2008 sur les parcelles contiguës de gauche (BzG) et de droite (BzD) (Fig. 4)

La répartition des cochenilles sur la parcelle a été suivie depuis 2006 par un contrôle individuel des ceps au printemps. Les pousses, l'écorce et les rameaux de chaque cep ont ainsi été inspectés pour évaluer la quantité d'individus des différentes espèces de cochenilles présentes.



### III-1-2- Épidémiologie de la parcelle

On constate une progression fulgurante et constante de la prévalence de l'enroulement sur la parcelle de Bonzon. Le pourcentage de plantes infectées croît de 5% en 2004 à 70% en 2009 (Fig. 5). La répartition spatiale des ceps infectés est représentée sur une cartographie où chaque case correspond à cep de vigne (Fig. 6). Sur cette cartographie, la couleur de chaque cep symptomatique correspond à l'année de première expression des symptômes. On peut observer que l'épidémie s'étend depuis trois foyers principaux, situés aux bordures, le long des parcelles voisines contiguës. L'extension de ces trois foyers est constante et seule la partie ouest de la parcelle reste à l'écart de la progression de l'épidémie.

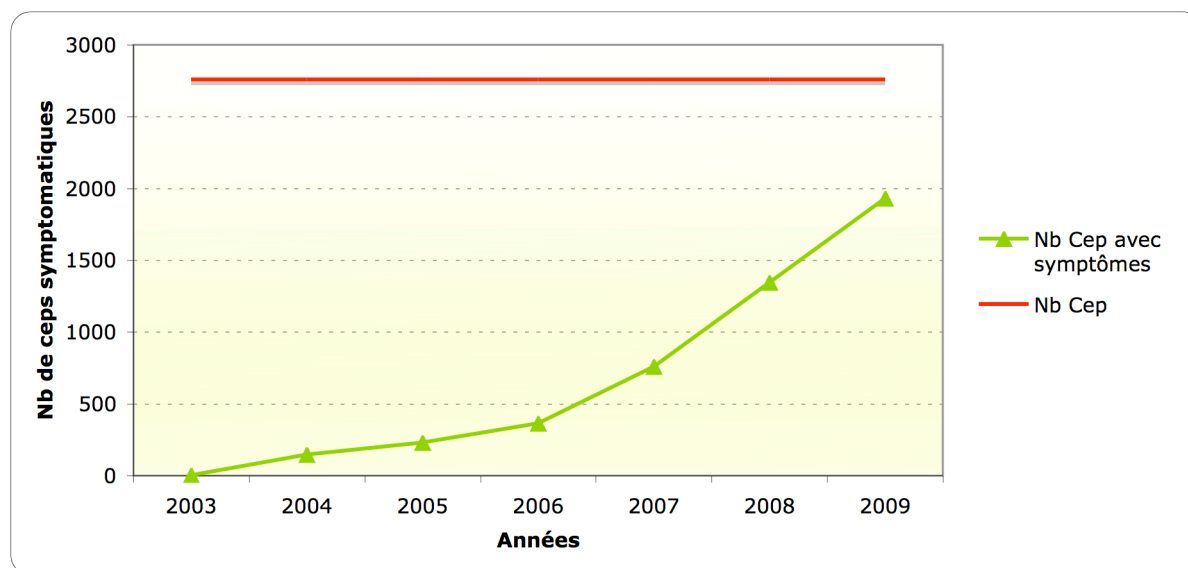


Fig. 5 : évolution de la prévalence de l'enroulement viral sur la parcelle de Bonzon entre 2003 et 2009

Deux espèces de cochenilles farineuses (*P. aceris* et *H. bohemicus*) et deux espèces de cochenilles à bouclier (*P. corni* et *P. vitis*) ont été observées durant les quatre années de notation. *P. aceris* est l'espèce de cochenille la plus abondante et a infesté 50 % des ceps en 2009. Elle est également très abondante dans les deux parcelles contiguës (Le Maguet et al., 2009)

On peut remarquer en cartographiant les ceps que la distribution spatiale des zones infestées par les cochenilles suit de manière sensible celle des zones atteintes d'enroulement, notamment au niveau des trois foyers principaux (Fig. 7). **Ainsi, la progression rapide des symptômes d'année en année sur la parcelle d'étude semble être directement liée à la dispersion du GLRaV-1 par *P. aceris*.**

Ce constat constitue ainsi la base de l'analyse spatiotemporelle qui va suivre.

## III-2- Démarche de l'analyse

### III-2-1- Jeu de données à analyser

Pour mener l'analyse, on dispose de données relatives aux symptômes et aux cochenilles.

#### **Pour les symptômes**

- de 2003 à 2009 : notations de la présence et de sévérité des symptômes pour chaque cep
- pour l'année 2008 : une notation de la présence pour les deux parcelles contiguës

#### **Pour les cochenilles :**

- de 2006 à 2008 : notations de la présence des cochenilles (par espèces ou pour toutes les espèces), des estimations du nombre de cochenilles par cep.

Pour montrer que l'infestation de la parcelle par *P. aceris* est la cause de la progression de l'enroulement viral, une analyse spatiotemporelle en trois étapes a été conduite.

### III-2-2- Raisonnement

On a constaté une progression rapide de l'épidémie d'enroulement sur la parcelle, principalement à partir de trois foyers de contamination principaux. De même, il semble que les observations indiquent que l'espèce *P. aceris* infeste les ceps également sur trois zones principales.

**Étape 1 :** Y a-t-il une agrégation des ceps symptomatiques ?  
Y a-t-il une agrégation des ceps infestés par les cochenilles ?

On observe ensuite que les foyers correspondant aux zones les plus contaminées par l'enroulement et aux zones les plus infestées par *P. aceris* semblent se superposer.

**Étape 2 :** Y a-t-il une corrélation entre la présence de la maladie et la présence de cochenilles (suivant les années, les espèces) ?

La parcelle de Bonzon a été implantée dans un vignoble préexistant déjà largement contaminé par l'enroulement viral. Les parcelles voisines du vignoble d'étude, plantées en pinot noir ou en chardonnay, sont toutes contaminées par les GLRaV-1 et/ou -3. En outre, les observations de symptômes ont montré que les deux parcelles contiguës présentent une prévalence importante de l'enroulement avec près de 100% de ceps contaminés pour la parcelle située à droite.

Avant sa plantation en 2001, la surface de la parcelle est restée sans culture pendant plus d'une dizaine d'années (prairie de fauche), conformément à la réglementation en vigueur pour les parcelles de pieds mères de greffons. Les espèces de cochenilles observées n'infestant que les plantes ligneuses, on peut exclure toute infestation préalable ou toute contamination du sol avant la plantation de la parcelle. En outre, des tests de détection de virus entrepris sur les parcelles mères ayant fourni les clones de pinot noir et les porte-greffes ont permis de confirmer l'absence d'enroulement dans le matériel végétal à la plantation. La contamination post-plantation à partir du vignoble environnant se trouvant ainsi confirmée, la présence d'un organisme vecteur dispersant l'enroulement d'une parcelle à l'autre constitue l'hypothèse la plus vraisemblable pour expliquer la situation actuelle.

**Étape 3 :** L'épidémie provient-elle des parcelles contiguës ?



## IV – Etude pratique

### IV-1- Analyse de l'agrégation des symptômes et des cochenilles dans la parcelle

#### IV-1-1- Outils et méthodes

Dans un premier temps j'ai procédé à des recherches bibliographiques sur les outils et méthodes statistiques utilisables pour mener à bien l'analyse spatiotemporelle des données. Nous avons décidé d'utiliser le logiciel de traitement statistique R, car il est relativement complet et facile d'accès de part sa gratuité (Cornillon et al., 2010, [www.r-project.org](http://www.r-project.org)).

En ce qui concerne les méthodes statistiques, mes premières recherches m'ont mené vers des méthodes d'analyse statistique spatiale telles que la fonction K de Ripley, le calcul de variogramme ou semi-variogramme empirique.

J'ai donc dans un premier temps exploré les packages de R nommés « spatstat » et « geoR ».

Quelques essais ont été réalisés à l'aide de « spatstat ». L'idée était de considérer les ceps malades de la parcelle modèle comme des processus ponctuels spatiaux et de tester la répartition de ces ceps avec comme hypothèse nulle une répartition aléatoire issue d'un processus de Poisson. Les premiers résultats obtenus avec les fonctions du package (ex : test de Kolmogorov-Smirnov, « quadrat counting tests », graphique de la fonction K de Ripley).

Cependant il s'est avéré que ces méthodes ne sont pas exactement adaptées à nos données, les résultats obtenus étaient donc faussés. Toutes ces méthodes consistent en effet à comparer à chaque fois la structure de nos données à une structure totalement aléatoire (processus de Poisson). Or en aucun cas la structure de la répartition des ceps symptomatiques dans la parcelle ne peut être totalement aléatoire, du fait de la plantation de celle-ci sous forme de grille rectangulaire à maille régulière le long des rangs.

Une autre méthode consistant à étudier un jeu de données sur une grille a été ainsi envisagée (Peyrard et al., 2005). Les auteurs présentent un certain nombre d'outils d'analyse statistique spatiale à l'aide de tests de permutation. J'ai donc utilisé certaines de ces méthodes pour analyser les données de l'épidémiologie de l'enroulement de la vigne.

En ce qui concerne l'implémentation sous R, j'ai récupéré les fonctions relatives aux exemples de la publication auprès d'un des auteurs. Certaines ont été réutilisées pour nos propres données, d'autres modifiées ou bien créées. Pour le reste, certaines fonctions classiques de R auront été utiles (voir ANNEXES 1, 2, 3).

## IV-1-2- Utilisation du variogramme et des tests de permutations

On considère un jeu de données schématisé sous forme de grille régulière contenant I lignes et J colonnes. Les données sont recueillies aux nœuds de la grille. Si on appelle X notre grille, on notera  $X_{(i,j)}$  la valeur de X à la ligne i et la colonne j, et d le pas de déplacement (Fig. 8).

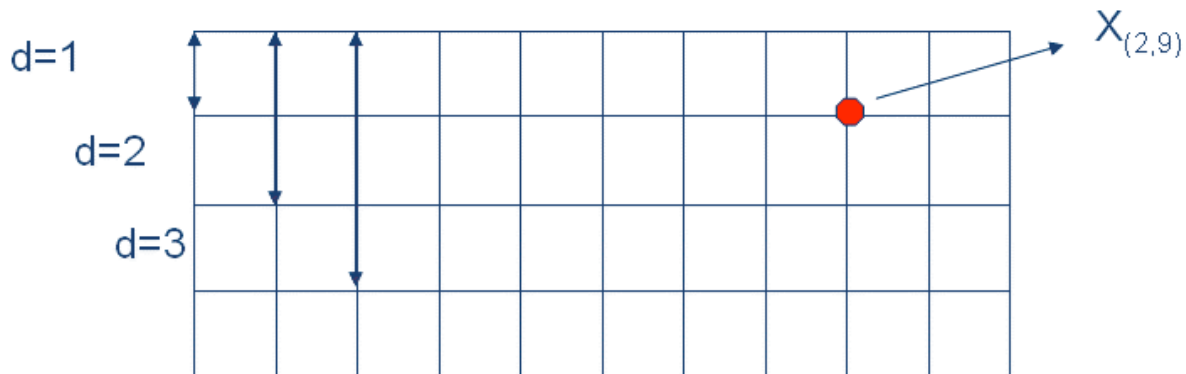


Fig. 8 : exemple d'une grille de donnée à maille régulière

On cherche à détecter une structure spatiale dans notre grille, une agrégation des données de même valeur. On va effectuer des tests de permutations pour tester l'indépendance des données par rapport à leur placement sur la grille.

Le **principe des tests de permutation** est de travailler à partir des seules données observées, de façon à éviter de formuler des hypothèses sur la distribution de la variable mesurée. Ayant une statistique de test, on la calcule sur les données observées puis on la recalcule pour chacune des permutations (permutation = redistribution des données au hasard). Le point clé de la méthode est que sous l'hypothèse nulle toute permutation des données observées est équiprobable dans l'ensemble des permutations correspondant à l'hypothèse testée.

Dans notre contexte, les critères servant de statistique de test correspondent au variogramme et au covariogramme, et les différents types de permutation dépendent des hypothèses que l'on souhaite tester. On peut alors tracer sur un graphique, le variogramme (ou le covariogramme) des données observées, ainsi qu'une enveloppe de confiance déterminée pour un nombre N de permutations et un seuil  $\alpha=5\%$ .

On distingue plusieurs types de tests de permutation : dans cette partie, on s'intéresse au test d'indépendance totale et au test d'indépendance entre lignes.

