



**HAL**  
open science

# Modélisation spatio-temporelle des épidémies : comment diversifier les paysages agricoles ?

Manon Couty

► **To cite this version:**

Manon Couty. Modélisation spatio-temporelle des épidémies : comment diversifier les paysages agricoles ?. Environmental Sciences. 2023. hal-04694185

**HAL Id: hal-04694185**

**<https://hal.inrae.fr/hal-04694185v1>**

Submitted on 11 Sep 2024

**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.

---

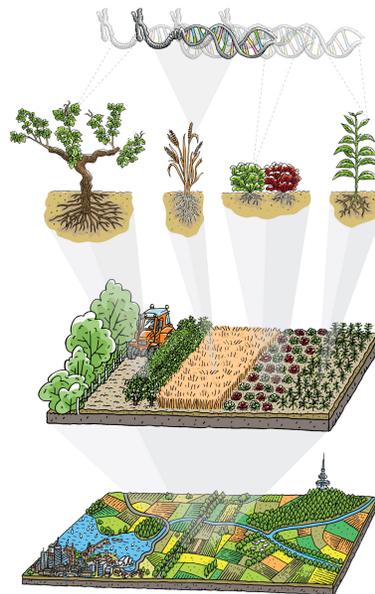
# Rapport de stage

---

## Modélisation spatio-temporelle des épidémies : comment diversifier les paysages agricoles ?

Manon Couty

1er Mars - 25 Août 2023



Rimbaud L, et al. 2021  
Annu. Rev. Phytopathol. 59:125-52

INSA Lyon - 5BIM - Biosciences

---

Unité Biostatistique et Processus Spatiaux (BioSP) - INRAE Avignon

Encadrants : Julien Papaix, Loup Rimbaud, Jean-François Rey

## Remerciements

Je tiens à remercier tout particulièrement Julien Papaïx qui a su m'encadrer avec bienveillance et sollicitude sur toute la durée de mon stage et qui a pu m'apporter de nombreuses connaissances et de précieux conseils !

Je remercie également Loup Rimbaud et Jean-François Rey pour leur disponibilité et leur implication ainsi que Marta Zaffaroni, qui m'ont aidé à découvrir et prendre en main *landsepi*. Merci aussi à Julien et Loup pour leur contribution à ce rapport.

J'adresse mes remerciements à toute l'équipe de BioSP qui a su m'accueillir chaleureusement et aux moyens mis à disposition par l'unité (cluster de calculs).

Dédicace à l'équipe des jeunes (stagiaires, doctorants et ingénieurs) et à ces bons moments partagés sous le soleil de Provence.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Agriculture et épidémies . . . . .	1
1.2	Coexistence entre pathotypes . . . . .	2
1.3	Objectifs du stage . . . . .	3
<b>2</b>	<b>Matériel &amp; Méthodes</b>	<b>4</b>
2.1	Le modèle . . . . .	4
2.2	Déroulé d'une simulation . . . . .	6
2.2.1	Dynamique spatio-temporelle . . . . .	6
2.2.2	Stratégies de résistance . . . . .	6
2.2.3	Paramétrage de l'agent pathogène . . . . .	7
2.2.4	Sorties des simulations . . . . .	8
2.3	Modifications du modèle . . . . .	9
2.4	Plan des simulations . . . . .	10
2.4.1	Cas d'études . . . . .	11
2.5	Analyses statistiques . . . . .	13
2.5.1	Cas d'étude 1 . . . . .	13
2.5.2	Cas d'étude 2 . . . . .	14
<b>3</b>	<b>Résultats</b>	<b>16</b>
3.1	Cas d'étude 1 . . . . .	16
3.1.1	Analyse de sensibilité . . . . .	16
3.1.2	Prédiction de la structure de la population de P3 . . . . .	17
3.1.3	Effet du trait d'histoire de vie . . . . .	20
3.1.4	Effet du taux de mutation . . . . .	20
3.1.5	Agrégation et fréquence locales . . . . .	20
3.2	Cas d'étude 2 . . . . .	21
3.2.1	Sorties du modèle . . . . .	21
3.2.2	Agrégation et fréquence locales . . . . .	26

<b>4</b>	<b>Discussion</b>	<b>27</b>
<b>5</b>	<b>Perspectives</b>	<b>29</b>
<b>6</b>	<b>Conclusion</b>	<b>30</b>
<b>7</b>	<b>Annexes</b>	<b>31</b>
7.1	Cas d'étude 1 : Introduction et établissement du généraliste . . . . .	31
7.1.1	<code>summary()</code> des GLM . . . . .	31
7.1.2	Boxplot de la variance de la fréquence de $P3$ . . . . .	32
7.1.3	Exemples de simulations . . . . .	32
7.1.4	Diagrammes en barres . . . . .	34
7.1.5	Graphiques de contours . . . . .	34
7.1.6	Boxplots comparant les traits d'histoire de vie . . . . .	35
7.1.7	<code>summary()</code> du modèle mixte . . . . .	36
7.2	Cas d'étude 2 : Coexistences entre pathotypes . . . . .	36
7.2.1	<code>summary()</code> du modèle de régression multinomiale . . . . .	36
7.2.2	Coexistences . . . . .	37
7.2.3	Exemples de simulations . . . . .	38
7.3	Section modifiée de la vignette 1 de <code>landsepi</code> . . . . .	39