### IV-1-2-1- Test d'indépendance totale

On effectue un test de permutation totale lorsqu'on veut étudier de manière globale s'il y a agrégation des données. Plus précisément on veut savoir si les points  $X_{(i,j)}$  qui sont spatialement proches ont des valeurs similaires (ou très proches).

Le principe du test de permutation totale est de mélanger aléatoirement toutes les valeurs des  $X_{(i,j)}$  pour tout i, j et pour un nombre N de permutations. On utilise le critère du variogramme le long de la ligne. La formule du variogramme le long de la ligne est la suivante :

$$C(d) = \frac{1}{I(J-d)} \sum_{i=1}^I \sum_{j=1}^{J-d} (X_{(i,j)} - X_{(i,j+d)})^2$$

Remarque : on choisit d'utiliser un variogramme le long de la ligne ou entre les lignes suivant le cadre de l'étude. Dans notre cas, on privilégiera celui le long de la ligne.

#### IV-1-2-2- Test d'indépendance entre les lignes

Lorsque l'on a montré qu'il existe une structure spatiale à l'aide d'un test de permutation totale, on peut se demander s'il y a une dépendance entre les lignes. Dans ce cas on utilisera le variogramme entre les lignes, dont la formule mathématique est la suivante :

$$C(d) = \frac{1}{J(I-d)} \sum_{j=1}^J \sum_{i=1}^{I-d} (X_{(i,j)} - X_{(i+d,j)})^2$$

Le principe est d'effectuer des permutations sans changer la structure au sein même d'une ligne. Pour cela, on peut effectuer deux types de tests de permutation :

- Test d'indépendance des lignes par permutations aléatoires des lignes entre elles
- Test d'indépendance des lignes par rotations aléatoires des lignes

#### Principe de permutation des lignes :

On permute les lignes de la grille entre elles de manière aléatoire. La structure des données au sein même d'une grille reste alors inchangée (Fig. 9).

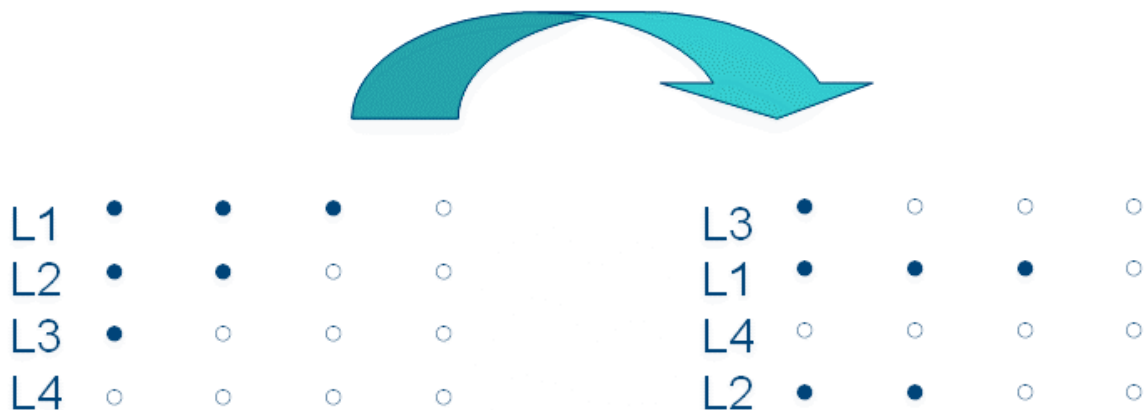


Fig. 9 : schéma pour une grille de 4 lignes (avant et après la permutation):

### Principe de rotation des lignes :

On effectue une rotation de pas aléatoire et indépendant d'une ligne à l'autre. De nouveau, la structure au sein même d'une ligne reste inchangée (Fig. 10).

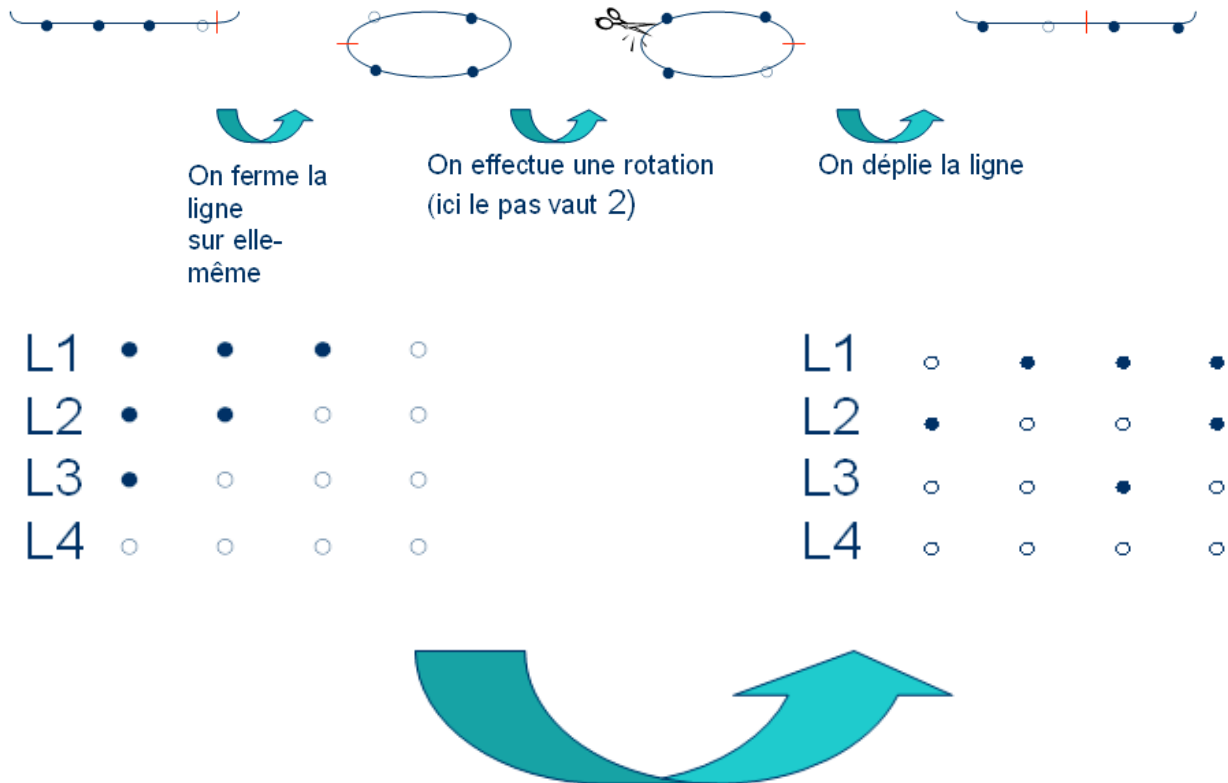


Fig. 10 : (haut) schéma explicatif de rotation d'une ligne (bas) schéma d'une translation par rotation sur une grille de quatre lignes :

### IV-1-3- Étude de l'agrégation des symptômes dans la parcelle modèle

On considère notre parcelle modèle, constituée de 18 rangs contenant 156 ceps chacun (Fig. 11). On veut tester l'indépendance spatiale des ceps porteurs de symptômes de l'enroulement viral de la vigne. Pour cela, on code nos données de telle manière qu'elles soient compatibles avec les méthodes d'analyse précédentes. On note avec un « 0 » un cep sans symptôme et avec un « 1 » un cep présentant des symptômes.

L'analyse des données de symptômes a ainsi été conduite pour chaque année mais par un souci de concision, seuls les résultats de l'analyse pour l'année 2009 (Fig. 12) sont présentés ici.



Fig. 12 : représentation graphique sur R de la répartition des ceps symptomatiques pour l'année 2009

#### IV-1-3-1- Test d'indépendance par permutation totale des ceps

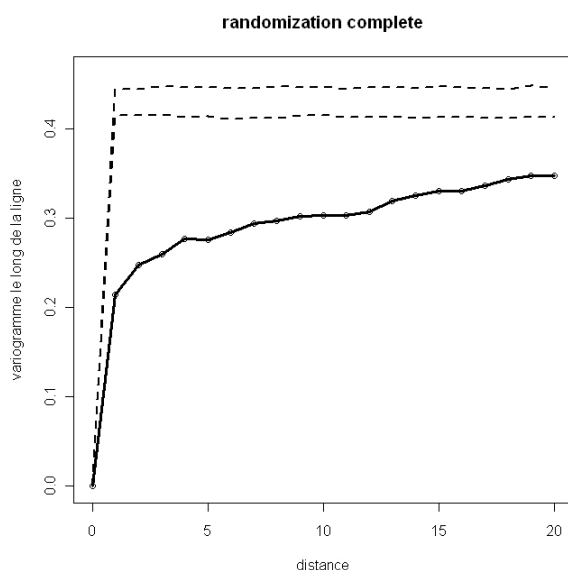


Fig. 13 : test d'indépendance totale des données symptômes pour l'année 2009. La courbe en trait plein correspond aux valeurs observées et les courbes en pointillés correspondent aux limites de confiance à 2.5% et 97.5%

On remarque que quelque soit le pas de distance  $d$ , le variogramme est nettement inférieur à l'enveloppe d'indépendance (Fig. 13). L'écart se réduit quand la distance augmente. On en déduit qu'il y a une structure dans la parcelle : **il existe donc une dépendance spatiale entre les ceps symptomatiques.**

#### IV-1-3-2- Test d'indépendance inter-rang

En observant la vue d'ensemble de la présence de symptômes (Fig. 11), on remarque que les symptômes se déclarent d'abord sur les bords (gauche et droite) puis se propagent entre les rangs vers le cœur de la parcelle. Pour cette raison, nous allons préférer un test de permutation par rotation pour éviter de casser la structure inter-rang.

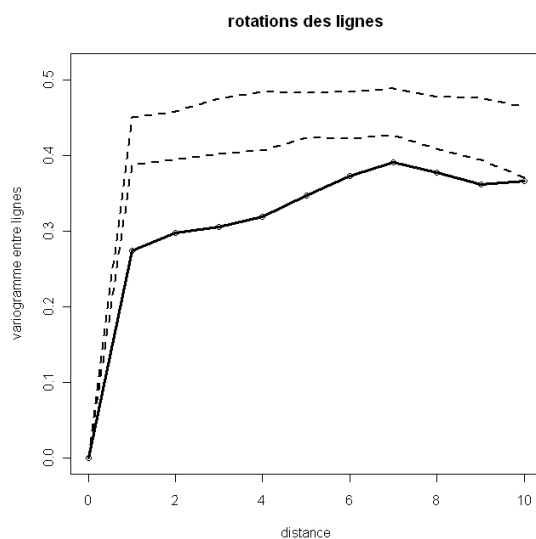


Fig. 14 : test d'indépendance par rotation des rangs pour les données « symptômes » en 2009

On remarque qu'il y a une dépendance entre les rangs jusqu'à  $d=6$  (Fig. 14) ce qui s'explique clairement par la forme des foyers de symptômes qui globalement ne dépasse pas un tiers de la parcelle (dans le sens inter-rang).

#### IV-1-4- Étude de l'agrégation des cochenilles dans la parcelle modèle

On procède de la même manière que dans la partie précédente pour tester l'indépendance de la répartition des cochenilles dans la parcelle. On note avec un « 1 » un cep infesté par des cochenilles, avec un « 0 » un cep sans cochenille. De plus, on va faire la même analyse en ne considérant que l'espèce *P. aceris*, car elle est l'espèce la plus largement représentée et on considèrera par la suite que les autres espèces sont négligeables.

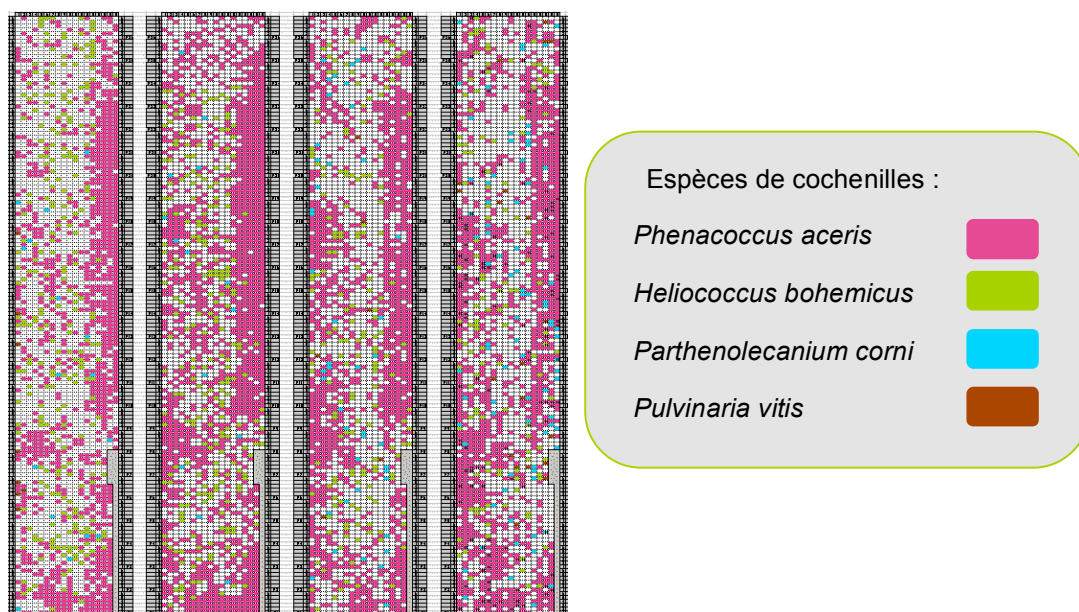


Fig. 15 : cartographies de la parcelle modèle de Bonzon, représentant la répartition des ceps infestés par les cochenilles de 2006 à 2009

##### IV-1-4-1- Test d'indépendance totale des ceps des données « cochenilles 08 » et « P. aceris 08 »

L'analyse des données de cochenille a été conduite pour chaque année, les résultats de l'année 2008 sont présentés ici.

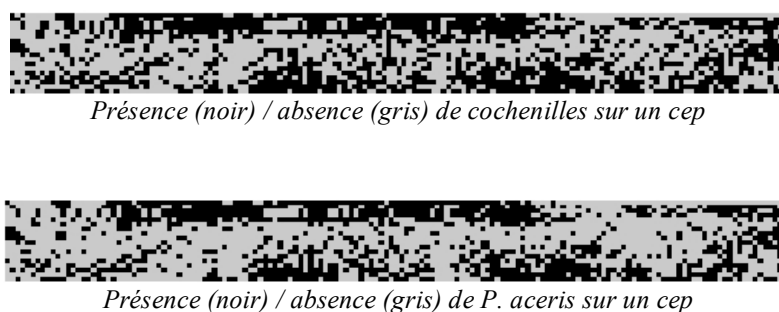


Fig. 16 : représentation graphique sur R de la répartition des ceps infestés par les quatre espèces de cochenilles (haut) et par *P. aceris* seulement (bas)

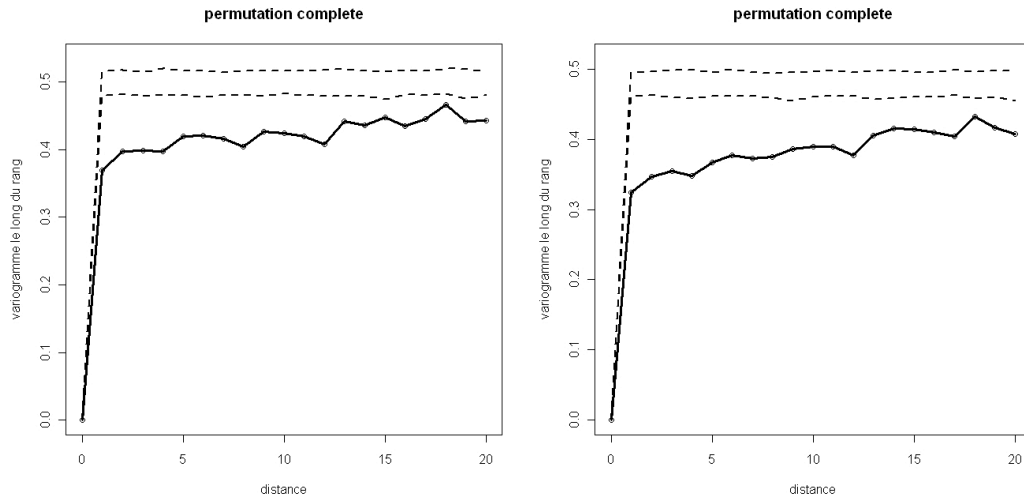


Fig. 17 : test d'indépendance par permutation complète pour les données « cochenilles 2008 » (à gauche) et « *P. aceris* 2008 » (à droite)

Le test d'indépendance montre bien que sur les données concernant uniquement l'espèce *P. aceris*, la courbe de variogramme est légèrement inférieure à celle prenant en compte toutes les espèces. On remarque en effet sur les cartes de présence qu'il y a **agrégation de la présence de cochenilles dans la parcelle** (Fig. 17). On constate qu'il y a moins de données éparpillées pour *P. aceris*, les ceps infestés par cette espèce formant des paquets plus homogènes. C'est pour cela que la courbe du variogramme est inférieure.

#### IV-1-4-2- Test d'indépendance spatiale entre les rangs

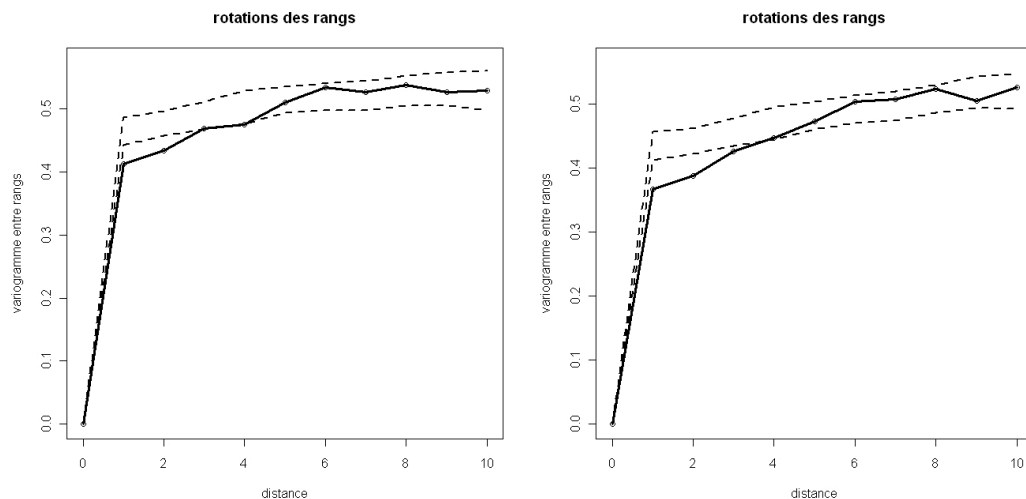


Fig. 18 : test d'indépendance par rotation des rangs pour « cochenilles 2008 » (gauche) et pour « *P. aceris* 2008 » (droite)

Le test d'indépendance spatiale entre les rangs n'est significatif que pour les petites distances. Plus précisément pour  $d=1, 2, 3, 4$ . Cependant, il est un peu limite pour  $d=3, 4$ . Au-delà, la courbe du variogramme sur les données observées est dans l'enveloppe d'indépendance. **Il y a donc une dépendance entre les rangs pour une distance inférieure à  $d=5$ .**

Ceci s'explique par la taille des groupes de ceps infestés par les cochenilles, ceux-ci sont d'une longueur assez faible lorsque l'on regarde dans la direction inter-rang (Fig. 16), en notant que le critère calcule les différences entre les rangs.

La notion d'agrégation des ceps infestés par les cochenilles indique qu'il existe une dépendance spatiale entre eux. Cependant, il est nécessaire d'apporter une explication d'ordre biologique pour pouvoir utiliser le terme foyer. Les cochenilles sont des insectes aptères (sans ailes) et se dispersent principalement de proche en proche sur les plantes. Dans un vignoble, elles se déplacent préférentiellement le long du rang (fils de conduite des vignes). Ainsi, compte tenu des caractéristiques de dispersion des cochenilles et des résultats obtenus dans l'analyse précédente, on peut conclure que les ceps infestés par *P. aceris* dans la parcelle modèle sont agrégés en foyer.

Remarque : Les quatre espèces de cochenille présentes sur la parcelle modèle sont connues pour être capable de transmettre l'enroulement viral. Cependant, *P. aceris* étant l'espèce majoritaire sur la parcelle, les tests suivants seront effectués seulement à partir des données sur cette espèce.

## IV-2- Corrélation entre la présence de *P. aceris* et de l'enroulement viral dans la parcelle modèle

### IV-2-1- Test d'indépendance dans le cadre non-spatial

Dans un premier temps, on se place dans un cadre non-spatial afin d'étudier s'il y a une dépendance entre le nombre de ceps infestés par *P. aceris* et le nombre de ceps infectés par la maladie de l'enroulement viral.

Connaissant le nombre de ceps infestés par *P. aceris* et celui présentant des symptômes pour les années 2006 à 2009, nous allons utiliser un test de  $\chi^2$  pour tester l'indépendance (Confais et al., 2005).

Dans notre cas, les données se résument en un tableau de contingence à deux lignes et deux colonnes, ce qui simplifie la formule de calcul du  $\chi^2$ .

Rappelons la formule dans le cas d'une table 2x2:

$$\chi^2 = N \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

avec :

$$\left\{ \begin{array}{l} n_{ij} : \text{effectif de la case ligne } i \text{ et colonne } j \\ n_{i.} : \text{effectif total sur la ligne } i \\ n_{.j} : \text{effectif total sur la colonne } j \\ N : \text{effectif total} \end{array} \right.$$



On obtient les résultats suivants pour chaque année (analyse avec logiciel R) :

#### ANNEE 2006

| SYMPTOMES<br>COCHENILLES ( <i>P. aceris</i> ) | PRESENCE | ABSENCE |
|---|----------|---------|
| PRESENCE                                      | 214      | 744     |
| ABSENCE                                       | 150      | 1649    |

$\chi^2 = 105.7$   
**p-valeur = 0.00**

#### ANNEE 2007

| SYMPTOMES<br>COCHENILLES ( <i>P. aceris</i> ) | PRESENCE | ABSENCE |
|---|----------|---------|
| PRESENCE                                      | 484      | 726     |
| ABSENCE                                       | 150      | 1273    |

$\chi^2 = 308.8$   
**p-valeur = 0.00**

#### ANNEE 2008

| SYMPTOMES<br>COCHENILLES ( <i>P. aceris</i> ) | PRESENCE | ABSENCE |
|---|----------|---------|
| PRESENCE                                      | 717      | 393     |
| ABSENCE                                       | 629      | 1018    |

$\chi^2 = 183.96$   
**p-valeur = 0.00**

#### ANNEE 2009

| SYMPTOMES<br>COCHENILLES ( <i>P. aceris</i> ) | PRESENCE | ABSENCE |
|---|----------|---------|
| PRESENCE                                      | 921      | 212     |
| ABSENCE                                       | 1009     | 615     |

$\chi^2 = 115.74$   
**p-valeur = 0.00**

Ainsi, pour les quatre années, on obtient une « p-valeur » approximativement nulle, donc on rejette significativement l'hypothèse d'indépendance (au seuil de risque  $\alpha = 5\%$ ).

On peut remarquer que les valeurs du  $\chi^2$  sont à chaque fois très élevées. Pour rappel le quantile d'une loi du khi<sup>2</sup> à un degré de liberté au seuil de risque 5% vaut 3.84.

## IV-2-2- Test d'indépendance dans le cadre spatial

Maintenant que l'on a montré qu'il y a un lien entre la présence de cochenilles (*P. aceris*) et la présence de symptômes dans notre parcelle modèle, on se propose d'étudier s'il y a une dépendance dans le cadre spatial. Pour cela, on va de nouveau utiliser les tests de permutation.