# Présentation de l'INRAE PACA et l'unité BioSP

L'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement ou INRAE est un établissement public à caractère scientifique et technologique (EPST) qui dépend du ministère de l'Enseignement Supérieur et de la Recherche ainsi que du ministère de l'Agriculture et de la Souveraineté Alimentaire. L'INRAE est créé en 2020, suite à la fusion de deux institutions : l'Institut National de la Recherche Agronomique (INRA), fondé en 1946, et l'Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA), fondé en 2012. 18 centres sont implantés sur la France, constitués de 272 unités de recherche, ce qui représente environ 12 000 personnes. L'INRAE occupe une place de leader au niveau mondial sur les disciplines de sciences végétales, animales ainsi qu'agricoles, alimentaires et environnementales.

Le centre INRAE Provence-Alpes-Côte d'Azur, où j'ai réalisé mon stage, se concentre principalement sur des problématiques liées au territoire méditerranéen telles que l'impact sur l'agriculture et les forêts du changement climatique et l'adaptation à ces changements, la transition agroécologique, la santé des plantes, les risques naturels ou encore la gestion des ressources hydriques.

L'unité de Biostatistique et processus SPatiaux (BioSP), implantée sur le site d'Agroparc à Avignon, est dépendante des départements MathNum (Mathématiques et Numérique) et SPE (Santé des Plantes et Environnement). Ses travaux de recherche principaux sont à l'interface entre statistique, systèmes dynamiques et écologie-épidémiologie. L'unité est composée d'une équipe Recherche et d'une équipe OPE (équipe Opérationnelle à l'INRAE pour la Plateforme d'Epidémiosurveillance en santé végétale).

# 1 Introduction

## 1.1 Agriculture et épidémies

Depuis l'avènement de l'agriculture au néolithique et la domestication des plantes qui en découle, l'Homme a cherché à maximiser la production agricole ou en améliorer la qualité. Cette maximisation a débuté de façon empirique en sélectionnant les graines des individus les plus favorables (caractères d'intérêt comme le goût, la taille, le rendement...) et les plus adaptés pour la prochaine récolte. Cette sélection dite "massale" a peu à peu fait place aux hybridations ou croisements variétaux grâce aux progrès dans le domaine de la génétique au XIX siècle.

En réponse à la mise en place de cultures, l'Homme a également domestiqué de manière non intentionnelle les populations de bio-agresseurs inféodés à ce nouvel environnement. Les bio-agresseurs désignent les virus, bactéries, champignons ou insectes nuisibles aux cultures. En effet, la mise en place de cultures procure aux agents pathogènes un milieu stable et homogène par rapport au compartiment sauvage, puisque des plantes proches d'un point de vue géographique le sont également d'un point de vue génétique (plantes similaires sur un même champ) ce qui est propice au développement d'épidémies (Stukenbrock et al., 2007).

Cette problématique a cependant prit une dimension bien supérieure depuis l'intensification de la production agricole au lendemain des guerres mondiales. Cette intensification s'est accompagnée de la généralisation des monocultures et du développement et l'utilisation massive d'intrants chimiques (fertilisants et produits phytosanitaires comme les insecticides, fongicides...) qui ont permis de sécuriser la production face aux risques épidémiques. Bien que très efficace, la protection chimique se révèle être dévastatrice sur l'environnement (Bonney, 2012) et peu pérenne puisque les agents pathogènes ont la capacité de s'adapter aux substances actives utilisées (Leroux et al., 2011 ; Omrane et al., 2017).

Un autre moyen de lutte face aux épidémies, lui-aussi déployé massivement, se fait via l'immunité végétale, c'est-à-dire en déployant des variétés résistantes aux maladies. Il existe deux types de résistance chez les plantes : les résistances qualitatives (monogéniques)(Flor, 1956) qui confèrent une résistance totale aux agents pathogènes, et les résistances quantitatives (polygéniques) (Lannou, 2012) qui offrent une protection partielle. Ces dernières

peuvent agir de différentes manières sur l'agent pathogène et ont comme conséquence de réduire la sévérité des symptômes en limitant le développement de l'agent pathogène. Malheureusement, là aussi les agents pathogènes peuvent s'adapter et contourner les résistances déployées (Ehrlich et al., 1964). Ainsi, les gènes de résistance étant une ressource précieuse et limitée, il est nécessaire de les déployer de façon raisonnée à différentes échelles spatio-temporelles. Une des étapes préliminaires est alors de comprendre comment se structurent les dynamiques des épidémies selon le paysage, plus ou moins hétérogène. En effet, de cette structuration va dépendre le potentiel adaptatif des agents pathogènes.