### IV-2-2-1- Test d'indépendance spatiale de deux grilles par translations

On nomme X et Y deux grilles et les données correspondant aux mêmes points de ces deux grilles. Si l'une de ces grilles ne présente aucune structure, alors on considère qu'elles sont indépendantes. Sinon, on peut tester l'indépendance entre deux grilles. Le principe général d'un test d'indépendance entre deux grilles repose sur le fait que, même s'il existe une structure spatiale dans chaque image, ces structures sont disposées au hasard l'une par rapport à l'autre. S'il y a une structure spatiale dans les grilles, on va chercher à la conserver dans le test, sinon ce dernier mélangera test d'indépendance entre grilles et test d'indépendance entre deux points de la même grille. On va donc effectuer un test de permutation en utilisant la méthode de translation. On va ainsi traduire l'une des deux images pour la comparer à l'autre en utilisant le critère du covariogramme, dont la formule est donnée ci-dessous :

$$C(d) = \frac{1}{I(J-d)} \sum_{i=1}^I \sum_{j=1}^{J-d} (X_{(i,j)} - X_{(i,j+d)})(Y_{(i,j)} - Y_{(i,j+d)})$$

On effectue une **translation de la grille** le long de la ligne, de la colonne ou les deux. Plus précisément, on découpe la grille en deux dans l'un ou l'autre sens suivant le cas, puis on translate les deux parties de la grille obtenue (Fig. 19)

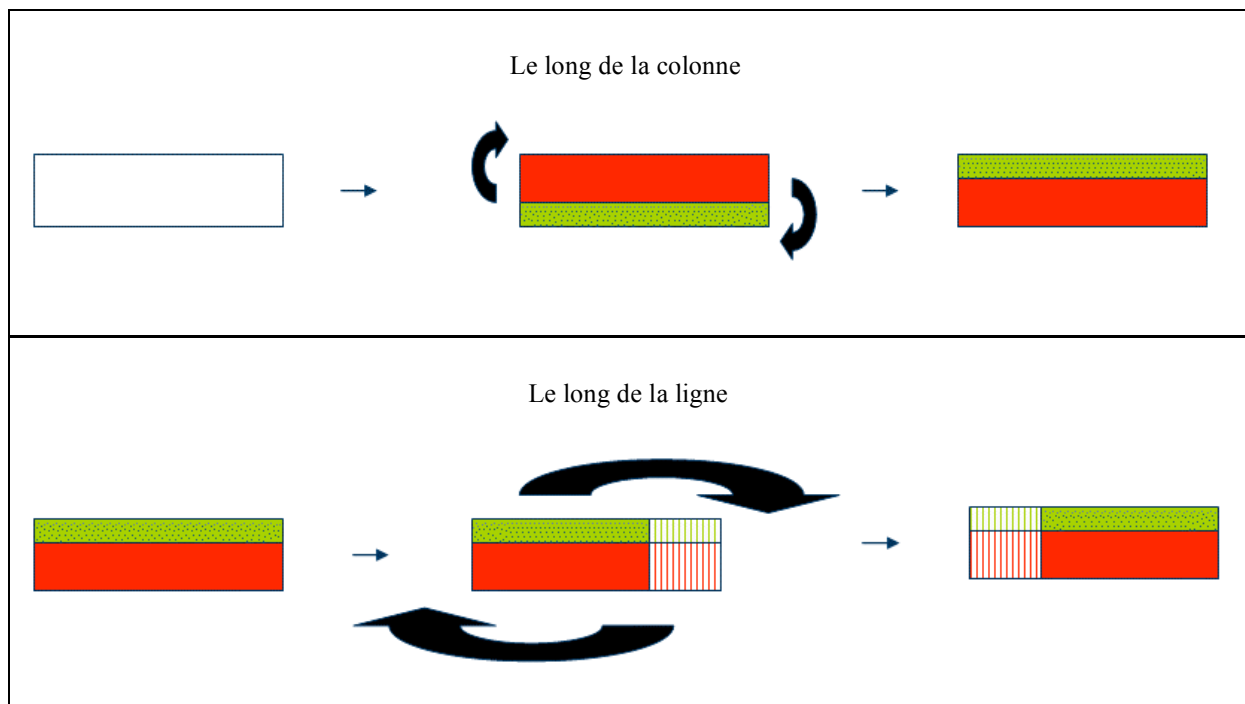
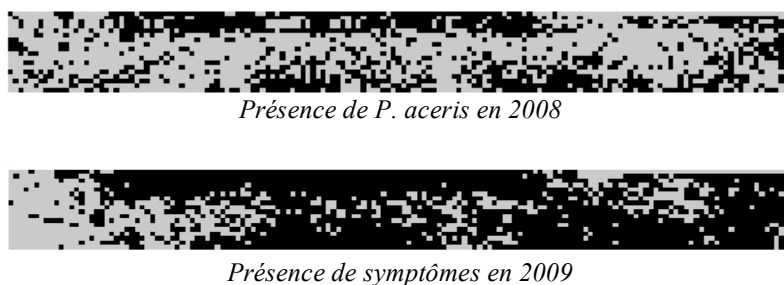


Fig. 19 : schéma explicatif d'une translation dans les deux sens

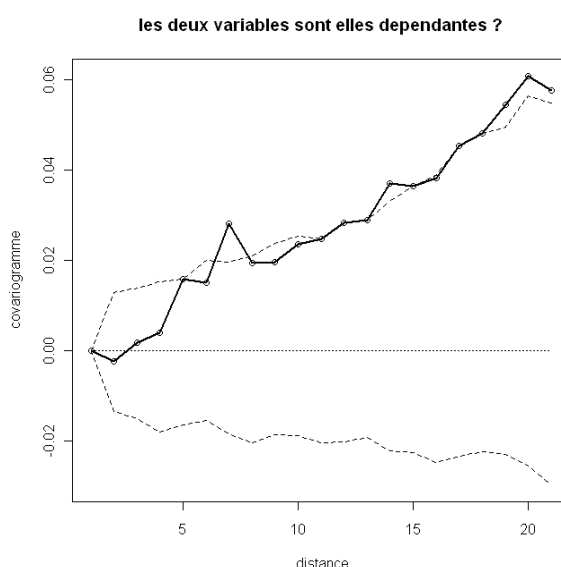
Il y a plusieurs possibilités pour tester l'indépendance entre deux grilles. Si les deux grilles possèdent des structures dans les deux sens, on effectuera les translations dans les deux sens, si elles ne possèdent qu'une structure dans un sens, on effectuera la translation que dans un sens. On pourra de plus effectuer toutes les translations possibles ou bien juste un nombre N de permutations.

*IV-2-2-2- Test d'indépendance spatiale entre la présence de symptômes et de P. aceris*

On a montré précédemment que les données « symptômes2009 » et « cochenilles2008 » sont structurées en foyer. On va donc tester l'indépendance entre les deux grilles à l'aide d'un test de permutation en utilisant la méthode de translation le long du rang (Fig. 19 : translation le long de la ligne). Les foyers se développant des bordures vers le cœur de la parcelle, nous ne voulons pas casser cette structure. On effectuera toutes les translations le long du rang, soit un total de 156 translations. On va donc traduire la grille correspondant aux données « P. aceris 2008 » et la grille correspondant aux données « symptômes2009 » en utilisant le critère du covariogramme le long du rang (Fig. 20).



*Fig. 20 : représentation graphique sur R de la répartition des cepts infestés par P. aceris en 2008 (haut) et de la présence de symptômes en 2009 (bas)*



*Fig. 21 : test d'indépendance par translation des données « P. aceris 2008 » et « symptômes 2009 »*

On remarque que la courbe du covariogramme est dans l'enveloppe d'indépendance pour les courtes distances. Pour  $d=7$  et à partir de  $d=11$  on observe un rejet de l'hypothèse d'indépendance. Donc à partir de  $d=11$ , on peut dire qu'il y a une dépendance entre les structures de la présence de symptômes et de la présence de *P. aceris*. Pour comprendre pourquoi la courbe du covariogramme est dans l'enveloppe d'indépendance pour les petites distances, il faut remarquer que les groupes de ceps infestés qui concordent avec les groupes de ceps infestés sont généralement d'assez grande taille, d'autant plus que le critère compare les données le long des rangs (Fig. 20).

### IV-2-3- Modélisation

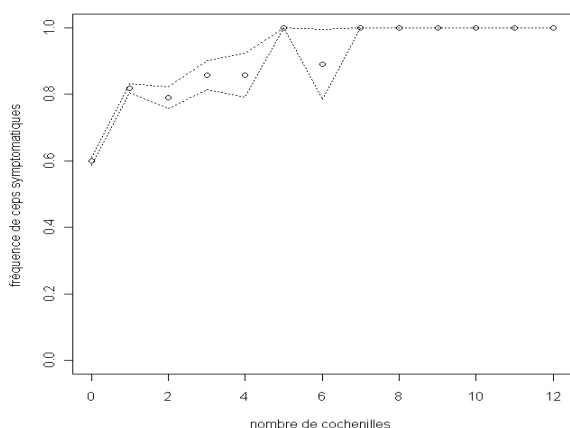
Nous avons montré précédemment qu'il existe un lien entre la présence de l'espèce de cochenille *P. aceris* et la présence de symptômes d'enroulement viral dans la parcelle modèle. Pour aller plus loin, on souhaite mettre au point un modèle visant à **expliquer la présence des symptômes par la présence de cochenilles**. On dispose aussi de données quantitatives pour les cochenilles (*P. aceris*). Plus précisément, pour chaque cep de la parcelle on dispose d'une approximation du nombre de cochenilles présentes. On parle d'une approximation car il est difficile voire impossible de connaître le nombre exact d'insecte infestant un cep, étant donnée la difficulté que représente l'observation des cochenilles sur le terrain (il peut y en avoir aussi sous l'écorce du cep par exemple).

Ainsi, nous allons faire l'hypothèse que **l'approximation du nombre de cochenilles représente le réel**. On suppose donc une proportionnalité de cette approximation. (par exemple : 4 cochenilles observées pourraient représenter une quarantaine en réalité, 10 en représenterait une centaine).

On conduit l'étude sur les données correspondant aux données de *P. aceris* en 2008 et la présence de symptômes en 2009. Comme précédemment, on choisit volontairement une année d'écart car on considère que la durée de latence de la maladie après inoculation par l'insecte est d'un an.

#### IV-2-3-1- Régression logistique

Les valeurs du nombre de cochenilles par cep sont comprises entre 0 et 12. Pour cela, on va dans un premier temps calculer les fréquences de ceps symptomatiques (les ceps malades) en fonction du nombre de cochenilles présentes sur celui-ci (Fig. 22).



- abscisse : nombre  $k$  de cochenilles
- ordonnée : fréquence  

$$F = \frac{\text{nombre ceps malades}}{\text{nombre de ceps}}$$
- enveloppe : + ou - l'écart-type de  $F$

Fig. 22 : fréquence du nombre de ceps symptomatiques pour chaque nombre  $k$  de cochenilles

A partir de ces fréquences, nous pouvons déduire les probabilités conditionnelles de notre modèle. Plus précisément, notons X la variable aléatoire représentant les cochenilles prenant ces valeurs dans  $\{0,1,\dots,12\}$  et Y celle représentant la présence de symptômes prenant ces valeurs dans  $\{0,1\}$ . On note  $p$  la probabilité conditionnelle suivante :

$$\left( \begin{array}{l} p = P(Y=1 | X=x) \\ 1-p = P(Y=0 | X=x) \end{array} \right)$$

On obtient une loi de probabilité. Chaque valeur k du nombre de cochenilles est associée à une valeur de probabilité  $p$  (Fig. 23).

| k   | 0     | 1     | 2     | 3     | 4     | 5 | 6     | 7 | 8 | 9 | 10 | 11 | 12 |
|-----|-------|-------|-------|-------|-------|---|-------|---|---|---|----|----|----|
| $p$ | 0.599 | 0.817 | 0.789 | 0.857 | 0.857 | 1 | 0.889 | 1 | 1 | 1 | 1  | 1  | 1  |

Fig. 23 : tableau des probabilités  $p$  en fonction de  $k$

A la vue des données, on décide d'ajuster un modèle de régression logistique (Rakotomalala, 2009). L'équation du modèle est la suivante :

$$\ln\left(\frac{p}{1-p}\right) = a + bX$$

Avec  $\left( \begin{array}{l} p = P(Y=1 | X=x) \\ 1-p = P(Y=0 | X=x) \\ a, b \text{ les coefficients à estimer} \end{array} \right)$

On va tout d'abord tracer la fonction  $p \rightarrow \log(p/(1-p))$  en fonction de  $k$  pour justifier l'utilisation de la fonction « LOGIT » pour nos données (Fig. 24).

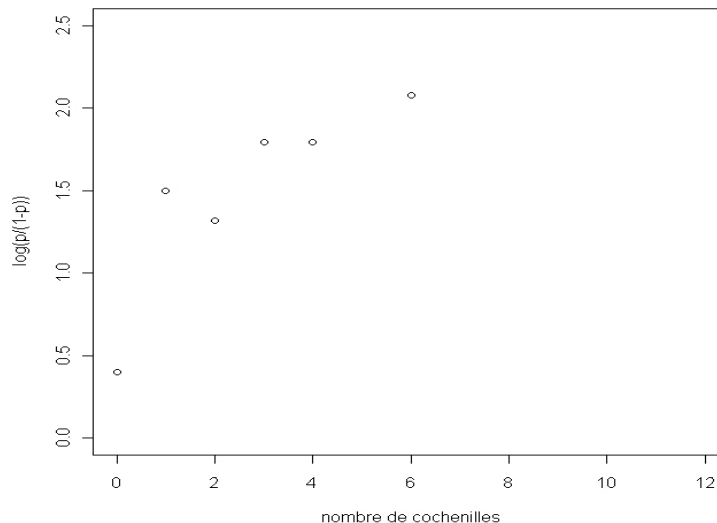


Fig. 24 : graphique de la fonction Logit sur nos données

A la vue du graphique ci-dessus, on considère que l'on peut ajuster un modèle logistique avec la fonction de transformation « LOGIT » à nos données. On utilise le logiciel R pour effectuer la régression logistique (Fig. 25).

```
> reglog<-glm(symptome09~cochenilles08,data=tab,family=binomial())
> summary(reglog)
Call:
glm(formula = symptome09 ~ cochenilles08, family = binomial(),
     data = tab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0093  -1.3847   0.7421   0.9834   0.9834

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.47516    0.04820   9.859  <2e-16 ***
cochenilles08 0.67367    0.06747   9.985  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3488.8 on 2807 degrees of freedom
Residual deviance: 3345.9 on 2806 degrees of freedom
AIC: 3349.9

Number of Fisher Scoring iterations: 5
```

Fig. 25 : sortie R de la régression logistique

La p-valeur associée à la variable « cochenille » vaut approximativement 0 (Fig. 25). On en déduit qu'il y a un effet significatif (au seuil de risque 5%) du facteur « cochenilles08 ».

#### IV-2-3-2- Test d'indépendance spatiale des symptômes conditionnellement au nombre de cochenilles (*P. aceris*) observées

On s'intéresse maintenant au test de l'indépendance de la présence de symptômes conditionnellement au nombre de cochenilles. Ce qui signifie que **sous l'hypothèse d'indépendance, seules les cochenilles présentes sur le cep expliquent la présence de symptômes.**

Pour cela on va effectuer des simulations à l'aide des probabilités  $p$  (Fig. 23).

#### Principe d'une simulation :

On connaît pour chaque cep de la parcelle le nombre  $k$  de cochenilles et la probabilité  $p$  d'avoir des symptômes sachant  $k$ . A partir de ces deux données, on va simuler pour chaque cep la valeur symptôme et donc obtenir une carte simulée des données symptômes.

Pour tester l'indépendance, on va donc effectuer 200 simulations pour calculer l'enveloppe d'indépendance (au seuil de risque 5%) de la même manière que précédemment avec un variogramme le long du rang. On comparera donc le graphique du variogramme des données symptômes observées à l'enveloppe de confiance calculée à partir des simulations.

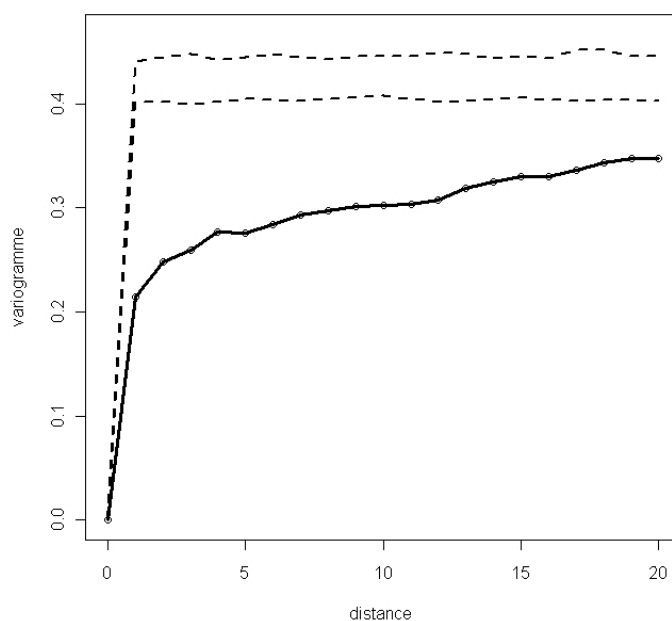


Fig. 26 : test d'indépendance spatiale de la présence de symptômes conditionnellement au nombre de cochenilles par cep

On remarque que pour  $d$  allant de 1 à 20 on a un net écart à l'indépendance (Fig. 26). Cela signifie qu'il y a une dépendance autre que les cochenilles présentes sur le cep qui explique la variable symptôme. **Malgré l'effet significatif des cochenilles sur la présence de symptômes, le test précédent nous informe qu'il reste une dépendance.** Plusieurs raisons peuvent expliquer la persistance d'une dépendance.

- si l'on regarde de plus près les données brutes, on trouve quelques ceps présentant un nombre important de cochenilles (*P. aceris*) mais non porteurs de symptômes. Il est ainsi possible que les premiers symptômes n'apparaissent que plus tard, c'est-à-dire deux ou trois ans après l'inoculation par l'insecte. On ne dispose en effet que peu de données chiffrées concernant la latence de la maladie dans un cep. Cette latence est en effet liée à de nombreux facteurs pédoclimatiques qui compliquent la détection de la maladie et les observations sur le terrain.

- selon le tableau de probabilité (Fig. 23), on remarque que la probabilité associée au nombre 0 de cochenilles vaut 0.599. Ce qui signifie que près de 60% des ceps non infestés par les cochenilles sont malades. Il n'est donc pas surprenant que le test d'indépendance spatiale des symptômes conditionnellement au nombre de cochenilles rejette l'hypothèse d'indépendance. Il se peut donc que des cochenilles aient transmis le virus à la plante une ou plusieurs années auparavant et qu'elles n'apparaissent plus dans la notation pour l'année 2008. Les populations de cochenilles sont en effet régulées par un cortège d'insectes prédateurs et parasitoïdes qui peut venir à bout d'une infestation de cochenilles sur un cep.

- les cochenilles des ceps voisins le long du rang ont pu transmettre le virus (les rameaux entre deux ceps peuvent se chevaucher suivant leurs longueurs). Les notations de cochenilles ayant lieu au printemps, le stade le plus mobile de l'insecte (larve de stade 1) peut ainsi facilement se disperser à l'été le long des rameaux de vigne sans être prise en compte par les notations.

- de nombreux facteurs environnementaux peuvent expliquer également la présence d'une autre dépendance que celle des cochenilles sur le cep. Les cochenilles sont en effet souvent en symbiose avec des fourmis, lesquelles se nourrissent du miellat sucré excrété par les cochenilles, tout en leur fournissant des abris ainsi qu'une protection contre les ennemis naturels. Aussi, la présence de symptômes peut être également conditionnée par la répartition des différentes espèces de fourmis sur la parcelle modèle.

- des erreurs de notations de symptômes ont également pu être commises. Cependant, la justesse des notations sur le terrain est régulièrement vérifiée par des tests fiables de détection de virus dans les ceps.

### IV-3- L'enroulement viral provient-il des parcelles contiguës ?

Le lien entre la présence de cochenilles et l'enroulement viral ayant été démontré, on peut à présent rechercher l'origine de cette maladie. La parcelle modèle de Bonzon a été plantée en 2001. Celle-ci est entourée de deux parcelles, que l'on appellera BzG (parcelle de Bonzon gauche) et BzD (parcelle de Bonzon droite). Ces deux parcelles étaient porteuses de symptômes d'enroulement viral avant même que la parcelle modèle ne soit plantée. Naturellement, on peut se demander si la maladie provient donc des parcelles voisines.

Cette hypothèse sous-entend que les cochenilles ont migré des parcelles contiguës vers la parcelle d'étude en dispersant ainsi la maladie des anciennes parcelles sur notre vignoble.

Deux approches pour répondre à cette hypothèse :

- **approche cochenille** : difficile à mettre en œuvre car il est impossible d'avoir une idée exacte de l'infestation des parcelles environnantes
- **approche virus** :
  - relevé symptomatologique sur les parcelles adjacentes
  - étude de la population virale du virus majoritaire (GLRaV-1) sur la parcelle modèle et les parcelles BzG et BzD

#### IV-3-1- Test d'indépendance de la structure des foyers de symptômes

En supposant que l'épidémie provient des parcelles voisines, on peut penser que les ceps les plus malades (présentant le plus de symptômes) se trouveront d'abord sur les bords de la parcelle, étant accolés aux parcelles BzG et BzD. Étant donné que nous disposons des notations de la sévérité des symptômes, nous décidons d'effectuer un test d'indépendance spatiale de la structure des foyers de symptômes. Les notes s'échelonnent de 0 à 3 en fonction de leur sévérité, la note 0 correspondant à un cep sain.



### Principe des permutations :

On va permuer uniquement les ceps malades, c'est-à-dire ayant une note de sévérité de 1, 2 ou 3, de façon à conserver la position et la taille des groupes de ceps possédant une note différente de 0. Autrement dit on va procéder à un test d'indépendance totale par permutation mais avec des permutations restreintes aux ceps infectés.

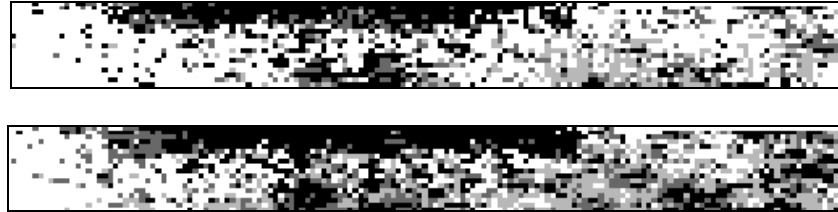


Fig. 27 : représentation graphique sur R de la répartition des ceps infectés par l'enroulement viral selon la sévérité des symptômes pour les années 2008 (haut) et 2009 (bas). Noir : note 3, gris foncé : note 2, gris clair : note 1, blanc : note 0 (sain)

### Commentaire :

On remarque à l'œil nu que les ceps les plus malades sont généralement regroupés ensemble. On va donc justifier cette impression avec un test de permutation.

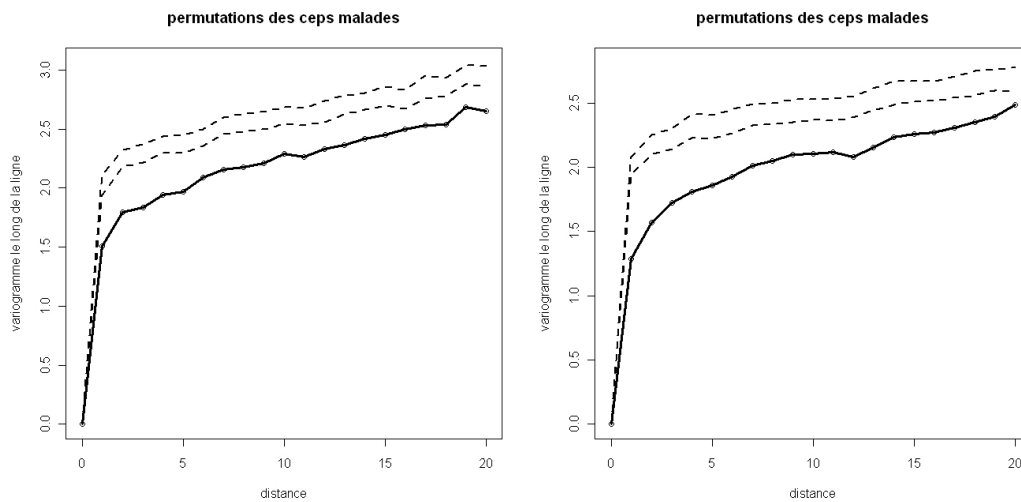


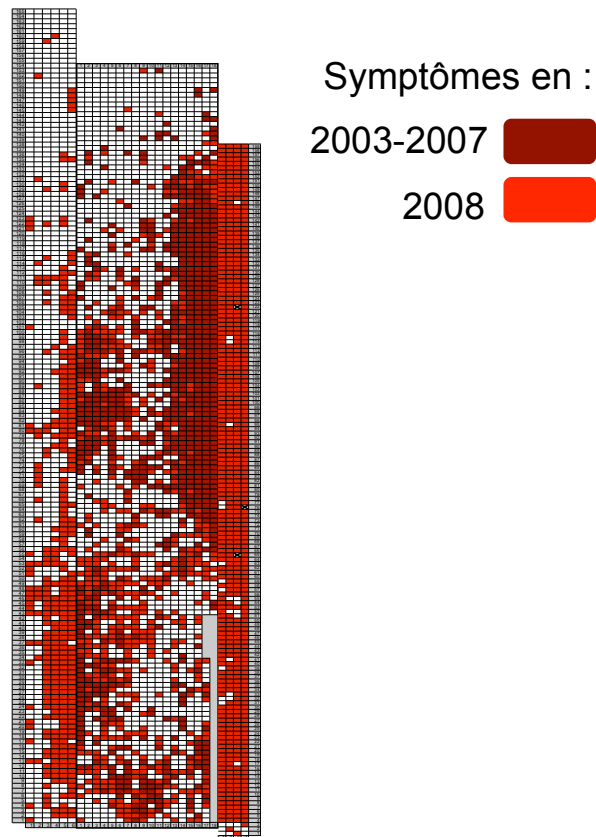
Fig. 28 : Test d'indépendance de la structure spatiale des ceps symptomatiques entre eux en fonction des notes de sévérité pour les années 2008 (gauche) et 2009 (droite)

### Commentaire :

La courbe du variogramme observées est inférieure aux enveloppes pour les années 2008 et 2009. Les ceps malades ne se répartissent pas indépendamment entre eux, des ceps proches ont globalement des notes plus semblables que sous le hasard car on a un fort écart à l'indépendance pour les distances  $d=1, 2, 3, 4, 5$ .