## 1.2 Coexistence entre pathotypes

L'un des enjeux majeurs pour comprendre la structuration de communautés des agents pathogènes réside dans la diversité des paysages agricoles, qui se composent de mosaïques de champs occupés par de multiples variétés. L'habitat d'un agent pathogène est donc constitué de sous-habitats formés par la diversité des plantes hôtes. Cette hétérogénéité du paysage implique des sous-populations d'agents pathogènes, comme différents génotypes d'un même agent pathogène (pathotypes), formant une communauté hétérogène. Une question fondamentale est alors de savoir quel génotype d'un agent pathogène va se développer sur quelle variété, étant donné la structuration du paysage. Il existe différents profils de pathotypes : les pathotypes spécialistes qui ont contourné un type de résistance et vont infecter en priorité une seule variété, et les pathotypes généralistes qui ont contourné toutes les résistances et considèrent donc le paysage de façon homogène. L'introduction d'un nouveau pathotype comme un généraliste dans un paysage peut se faire soit via une migration extérieure (champ voisin ou compartiment sauvage), soit par mutation des pathotypes déjà présents.

L'étape suivant l'introduction d'un pathotype est son établissement (éventuel) dans la population d'agents pathogènes. Si le paysage est homogène, il est possible de prédire si le pathotype pourra perdurer ou non dans la population à partir du taux de reproduction de base de l'agent pathogène :  $R_0$  (nombre de nouveaux cas d'infections attendu causées par un individu infecté dans une population d'hôtes sensibles de taille infinie) (With, 2002). Un  $R_0 > 1$  est synonyme d'un établissement tandis qu'un  $R_0 < 1$  mène à l'extinction du pathotype. Si plusieurs pathotypes sont en compétition pour le génotype d'un hôte, le pathotype ayant

le  $R_0$  le plus élevé sera celui ayant le plus de succès (Yorke et al., 1978). Cependant, dans un paysage hétérogène (ou fragmenté), composés de génotypes d'hôtes différents (différentes espèces ou variétés), la distance moyenne entre les hôtes de même génotype augmente. Ainsi, par des effets de dilution des propagules produites et de barrières à leur dispersion (Keesing et al., 2006), la propagation d'un pathotype donné sera limitée. Il existe deux scénarios possibles à long terme : la coexistence entre plusieurs pathotypes ou la persistance d'un seul pathotype dans le paysage. Une notion maintes fois démontrée est que  $n$  génotypes peuvent cohabiter sur  $n$  habitats (Karlin et al., 1972 ; Strobeck, 1979).

Il existe différents mécanismes de coexistence stable. Un des plus classiques, le "niche partitioning" ou partitionnement de niche (Chesson, 2000) consiste à avoir les pathotypes qui occupent des niches écologiques différentes (habitats et ressources différentes). Cette coexistence se traduit par la persistance des génotypes spécialistes (Melbourne et al., 2007) sur leur variété propre. Un autre mécanisme, nommé "habitat boundary polymorphism" (Débarre et al., 2011) est favorisé par une hétérogénéité spatiale et une dispersion limitée et conduit à la coexistence des différents pathotypes (persistance des pathotypes généralistes dans la population au côté des spécialistes).

### 1.3 Objectifs du stage

La problématique principale étudiée au cours de ce stage est l'étude des mécanismes impliqués dans la coexistence entre différents pathotypes dans des paysages plus ou moins hétérogènes. Le premier objectif est d'apporter des modifications à un modèle existant, implémenté dans le package R *landsepi* afin de pouvoir mieux définir l'inoculum de départ dans les simulations. L'étape suivante est d'utiliser le modèle modifié pour lancer des simulations et étudier l'établissement du pathotype généraliste ainsi que la coexistence entre pathotypes sur le long terme. Enfin, l'objectif final est d'effectuer des analyses statistiques sur les sorties des simulations pour mieux comprendre les mécanismes impliqués dans la coexistence entre pathotypes généraliste et spécialistes.

## 2 Matériel & Méthodes

### 2.1 Le modèle

Le package R *landsepi* (<https://CRAN.R-project.org/package=landsepi>) repose sur un modèle mathématique qui permet de simuler la propagation et l'évolution d'un agent pathogène dans un paysage agricole hétérogène. C'est un modèle spatio-temporel, stochastique, et démo-génétique (Rimbaud et al., 2018; Rimbaud et al., 2021).

*Landsepi* est un modèle à compartiments de type HLIR (Healthy-Latent-Infectious-Removed, ou SEIR (Susceptible-Exposed-Infectious-Recovered) (Figure 1). Au début d'une épidémie, les hôtes présents sont tous dans le compartiment  $H$  (tous sains), excepté les hôtes où l'agent pathogène a été inoculé, qui sont déjà infectieux (compartiment I). Les hôtes sains vont peu à peu se déplacer dans les compartiments suivants à mesure des infections et de la progression de l'épidémie. Ce modèle est régi par un système d'équations stochastiques, à pas de temps discret. Le modèle est démo-génétique dans le sens où il tient compte de l'évolution de l'agent pathogène et permet l'apparition de nouveaux pathotypes par mutations qui contournent les résistances présentes chez les plantes. Enfin, le modèle a une dimension à la fois spatiale et temporelle puisque la modélisation permet d'une part une étude au niveau du paysage et d'autre part un suivi sur plusieurs pas de temps et saisons de récoltes.