#### IV-3-2- Test d'indépendance spatiale de la structure des symptômes avec les données des parcelles voisines

On dispose aussi d'une notation de la présence des symptômes dans les parcelles voisines pour l'année 2008. C'est pourquoi on va prendre en compte les données des trois parcelles que l'on va regrouper de manière à ce qu'elles ne forment qu'une seule grille. On voudrait effectuer un test d'indépendance par permutation totale (voir IV-1-2-1-) pour voir si on détecte toujours une structure agrégée, après avoir rajouté les données de BzD et BzG. Cependant, la taille des parcelles voisines diffère de la parcelle modèle (Fig. 29).

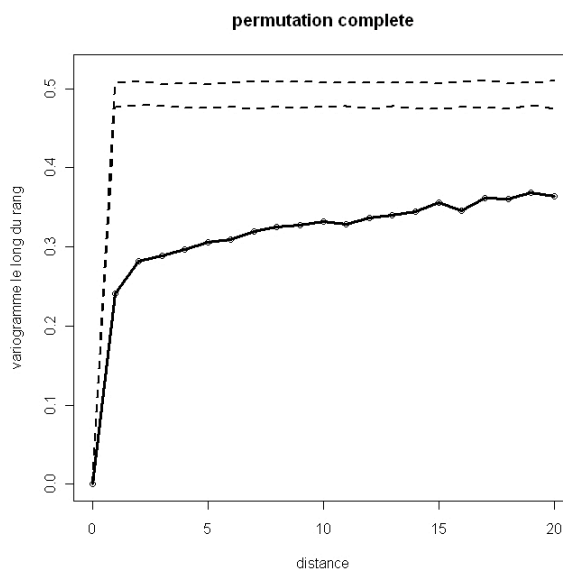


*Fig. 29 : carte de la parcelle modèle et des parcelles voisines avec une notation de la présence des symptômes*

On va donc tronquer les données de manière à obtenir un tableau de données sous forme de grille rectangulaire. On va retirer une partie des données de BzG et la parcelle (en haut), et une partie de BzD (en bas). On obtient alors une carte de symptômes rectangulaire (Fig. 30).



*Fig. 30 : représentation graphique sur R de la carte de la parcelle d'étude et des parcelles voisines modifiées pour le calcul du test d'indépendance*



*Fig. 31 : test d'indépendance totale sur les données symptômes de la parcelle modèle et des parcelles voisines pour l'année 2008*

On constate qu'il y a un net écart à l'indépendance pour ces données (Fig. 31). On en déduit que même en ajoutant les données des parcelles voisines, nos données sont agrégées. Visuellement, (Fig. 29) on remarque assez clairement qu'il y a **une concordance entre les foyers de la parcelle modèle et ceux de la parcelle BzG** qui forment deux groupes de ceps symptomatiques.

#### IV-3-3- Étude de la population virale : piste moléculaire

Le principe de cette étude consiste à analyser la variabilité de la population du virus majoritaire détecté sur la parcelle modèle de Bonzon et les deux parcelles adjacentes. Il existe en effet des différences naturelles au sein même d'une espèce d'enroulement viral. On parle ainsi d'isolats viraux.

Une comparaison des isolats du GLRaV-1 présents sur la parcelle modèle et sur les parcelles adjacentes a été effectuée afin de confirmer que les cochenilles ont acquis, puis inoculé ces mêmes isolats d'une parcelle à l'autre. On s'attend donc à trouver les mêmes populations sur la parcelle modèle, receveuse, que sur les parcelles adjacentes, donneuses.

Les premiers résultats obtenus indiquent que la plupart des isolats du GLRaV-1 présents sur les parcelles anciennes adjacentes sont également détectés dans la parcelle modèle. Cependant, d'autres isolats ont également été retrouvés sur la parcelle modèle, sans que l'on trouve de ceps infectés par ces mêmes isolats sur les parcelles donneuses. L'étude moléculaire montre donc ici ses limites.

Un nouvel échantillonnage est en cours afin de couvrir un plus large nombre de ceps afin d'assurer une représentativité plus juste de l'analyse.

## CONCLUSION DE L'ETUDE

L'analyse spatio-temporelle des données épidémiologiques de l'enroulement de la vigne nous a permis de montrer que l'espèce de cochenille *P. aceris* est une des causes de la dispersion de l'enroulement dans la parcelle d'étude de Bonzon. Nous avons montré dans un premier temps que la répartition des ceps symptomatiques et des ceps infestés par les cochenilles est structurée, avec en plus une concordance entre ces deux structures. Cette corrélation a été confirmée à l'aide d'un modèle qui montre très clairement que les cochenilles *P. aceris* sont une des causes de la présence de l'enroulement viral. Cependant nous avons vu aussi que d'autres facteurs expliquent la présence de la maladie. Une des perspectives de l'étude serait d'essayer de montrer quelles sont ces dépendances.

La maladie de l'enroulement provient des parcelles contiguës infectées. Quelques tests statistiques nous ont permis de justifier en partie ce constat. Mais la démonstration d'un point de vue uniquement statistique n'est pas possible, c'est pour cela qu'une étude d'épidémiologie moléculaire a été initiée.

Les outils et méthodes statistiques utilisés pour ce projet ont ainsi permis d'analyser en profondeur le jeu de données épidémiologique disponible pour une parcelle modèle. Ils ont justifié statistiquement les diverses observations faites à partir des cartographies tout en apportant de nouvelles perspectives quant à l'analyse de la responsabilité de *P. aceris* dans la dispersion de la maladie.

La parcelle de Bonzon ne constitue qu'un exemple de situation épidémiologique parmi beaucoup d'autres. Suivant l'espèce végétale cultivée, la maladie considérée et son mode de transmission (par un organisme vecteur ou non), ainsi que l'ensemble des facteurs agro-environnementaux, la dynamique de la dispersion de la maladie sera différente d'un cas à l'autre.

Il est donc nécessaire de lier en permanence les aspects biologiques et culturels à l'analyse de données épidémiologiques.

## BIBLIOGRAPHIE

- Boudon-Padieu, E., Ridé, M., and Walter, B. (2000): L'enroulement viral, pp. 47-55. In Féret (Ed.): *Maladies à virus, bactéries et phytoplasmes de la Vigne*.
- Confais, J., Grelet, Y., and M., L. G. (2005): La procédure FREQ de SAS : Tests d'indépendance et mesures d'association dans un tableau de contingence. *Modulad*, 188-242.
- Cornillon, P.-A., Guyader, A., Husson, F., Jégou, N., Josse, J., Kloareg, M., Matzner-Lober, E., and Rouvière, L. (2010): *Statistiques avec R, 2e édition augmentée*. Presses Universitaires de Rennes.
- Le Maguet, J. (2010): Un point sur l'enroulement viral. *Le vigneron Champenois* **2**, 49-59.
- Le Maguet, J., Herrbach, E., Hommay, G., Beuve, M., Boudon-Padieu, E., and Lemaire, O. (2009): Monitoring of grapevine leafroll-associated virus 1 (GLRaV-1) dispersion by the mealybug *Phenacoccus aceris*. 16th Meeting of ICVG.
- Peyrard, N., Calonnec, A., Bonnot, F., and Chadoeuf, J. (2005): Explorer un jeu de donnée sur grille par tests de permutation. *Revue de statistique appliquée* **53**, 59-78.
- Rakotomalala, R. (2009): *Pratique de la Régression Logistique - Régression Logistique Binaire et Polytomique*, Université Lumière Lyon 2.
- Sforza, R. (2000): *Cochenilles sur la vigne : bio-éthologie, impact agronomique, lutte et prophylaxie*. Stockel, J. Bordeaux.

# ANNEXE 1

Dans cette annexe, on trouvera le code ainsi que les fonctions codées pour R utilisées pour la partie A et C de l'étude pratique. Les fonctions de test de permutation ainsi que la fonction du variogramme ont été gracieusement mis à disposition par J. Chadoeuf, l'un des auteurs de la publication « analyse d'un jeu de données sur grille par tests de permutation » de N. Peyrard *et al.* J'ai ainsi légèrement modifié ces fonctions pour les appliquer à nos propres données.

Les données concernant la parcelle d'étude étaient sous différentes formes dans un fichier Excel. J'ai donc dans un premier temps effectué des transformations sous Excel, afin d'obtenir pour chaque année, chaque type de donnée (symptôme, cochenilles, note de sévérité, nb de cochenilles par cep) des tableaux rectangulaires de 156 lignes et 18 colonnes (les rangs).

C'est ensuite que j'utilisais le logiciel R pour l'analyse des données, en chargeant les fichiers Excel dans le logiciel.

Comme nous disposions d'un nombre important de tableaux de données et que les analyses étaient les mêmes pour les différents tableaux, j'ai écrit le code de manière générale pour qu'il soit applicable aux différents tableaux de données. De cette manière on peut utiliser le code ci-dessous comme une procédure dès lors que l'on veut appliquer l'un des différents tests pour un type de jeu de données.

```
#ANALYSE D'UN TABLEAU DE DONNEE (COCHENILLE OU SYMPTOME)  
#LE LONG DES RANGS  
#Le seuil pour les enveloppes de confiance est de 0.05
```

```
library(xlsReadWrite)  
#On récupère les tableaux excel  
  
(data<-  
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))  
data<- as.matrix(data)  
data<-t(data)  
data<-data[-1,]
```

```
#Affichage de l'image de la parcelle
```

```
image(t(data),col=grey(1:0))
```

```
#Fonction de calcul d'un variogramme le long du rang (attention, on utilise la  
transposée de la matrice, les rangs sont en lignes)
```

```
gam.fct <- fonction(tab,nd)  
{  
  nx <- dim(tab)[1]  
  ny <- dim(tab)[2]  
  du <- rep(0,nd+1)  
  for (i in 1:nd)  
  {  
    val1 <- tab[, (1+i):ny]  
    val2 <- tab[, 1:(ny-i)]  
    du[i+1] <- sum((val1-val2)**2)  
    du[i+1] <- du[i+1]/(nx*(ny-i))  
  }  
  return(du) }  
}
```

```
#permutation complete des points de la grille (les ceps)
```

```

perml.fct <- function(tab,gam.fct,n,nrep)
{
  nx <- dim(tab)[1]
  ny <- dim(tab)[2]
  vobs <- gam.fct(tab,n)
  nz <- length(vobs)
  tabres <- matrix(0,nz,nrep)
  for (i in 1:nrep)
  {
    tabres[,i] <- gam.fct(matrix(sample(tab),nx,ny),n)
  }
  for (i in 1:nz)
  {
    tabres[i,] <-sort(tabres[i,])
  }

  alphasur2 <-0.025
  k1 <- floor(alphasur2*nrep)
  k2 <- ceiling((1-alphasur2)*nrep)
  zv <- range(c(vobs,tabres))
  plot(0:n,vobs,ylim=zv,xlab="distance",ylab="variogramme le long du
rang",main="permutation complete")
  lines(0:n,vobs,lwd=3)
  lines(0:n,tabres[,k1],lwd=2,lty=2)
  lines(0:n,tabres[,k2],lwd=2,lty=2)
  return()
}

```

**#on affiche le variogramme le long des rangs et les enveloppes pour 200 permutations de tous les points**

```

x11()
perml.fct(data,gam.fct,20,200)

```

**#On permute les rangs entre eux**

```

perm2.fct <- function(tab,gam.fct,n,nrep)
{
  nx <- dim(tab)[1]
  ny <- dim(tab)[2]
  vobs <- gam.fct(t(tab),n)
  nz <- length(vobs)
  tabres <- matrix(0,nz,nrep)
  for (i in 1:nrep)
  {
    tabres[,i] <- gam.fct(matrix(t(tab[sample(nx),]),nx,ny),n)
  }
  for (i in 1:nz)
  {
    tabres[i,] <-sort(tabres[i,])
  }

  alphasur2 <-0.025
  k1 <- floor(alphasur2*nrep)
  k2 <- ceiling((1-alphasur2)*nrep)
  zv <- range(c(vobs,tabres))
  plot(0:n,vobs,ylim=zv,xlab="distance",ylab="variogramme entre
rangs",main="permutations des rangs")
  lines(0:n,vobs,lwd=3)
  lines(0:n,tabres[,k1],lwd=2,lty=2)
  lines(0:n,tabres[,k2],lwd=2,lty=2)
  return()
}