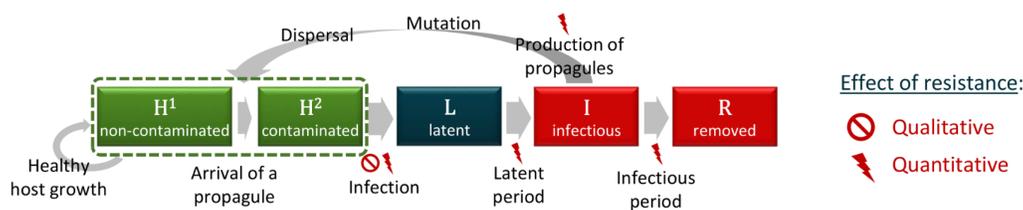


FIGURE 1 – Les hôtes sains  $H$  peuvent être séparés en  $H^1$  (non contaminés par les propagules) et  $H^2$  (contaminés). Les hôtes deviennent infectés (compartiment  $L$ ), puis infectieux (compartiment  $I$ ) après une période de latence. Les hôtes  $I$  produisent des propagules qui vont infecter d'autres hôtes. Après la période infectieuse, les hôtes deviennent épidémiologiquement inactifs (compartiment  $R$ ). Selon le type de résistances des plantes impliquées, il y a soit une protection totale avec une résistance qualitative qui permet de ne pas rentrer dans le compartiment  $L$  ou une protection partielle avec une résistance quantitative qui peut diminuer les effets de l'infection à plusieurs stades. (Rimbaud et al., 2018)

On note  $H_{i,v,t}$ ,  $L_{i,v,p,t}$ ,  $I_{i,v,p,t}$  et  $R_{i,v,p,t}$  les quantités respectives d'hôtes sains, latents,

infectieux et inactifs pour chaque champ  $i$ , variété  $v$ , pathotype  $p$ , au pas de temps  $t$  et  $Pr_{i,p,t}$  le nombre de propagules.

Concernant la croissance des hôtes, seuls les hôtes sains y contribuent. La croissance suit une fonction logistique :  $H_{i,v,t+1} = H_{i,v,t}[1 + \delta_v(1 - \frac{N_{i,v,t}}{K_{i,v}})]$  où  $\delta_v$  désigne le taux de croissance pour la variété  $v$ ,  $N_{i,v,t}$  le nombre total d'hôtes (somme sur tous les compartiments) et  $K_{i,v}$  la capacité de charge.

Le compartiment  $H$  est découpé en  $H^1$ , désignant les hôtes sains non-contaminés et  $H^2$  les hôtes sains contaminés par les propagules. Le nombre d'hôtes contaminés est établi par tirage dans une loi multinomiale avec comme paramètres  $H_{i,v,t}^{contaminable}$ , le nombre d'hôtes qu'il est possible de contaminer, et une proportion de propagules pour chaque pathotype :

$$[H_{i,v,t}^2]_{p=1,\dots,P} \sim Multinomial(H_{i,v,t}^{contaminable}, [\frac{Pr_{i,p,t}}{\sum_{p=1}^P Pr_{i,p,t}}]_{p=1,\dots,P})$$

Entre  $t$  et  $t + 1$ , les hôtes contaminés  $H_{i,v,p,t}^2$  subissent une infection selon une probabilité  $e_{v,p}$  :

$$H_{i,v,p,t}^2 \rightarrow L_{i,v,p,t+1} \sim Binomial(H_{i,v,t}^2, e_{v,p})$$

Les temps passés dans les compartiments  $L$  et  $I$  sont déterminés avec le tirage dans des lois *Gamma* permettant de tirer un temps dans l'intervalle  $[0; +\infty[$ . Pour le passage de  $L$  à  $I$ , on a  $\gamma_{exp_{v,p}}$  la valeur attendue et  $\gamma_{var}$  la variance :

$$(LI)_{v,p} \sim Gamma(\gamma_{exp_{v,p}}, \gamma_{var})$$

Le nombre de propagules produit par les hôtes infectés sont tirés selon une distribution de *Poisson* d'intensité  $r_{exp_{v,p}}$ , le nombre attendu de propagules.

$$Pr_{i,p,t} \sim Poisson(\sum_{v=1}^V r_{exp_{v,p}})$$

La mutation d'un pathotype à un autre se fait de proche en proche, c'est à dire via de petits sauts adaptatifs dans le cas des résistances quantitatives. Pour les résistances qualitatives, chaque pathotype peut gagner ou perdre en infectivité avec une certaine probabilité.

La dispersion des propagules dans le paysage se fait selon un noyau de dispersion (permettant des évènements de dispersion à longue distance) dépendant de la distance euclidienne entre les champs.

## 2.2 Déroulé d'une simulation

Sur le CRAN du package se trouve un exemple de simulation (vignette 1) et la liste de tous les paramètres.

### 2.2.1 Dynamique spatio-temporelle

Une des premières étapes à réaliser pour lancer une simulation est de choisir le nombre de saisons de récolte et le nombre de pas de temps par saison. Lancer une simulation sur plusieurs saisons permet d'étudier les effets "bottleneck" sur l'épidémie (goulot d'étranglements), qui correspondent à une réduction drastique de la population d'agents pathogènes causée par la réduction de celle des hôtes. Du fait de la dérive génétique qu'ils impliquent, ces goulots d'étranglements peuvent fortement moduler la composition génétique de la population des agents pathogènes et ainsi sa capacité d'adaptation à la résistance des plantes. Pour la définition du paysage, le choix est possible entre 5 paysages d'environ 150 champs. Une matrice de dispersion de l'agent pathogène est associée à ce paysage dépendant de la distance moyenne de dispersion.

### 2.2.2 Stratégies de résistance

On définit ensuite les différentes variétés présentes (aussi appelées cultivars) qui sont soit sensibles (la variété ne porte aucun gènes de résistance) soit résistantes (la variété porte un ou plusieurs gènes de résistance). Chaque variété a ses paramètres propres comme son taux de croissance ou sa densité maximale de plantation. Les gènes de résistance associés aux variétés résistantes peuvent être de type qualitatifs (Major Gene ou MG) et ciblent l'infectivité de l'agent pathogène ou quantitatifs (Quantitative Trait Loci ou QTL) et peuvent agir sur les quatre traits d'histoire de vie des agents pathogènes (baisser le taux d'infection, allonger la durée de la période de latence, baisser la durée de la période d'infectiosité, réduire la produc-

tion de propagules), ce qui a pour effet de diminuer leur agressivité (Rimbaud et al., 2021). Chaque gène a ses paramètres propres et détermine le nombre de niveaux d’agressivité de l’agent pathogène (deux pour un MG : adapté ou non à la résistance).

Ainsi, le modèle permet un déploiement des résistances à plusieurs échelles :

- à l’échelle du gène avec différents types de résistance
- à l’échelle de la plante avec du pyramidage (une variété peut porter plusieurs gènes)
- à l’échelle du champ avec les mélanges variétaux
- à l’échelle du paysage avec des mosaïques et un degré d’agrégation spatiale plus ou moins élevé (Figure 2)
- à l’échelle temporelle en intégrant des rotations dans le temps

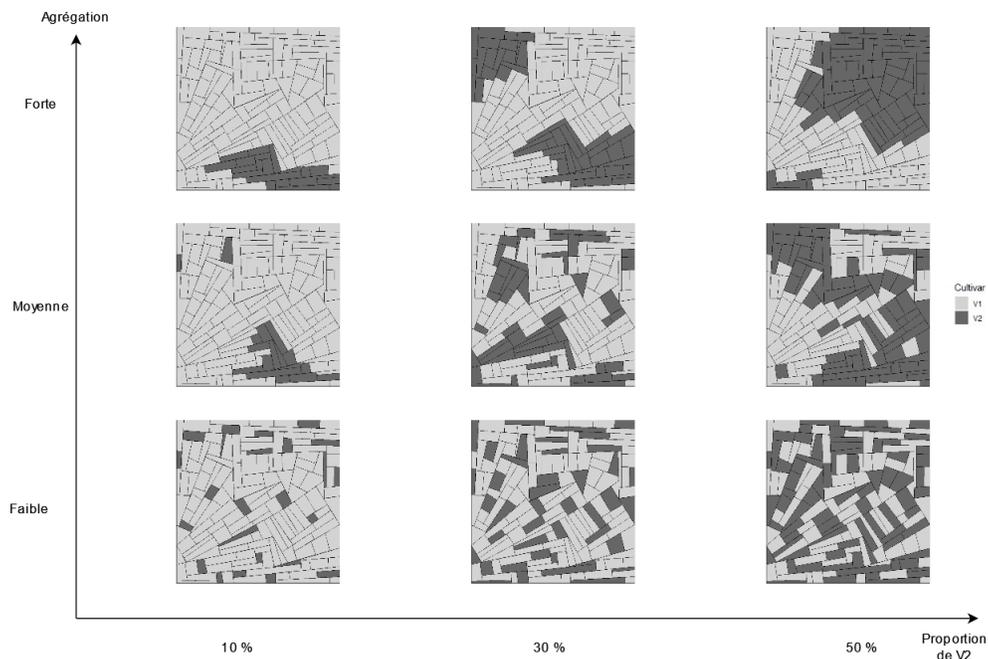


FIGURE 2 – Exemple de paysages générés avec *landsepi* avec 2 variétés V1 et V2, 3 niveaux d’agrégation spatiale et 3 proportions de V2. Chaque paysage contient environ 150 champs.

### 2.2.3 Paramétrage de l’agent pathogène

Une autre étape est le paramétrage de l’agent pathogène. Les paramètres associés sont liés à l’infectivité, la survie lors de l’inter-saison ou encore la reproduction. Dans le package *landsepi*, il est possible de choisir directement un type d’agent pathogène comme la rouille

brune des céréales, un champignon du genre *Puccinia* qui se reproduit par sporulation clonale (les paramètres sont fixés en conséquence).