```

**#On affiche le variogramme et les enveloppes pour 200 permutations des rangs**

```

x11()

```

```
perm2.fct(data,gam.fct,10,200)
```

### **#On fait une rotation des rangs**

```
perm2b.fct <- fonction(tab,gam.fct,n,nrep)
{
  nx <- dim(tab)[1]
  ny <- dim(tab)[2]
  vobs <- gam.fct(t(tab),n)
  nz <- length(vobs)
  tabres <- matrix(0,nz,nrep)
  for (i in 1:nrep)
  {
    t1 <- tab
    for (j in 1:nx)
    { id <- sample(ny,1)
      jd <- (1:ny)+id
      jd[jd>ny] <- jd[jd>ny]-ny
      t1[j,] <- t1[j,jd]
    }
    tabres[,i] <- gam.fct(t(t1),n)
  }
  for (i in 1:nz)
  {
    tabres[i,] <-sort(tabres[i,])
  }

  alphasur2 <-0.025
  k1 <- floor(alphasur2*nrep)
  k2 <- ceiling((1-alphasur2)*nrep)
  zv <- range(c(vobs,tabres))
  plot(0:n,vobs,ylim=zv,xlab="distance",ylab="variogramme entre
rangs",main="rotations des rangs")
  lines(0:n,vobs,lwd=3)
  lines(0:n,tabres[,k1],lwd=2,lty=2)
  lines(0:n,tabres[,k2],lwd=2,lty=2)
  return()
}
```

### **#On affiche le variogramme et les enveloppes pour 200 permutations (par #rotation des rangs**

```
X11()
perm2b.fct(data,gam.fct,10,200)
```

### **#Affichage des trois graphiques de variogramme**

```
par(mfrow=c(2,2))
perm1.fct(data,gam.fct,20,200)
perm2.fct(data,gam.fct,10,200)
perm2b.fct(data,gam.fct,10,200)
```

### **#raccourci, importation, affichage**

```
(data<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
data<- as.matrix(data)
data<-t(data)
data<-data[-1,]

x11()
image(t(data),col=grey(1:0))

x11()
perm1.fct(data,gam.fct,20,200)
x11()
```



```

perm2.fct(data,gam.fct,10,200)
x11()
perm2b.fct(data,gam.fct,10,200)

#affichage noir et gris
plot(c(1,156),c(1,156),type="n",xlab="",ylab="",xaxt="n",yaxt="n",bty="n")
image(1:156,1:18,t(data),add=T,col="grey")
image(1:156,1:18,t(data),add=T,col=0:1)

#ANALYSE DE LA STRUCTURE AVEC LES NOTES DE SEVERITE

#On récupère les tableaux excel

(data<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
data<- as.matrix(data)
data<-t(data)
data<-data[-1,]

#Affichage de l'image de la parcelle
x11()
plot(c(1,156),c(1,156),type="n",xlab="",ylab="",xaxt="n",yaxt="n",bty="n")
image(1:156,1:18,t(data),add=T,col=grey((3:0)/3))

#Fonction de calcul d'un variogramme le long du rang (attention, on utilise la
transposée de l'image, les rangs sont en lignes)

gam.fct <- fonction(tab,nd)
{
  nx <- dim(tab)[1]
  ny <- dim(tab)[2]
  du <- rep(0,nd+1)
for (i in 1:nd)
  {
val1 <- tab[, (1+i):ny]
val2 <- tab[, 1:(ny-i)]
du[i+1] <- sum((val1-val2)**2)
du[i+1] <- du[i+1]/(nx*(ny-i))
  }
return(du) }

#Permutation des seuls strictement positifs (=attaques)

perm3.fct <- fonction(tab,gam.fct,n,nrep)
{
  nx <- dim(tab)[1]
  ny <- dim(tab)[2]
vobs <- gam.fct(tab,n)
nz <- length(vobs)
tabres <- matrix(0,nz,nrep)
for (i in 1:nrep)
  {
t1 <- tab
t1[tab>0] <- sample(tab[tab>0])
tabres[,i] <- gam.fct(t1,n)
  }
for (i in 1:nz)
  {
tabres[i,] <-sort(tabres[i,])
  }

alphasur2 <-0.025
k1 <- floor(alphasur2*nrep)
k2 <- ceiling((1-alphasur2)*nrep)

```

```

zv <- range(c(vobs,tabres))
plot(0 :n,vobs,ylim=zv,xlab="distance",ylab="variogramme le long de la
ligne",main="permutations des ceps malades")
lines(0:n,vobs,lwd=3)
lines(0:n,tabres[,k1],lwd=2,lty=2)
lines(0:n,tabres[,k2],lwd=2,lty=2)
return()
}
X11()
perm3.fct(data,gam.fct,20,200)

```

**#procédure pour faire les tableaux des données cumulées à l'aide de R**  
**#Exemple pour les données symptômes 2004 et symptômes 2005**

**#On récupère les tableaux excel**

```

(symptome04<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
data<- as.matrix(data)
data<-data[-1,]
(symptome05<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
data<- as.matrix(data)
data<-data[-1,]

```

**#cumul de l'année 2004 et 2005**

```
s45<-symptome04+symptome05
```

**#on convertit la matrice en tableau excel**

```
write.xls(s45,"cumul _symp45.xls",colNames=TRUE,sheet=1,from=1,rowNames=NA)
```

Voici pour exemple le type de tableau utilisé pour l'analyse de données. Ce tableau correspond à la présence de symptômes pour l'année 2009.

| rang1 | rang2 | rang3 | rang4 | rang5 | rang6 | rang7 | rang8 | rang9 | rang10 | rang11 | rang12 | rang13 | rang14 | rang15 | rang16 | rang17 | rang18 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 1      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 1      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      |
| 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0      | 1      | 0      | 0      | 0      | 0      | 0      | 1      | 1      |
| 1     | 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 1      | 1      | 1      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      | 1      |
| 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 1      | 0      | 0      | 1      | 0      | 0      | 0      | 1      | 1      |
| 0     | 0     | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 0      | 0      | 1      | 0      | 1      | 0      | 0      | 1      | 1      |
| 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 1     | 0      | 1      | 1      | 1      | 0      | 0      | 1      | 0      | 1      |
| 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      |
| 0     | 0     | 1     | 0     | 0     | 0     | 1     | 1     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      | 0      |
| 0     | 0     | 0     | 0     | 0     | 1     | 1     | 1     | 1     | 0      | 0      | 0      | 0      | 1      | 0      | 1      | 1      | 1      |
| 0     | 1     | 0     | 0     | 0     | 1     | 1     | 1     | 1     | 0      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 0      |
| 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 1      | 0      | 1      | 1      | 0      | 1      | 1      | 1      | 0      |
| 0     | 1     | 0     | 1     | 0     | 0     | 1     | 0     | 0     | 0      | 1      | 1      | 0      | 1      | 1      | 1      | 1      | 0      |
| 0     | 1     | 1     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |
| 0     | 0     | 1     | 1     | 0     | 0     | 0     | 1     | 0     | 0      | 0      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |
| 0     | 1     | 1     | 0     | 1     | 0     | 0     | 1     | 0     | 0      | 1      | 1      | 1      | 1      | 0      | 1      | 1      | 1      |
| 1     | 1     | 1     | 1     | 1     | 0     | 0     | 0     | 0     | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |
| 1     | 1     | 1     | 1     | 0     | 0     | 1     | 0     | 0     | 0      | 0      | 1      | 1      | 1      | 1      | 1      | 1      | 1      |





## ANNEXE 2

Dans cette annexe on trouvera le code correspondant aux sections 2.a. et 2.b. de l'étude pratique, en particulier le test du  $\chi^2$  ainsi que l'étude de deux jeux de données avec le critère du covariogramme. Les fonctions de covariogramme et de test de permutations associés ont été récupérées auprès d'un des auteurs de la publication « analyse d'un jeu de données sur grille par tests de permutation » de N. Peyrard *et al.* Celles-ci ont été légèrement modifiées par moi-même pour les appliquer à nos propres données.

### **#test du Khi 2**

```
m1<-matrix(c(214,150,744,1649),nrow=2)
m2<-matrix(c(484,150,726,1273),nrow=2)
m3<-matrix(c(717,629,393,1018),nrow=2)
m4<-matrix(c(921,1009,212,615),nrow=2)
chisq.test(m1)
chisq.test(m2)
chisq.test(m3)
chisq.test(m4)
```

### **#Analyse de deux grilles par tests de permutation à l'aide du critère du covariogramme**

#### **#on récupère les fichiers**

```
(tab1<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
```

```
(tab2<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
```

#### **#On convertit les tableaux en matrice, et on affiche un aperçu des parcelles**

```
m1<-as.matrix(tab1)
m2<-as.matrix(tab2)
m1<-m1[,-1]
m2<-m2[,-1]
```

```
x11()
par(mfrow=c(2,1))
image(m1,col=grey(1:0))
image(m2,col=grey(1:0))
```

#### **#Fonction de translation dans un sens seulement, le long du rang (dans notre cas 156 translations)**

```
translc.fct<- fonction(mat,i1,i2)
{
n1 <- dim(mat)[1]
n2 <- dim(mat)[2]
mu <- mat
j <- c(i1:n1,1:i1-1)
if(i1==1) j <- 1:n1
for (i in 1:n2)
{
```

```

mu[,i] <- (mat[,i])[j]
}

return(mu)
}

#test
t12c.fct <- function(mat1,mat2,nd){
vobs <- covarl.fct(mat1,mat2,nd)
nrep <- dim(mat1)[1]*dim(mat1)[2]
mg <- matrix(0,length(vobs),nrep)
i <-0
for (i1 in 1:dim(mat1)[1]){
  for (i2 in 1:dim(mat1)[2]){
    i <- i+1
mg[,i] <- covarl.fct(translc.fct(mat1,i1,i2),mat2,nd)
  }}
for (i in 1:length(vobs)){
mg[i,] <- sort(mg[i,])
}
ind1 <-floor(2.5*nrep/100)
ind2 <-ceiling(97.5*nrep/100)
gmin <- mg[,ind1]
gmax <- mg[,ind2]
gmoy <- apply(mg,1,mean)
cr <- range(c(gmax,gmin,vobs))
plot(vobs,ylim=cr,xlab= " distance",ylab= "covariogramme",main="les deux variables
sont elles dependantes ?")
lines(vobs,lwd=2)
lines(gmin,lty=2)
lines(gmax,lty=2)
lines(gmoy,lty=3)
pv <- vobs*0
for (i in 1:length(vobs)){
pv[i] <- length(mg[i,][mg[i,]>=vobs[i]])/length(mg[i,])
}
return(pv)
}

X11()
print(t12c.fct(m1,m2,20))

```

## ANNEXE 3

Dans cette annexe on trouvera le code correspondant à la section 2.c. de l'étude pratique, c'est-à-dire le code utilisé pour le modèle. Les fonctions utilisées ci-dessous ont été écrites par moi-même (fonctions de comptages, simulations, tests).

```
#CODE R pour le modèle logistique de la variable « symptomes » (présence/absence)  
#expliquée par la variable « cochenilles » (en effectif)
```

```
#on récupère les données
```

```
(tab1<-  
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
```

```
(tab2<-  
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
```

```
#On convertit les tableaux en matrice
```

```
symp09<-as.matrix(tab1)  
aceris08<-as.matrix(tab2)  
symp09<-symp09[,-1]  
aceris08<-aceris08[,-1]
```

```
#fonction qui renvoie le nombre de ceps ayant k cochenilles, le nombre de ceps avec  
#symptomes ayant k cochenilles, ainsi que la fréquence de ceps symptomatiques ayant  
#k cochenilles
```

```
count.fct<-function(ts,ta,k){  
  nb_symp<-0  
  nb_tt<-0  
  
  n1<-dim(ts)[1]  
  n2<-dim(ta)[2]  
  
  for(i in 1:n1){  
    for(j in 1:n2){  
      if (ts[i,j]==1 && ta[i,j]==k)  
        nb_symp<-nb_symp+1  
      if(ta[i,j]==k)  
        nb_tt<-nb_tt+1  
    }  
  }  
  freq<-nb_symp/nb_tt  
  print(nb_symp)  
  print(nb_tt)  
  print(freq)  
}
```

```
count.fct(symp09,aceris08,0)
```

```
#fonction qui calcul la fréquence de ceps symptomatiques en fonction du nombre de  
#cochenilles, renvoie le vecteur de fréquence pour chaque nombre de cochenilles, de  
#0 à kmax-1
```

```

freq.fct<-function(m1,m2,kmax){
n1<-dim(m1)[1]
n2<-dim(m1)[2]
freq<-c(rep(0,kmax))
tt<-c(rep(0,kmax))
for(k in 1:kmax){
  nb_symp<-0
  nb_tt<-0
for(i in 1:n1){
for(j in 1:n2){
  if (m1[i,j]==1 && m2[i,j]==k-1)
  nb_symp<-nb_symp+1
  if(m2[i,j]==k-1)
  nb_tt<-nb_tt+1
  }
  }
  tt[k]<-nb_tt
  freq[k]<-nb_symp/nb_tt
  }
}
return(freq)
}