Au début d'une simulation, l'agent pathogène inoculé n'est présent que sur les variétés sensibles et n'a pas encore contourné de résistances (agent pathogène avirulent). D'un point de vue biologique, le début d'une simulation coïncide alors avec le début d'une épidémie. L'inoculum de départ est défini par une probabilité qu'un hôte sensible soit infecté. Au cours de la simulation, des mutations peuvent survenir, faisant apparaître d'autres pathotypes qui amènent au contournement des résistances qualitatives (acquisition de gènes d'infectivité) et à l'érosion des résistances quantitatives (agressivité accrue). Néanmoins, un gain d'infectivité ou d'agressivité sur un hôte résistant se traduit souvent par une baisse d'agressivité sur les hôtes sensibles, appelée "coût en fitness".

#### 2.2.4 Sorties des simulations

De nombreux fichiers de sortie sont générés permettant d'étudier les simulations sous des aspects épidémiologiques ou évolutifs. Il est aussi possible de générer des graphes comme la dynamique de l'épidémie (Figure 3).

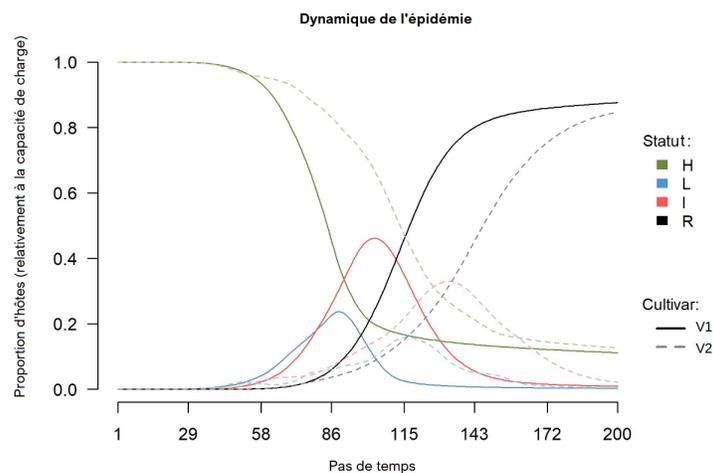


FIGURE 3 – *Dynamique d'une épidémie simulée avec landsepi sur une saison de 200 jours avec 2 variétés résistantes V1 et V2 ayant chacune un gène majeur de résistance. Il est possible de suivre l'évolution des hôtes dans chaque compartiment H,L,I et R pour chaque variété. Le pic d'infections de V2 arrive après celui de V1.*

Afin d'évaluer les différentes stratégies simulées, plusieurs critères sont possibles. D'un

point de vue épidémiologique, les meilleures stratégies sont celles où la proportion d'hôtes sains est maximale (rendement le plus élevé) et celle d'infectés minimale (prévalence de l'agent pathogène). D'un point de vue évolutif, il faut plutôt que les résistances soient les plus durables possible et donc que les agents pathogènes évoluent le plus lentement possible.

## 2.3 Modifications du modèle

La première étape de mon travail de stage a été de modifier le modèle décrit précédemment pour qu'il puisse répondre aux questions posées. La contrainte principale portait sur l'initialisation des simulations, qui correspondait à un contexte d'épidémie endémique (c'est-à-dire la présence d'un inoculum primaire généralisé dans l'ensemble du paysage). L'objectif des modifications était de modifier la paramétrisation de l'inoculum de départ afin d'avoir plus de flexibilité au niveau de la diversité des pathotypes présents et de la localisation de l'inoculum, et répondre ainsi à un plus large éventail de situations possibles. Les modifications que j'ai effectuées ont porté sur le code source du package *landsepi* (modèle défini dans des classes C++ et méthodes en R).

Une des modifications majeure dans le modèle a donc été de transformer la définition et la paramétrisation de l'inoculum de départ. Il était précédemment défini comme un scalaire, désignant la probabilité que les hôtes sensibles soient infectés au début de la simulation, sur tout le paysage. Cette définition a comme contrainte de ne pas pouvoir choisir les variétés infectées, les pathotypes présents ou la localisation de l'inoculum au début des simulations. Ainsi, l'inoculum est désormais entré comme un vecteur 3D (une matrice de dimensions NombreDeVariétés x NombreDePathotypes x NombreDeChamps) puis transformé en vecteur 1D dans la mémoire. Afin de faciliter la définition de l'inoculum pour l'utilisateur, il peut être défini via la fonction `setInoculum()` qui prend en input les pathotypes présents, les probabilités d'infections associées, les variétés infectées et les champs infectés. La plupart des paramètres sont optionnels de sorte que l'inoculum peut, si besoin, être défini de manière simple et rapide pour l'utilisateur. Les modifications effectuées permettent désormais d'inoculer différents pathotypes sur le paysage au début des simulations (avec leur probabilité d'infection associée) et donc que certaines résistances soient déjà contournées et qu'il n'y ait alors pas seulement les hôtes sensibles infectés. Elles permettent aussi de choisir les variétés

qui sont infectées au début des simulations et la localisation de l'inoculum, globale ou locale ainsi que d'avoir différents inoculums selon les champs et selon les variétés. De nouvelles vignettes (tutoriels) explicitant les fonctionnalités développées pendant le stage ont été intégrées au package (voir section modifiée de la vignette 1 en Annexes). La nouvelle version du package *landsepi* sera proposée au CRAN prochainement, afin qu'elle soit accessible à la communauté.

## 2.4 Plan des simulations

L'objectif est de lancer des simulations en se servant des modifications effectuées sur le modèle et d'étudier l'introduction et l'établissement d'un généraliste au sein d'un paysage hétérogène (cas d'étude 1), et la coexistence entre les différents pathotypes possibles (généraliste et spécialistes) sur le long terme (cas d'étude 2).

Au niveau de la structure spatiale, chaque champ est re-divisé en polygones plus petits nommés patchs (intersection d'une grille avec le paysage) afin de prendre en compte une dispersion à l'intérieur des champs (Figure 4). Le nombre de patchs est environ de 1180 (varie selon le paysage).

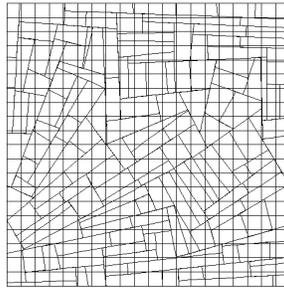


FIGURE 4 – Intersection d'une grille avec un paysage généré avec *landsepi* permettant de diviser les champs en unités plus petites nommés patchs. Cela permet de passer d'environ 150 champs à 1180 patchs.

Deux variétés (ou cultivars)  $V1$  et  $V2$  sont présentes dans le paysage, ayant chacune un gène de résistance majeur (respectivement MG1 et MG2). Par conséquent, l'agent pathogène ayant deux niveaux d'agressivité par rapport à chaque gène (avirulent ou virulent), cela induit l'existence de 4 pathotypes (avirulent, adapté à MG1, adapté à MG2 et adapté à MG1 et MG2). Etant donné qu'il n'y a pas de variété sensible, le pathotype avirulent, qui sera

quasi-inexistant, n'est pas considéré.  $P1$  et  $P2$  sont donc, respectivement, les pathotypes spécialistes de  $V1$  et  $V2$ , et  $P3$  le pathotype généraliste qui a contourné les 2 gènes de résistance. Le pathotype  $P3$  est donc adapté aux 2 variétés et perçoit le paysage comme homogène.

### 2.4.1 Cas d'études

Cas 1 : On s'intéresse à l'introduction et l'établissement du généraliste dans la population. Les pathotypes présents au début des simulations sont les spécialistes  $P1$  et  $P2$ . Le généraliste  $P3$  est alors introduit par mutation des spécialistes. La croissance des hôtes est fixée à 0 (pour étudier spécifiquement la phase d'établissement). A l'état initial, pour chaque simulation, un patch cultivé avec  $V1$  et un patch cultivé avec  $V2$  sont choisis au hasard parmi tous les patches. Sur chaque patch, on calcule la probabilité d'infection selon le nombre d'hôtes présents pour avoir en moyenne 10 infections de chaque pathotype.

Chaque simulation est effectuée sur une seule saison de 400 pas de temps. On fait varier différents paramètres dans les simulations afin de voir leur impact sur l'épidémie (Figure 5). Cela représente plus de 80 000 simulations, chaque simulation prenant environ 2 minutes.

Paramètre ( <i>nom_variable</i> )	Valeur
Niveau d'agrégation spatiale ( <i>aggreg</i> )	Faible(1), Moyenne(2), Forte(3)
Proportion de la variété V2 ( <i>prop</i> )	10%, 30%, 50%
Dispersion du pathogène ( <i>disp</i> )	2.5%, 10% et 25% du paysage
Efficacité ( <i>R_efficiency</i> )	0.9, 0.8, 0.7, 0.6, 0.5
Coût en fitness ( <i>fitness_cost</i> )	0.2, 0.15, 0.1, 0.05
Trait d'histoire de vie ciblé ( <i>target_trait</i> )	Taux d'infection (IR), Durée de latence (LAT), Durée d'infectiosité (IP), Production de propagules (PR)
Taux de mutation ( <i>mutation_rate</i> )	1e-5, 1e-3
Paysage	1:5
Allocation des variétés	1:2
Répétition	1:2

FIGURE 5 – Tableau des paramètres variables dans les simulations lancées avec *landsepi* pour le cas d'étude 1, étudiant l'introduction et l'établissement du pathotype généraliste. Pour chaque paramètre, on donne le nom de la variable associé et les valeurs des différentes variations. Au total, cela représente plus de 80 000 simulations différentes.

Les variables associées au paysage sont le niveau d'agrégation spatiale et la proportion de

la variété  $V2$  présente (Figure 2). Les variables associées à l'agent pathogène sont sa dispersion moyenne dans le paysage et le taux de mutation (qui permet l'apparition du généraliste). Les variables associées aux gènes de résistance sont l'efficacité (perte d'infectiosité d'un pathotype non adapté à la résistance ( $P2$  sur  $V1$  et  $P1$  sur  $V2$ )), le coût en fitness (perte d'infectiosité suite à l'adaptation à une résistance sur un hôte qui ne possède pas cette résistance ( $P3$  sur  $V1$  et  $V2$ )) (Figure 6) et le trait d'histoire de vie ciblé (effet de la résistance sur le cycle de vie de l'agent pathogène). Bien que les résistances implémentées soient qualitatives, ciblant en général l'infectivité du pathogène, on implémente ici des résistances sur les différents traits d'histoire de vie du pathogène.