```

**#fonction qui compte le nombre de ceps en fonction du nombre de cochenilles**

```

nb_tt.fct<-function(m1,m2,kmax){
n1<-dim(m1)[1]
n2<-dim(m1)[2]
tt<-c(rep(0,kmax))

for(k in 1:kmax){
  nb_tt<-0
for(i in 1:n1){
for(j in 1:n2){
  if(m2[i,j]==k-1)
  nb_tt<-nb_tt+1
  }
}
tt[k]<-nb_tt
}
print(tt)
}

```

**#Graphique des fréquences pour un nombre K de cochenilles (P. acéris)**

**#calcul des probabilités d'avoir des symptômes sur un cep pour chaque nombre k #de  
#cochenilles, nombre total de ceps avec k cochenilles**

```

proba<-freq.fct(symp09,aceris08,13)
nb_tt<-nb_tt.fct(symp09,aceris08,13)
nb_coch<-c(0:12)

```

**#ecart-type de la loi des fréquences**

```

variance<-(proba*(1-proba))/nb_tt

```

**#graphique des probabilités avec une enveloppe de + ou - l'écart-type**

```

plot(proba~nb_coch,xlim=c(0,12),ylim=c(0,1),xlab="nombre de cochenilles",ylab="
fréquence de ceps symptomatiques ")
ecart_t<-sqrt(variance)
lines((proba+ecart_t)~c(0:12),lty=3)
lines((proba-ecart_t)~c(0:12),lty=3)

```

**#graphique de la fonction logistique**

```

plot(log(proba/(1-proba)),xlim=c(1,12),ylim=c(0,2.5))

```

**#Graphique du variogramme et simulation pour tester l'indépendance**

**#fonction de simulation : on simule les valeurs symptômes en sachant le nombre de**



**#cochenilles (observés) pour chaque cep de la parcelle à l'aide de la loi de probabilité calculé sur les données #observées**

```
simulation<-function(t_coch,vec_prob){
n1<-dim(t_coch)[1]
n2<-dim(t_coch)[2]
t<-matrix(0,n1,n2)
for (i in 1:n1){
for (j in 1:n2){
p<-t_coch[i,j]+1
pp<-vec_prob[p]
t[i,j]<-sample(c(1,0),1,p=c(pp,1-pp))
}}
print(t)
}
```

**#fonction du variogramme le long du rang**

```
gam.fct <- function(tab,nd)
{
nx <- dim(tab)[1]
ny <- dim(tab)[2]
du <- rep(0,nd+1)
for (i in 1:nd)
{
val1 <- tab[(1+i):nx,]
val2 <- tab[1:(nx-i),]
du[i+1] <- sum((val1-val2)**2)
du[i+1] <- du[i+1]/((nx-i)*ny)
}
return(du) }
```

**#on teste l'indépendance en effectuant un nombre « nrep » de simulation, on #trace #un variogramme et l'enveloppe de confiance**

```
test.fct<-function(t_symp,t_coch,vec_prob,gam.fct,nrep,nd){
vobs<-gam.fct(t_symp,nd)
mat_v<-matrix(0,length(vobs),nrep)

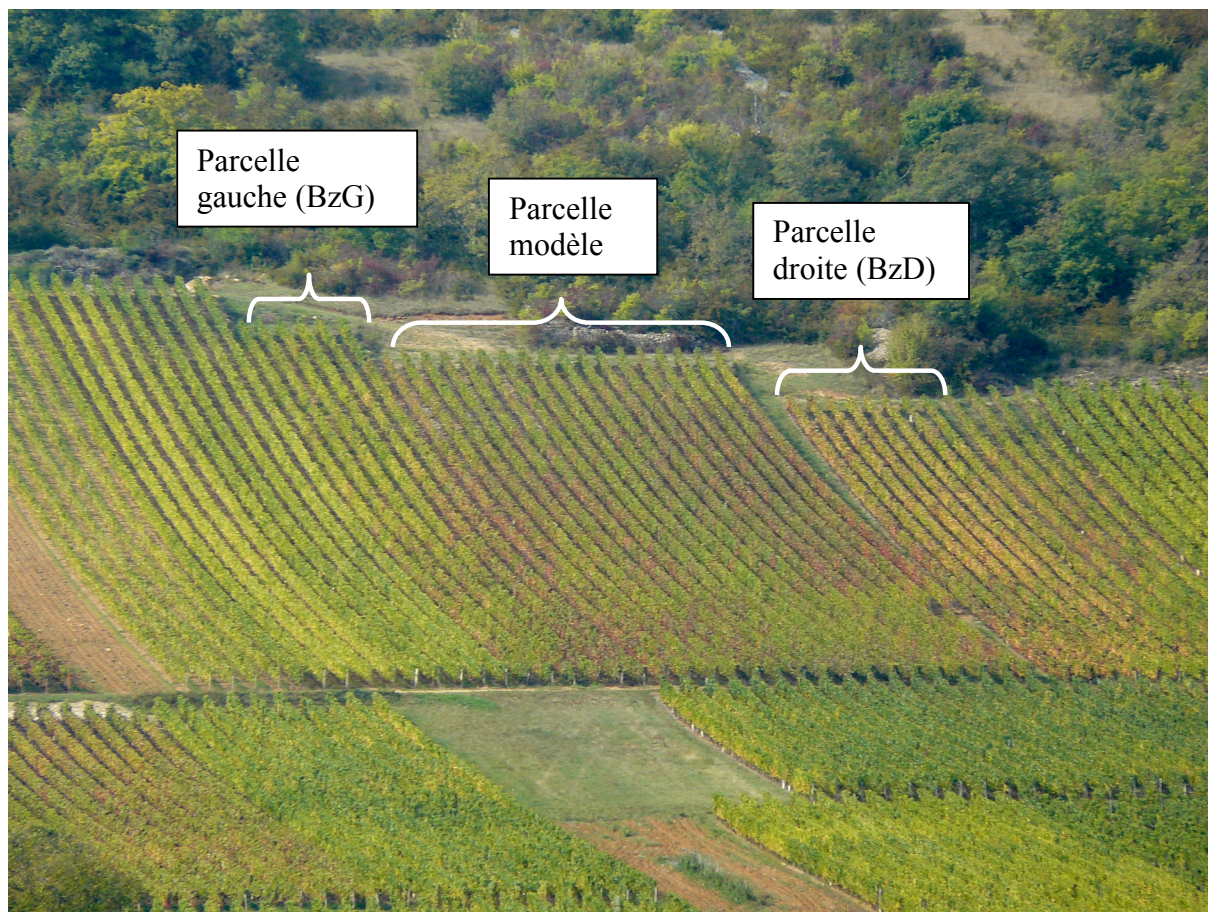
for (i in 1:nrep) {
mat_v[,i]<-gam.fct(simulation(t_coch,vec_prob),nd)
}
for (i in 1:length(vobs)) {
mat_v[i,]<-sort(mat_v[i,])
}

alphasur2<-0.025
k1<-floor(alphasur2*nrep)
k2<-ceiling((1-alphasur2)*nrep)
zv<-range(c(vobs,mat_v))
plot(0:nd,vobs,ylim=zv,xlab="distance",ylab=" variogramme ")
lines(0:nd,vobs,lwd=3)
lines(0:nd,mat_v[,k1],lwd=2,lty=2)
lines(0:nd,mat_v[,k2],lwd=2,lty=2)
return()
}
```

```
test.fct(symp09,aceris08,proba,gam.fct,200,15)
```

**#regression logistique des symptomes09 en fonction de P.aceris08  
#pour cette étape, on a utilisé un tableau de deux colonnes, avec les données pour #chaque cep**

```
(tableau<-
read.xls(file.choose(),colNames=TRUE,sheet=1,type="data.frame",from=1,rowNames=NA))
reglog<-glm(symptome09~aceris08,data=tab,family=binomial())
summary(reglog)
```



*Fig. 4 : photographie de la parcelle modèle de Bonzon et des parcelles de gauche et droite (JLM)*

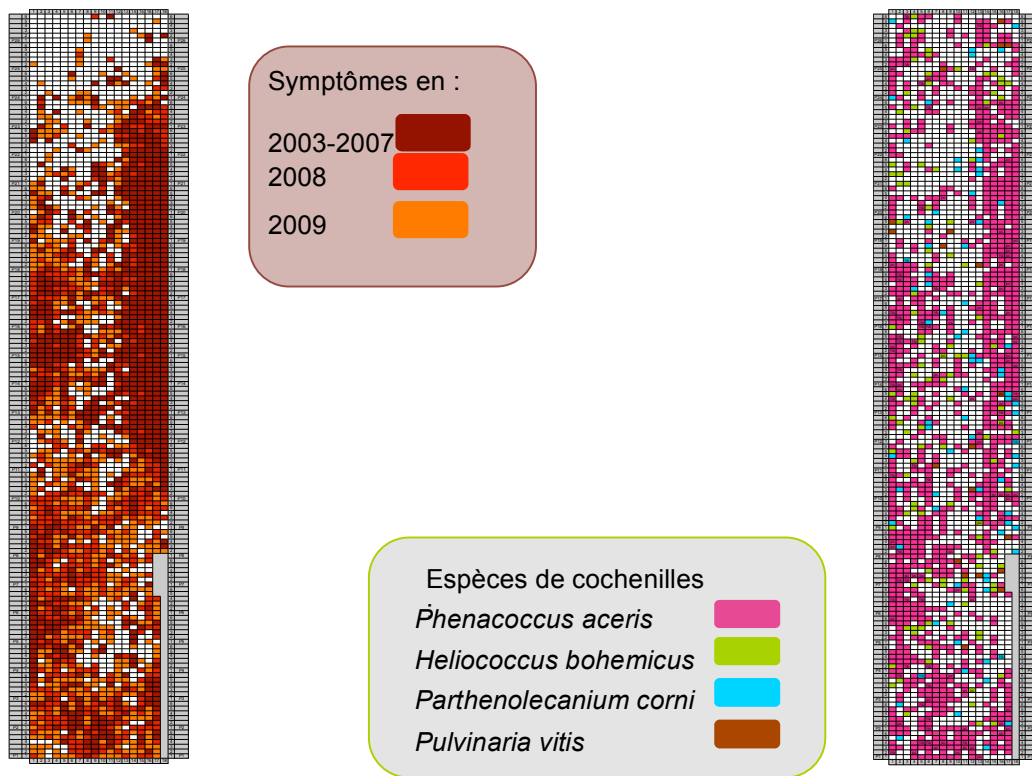
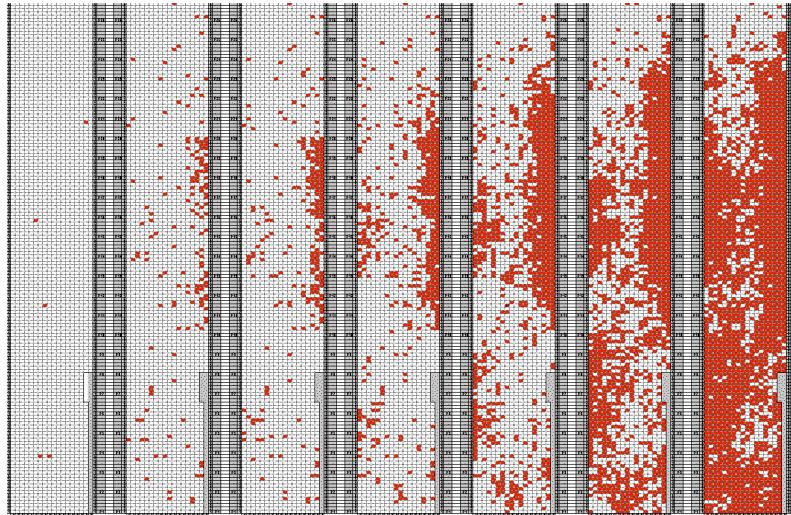


Fig. 6 : cartographie de la présence de symptômes par cep de 2003 à 2009

Fig. 7 : cartographie de la présence des différentes espèces de cochenilles par cep sur la parcelle modèle



*Fig. 11 : cartographies de la parcelle modèle de Bonzon, représentant la répartition des ceps symptomatiques de 2003 à 2009*