Pour chaque combinaison de paramètres d'intérêt, 20 simulations équivalentes (5 paysages différents, 2 allocations et 2 répétitions du modèle) sont lancées afin de considérer la stochasticité du modèle et la variabilité paysagère (les paramètres mentionnés n'influent pas sur l'épidémie).

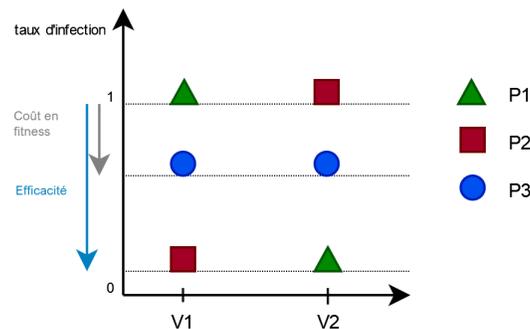


FIGURE 6 – Schéma explicitant les paramètres d'efficacité (perte d'infectiosité d'un pathotype non adapté à la résistance) et de coût en fitness (perte d'infectiosité suite à l'adaptation à une résistance sur un hôte qui ne possède pas cette résistance (ou virulence non nécessaire)) des gènes portés par les variétés  $V1$  et  $V2$ . Plus le coût en fitness est faible et l'efficacité forte, plus le généraliste  $P3$  est favorisé et les spécialistes  $P1$  et  $P2$  défavorisés.

Cas 2 : On s'intéresse ici à la coexistence éventuelle entre les 3 pathotypes. On considère qu'ils sont déjà tous présents dans la population et le taux de mutation est fixé à 0 en conséquence. La croissance des hôtes est fixée à 0.1 et l'inoculum est cette fois défini de manière globale : tous les patchs ont un inoculum de 1% pour chaque pathotype  $P1$ ,  $P2$ ,  $P3$ . Chaque simulation est effectuée sur une seule saison de 1000 pas de temps. Cela représente plus de 40 000 simulations, chaque simulation prenant environ 5 minutes. Les paramètres

variables sont les mêmes que le cas 1 (Figure 5) excepté le taux de mutation qui est nul dans ce cas.

## 2.5 Analyses statistiques

L'ensemble des analyses statistiques ont été réalisées sur R, version 4.2.2.

### 2.5.1 Cas d'étude 1

Pour le premier cas d'étude, la capacité du pathotype  $P3$  à s'établir dans le paysage est évaluée en calculant la fréquence du pathotype à la fin de la simulation dans chaque champ. Les compartiments  $L$ ,  $I$  et  $R$  sont sommés, ce qui permet de prendre en compte les infections sur toute la durée de l'épidémie. La fréquence moyenne de  $P3$  est calculée sur tous les champs ainsi que la fréquence maximale atteinte afin d'évaluer si le pathotype est distribué plutôt de façon diffuse dans le paysage (la moyenne et le maximum sont proches) ou plutôt localement (le maximum est bien supérieur à la moyenne). Cela est équivalent à calculer la variance de la fréquence de  $P3$  à l'échelle du champ où une variance élevée serait synonyme de l'établissement de populations locales et une variance faible une distribution plus homogène du généraliste dans le paysage. Ces fréquences moyennes et maximales sont définies comme les variables réponses de modèles statistiques (respectivement `mean_prop_P3` et `max_field_P3`).

Différents modèles ont été testés : Modèles Linéaires Généralisés (GLM) et Modèles Additifs Généralisés (GAM). Les GLM sont une généralisation des régressions linéaires en instaurant une fonction lien entre l'espérance du modèle probabiliste décrivant la distribution supposée des observations et le prédicteur linéaire (les variables explicatives). Les GAM sont une généralisation non linéaire des GLM et permettent la modélisation de relations non linéaires entre la variable réponse et les variables explicatives. Après comparaison des différents modèles, les modèles retenus sont des GLM avec régression polynomiale (une fonction polynomiale est construite pour chaque variable explicative) (package `orthopolynom`). La régression polynomiale est utilisée pour construire des espaces orthogonaux (polynômes de Legendre), permettant d'avoir une décomposition de la variance unique, indépendante de l'ordre d'apparition des variables dans le modèle. Les polynômes utilisés sont au maximum

de degré 2 (espace en 3 dimensions).

La fonction lien utilisée dans les GLM est une fonction *logit*, permettant de respecter la contrainte de l'intervalle  $[0, 1]$  pour une proportion. Un modèle est construit pour chaque niveau d'agrégation spatiale `agregg`, chaque taux de mutation `mutation_rate` et chaque trait d'histoire de vie ciblé `target_trait`. Le modèle suivant, faisant intervenir les variables explicatives proportion de la variété *V2* `prop`, efficacité de la résistance `R_efficiency`, coût en fitness `fitness_cost` et dispersion `disp`, est ajusté :

$$\begin{aligned} \text{logit}(\text{mean\_prop\_P3}) = & \beta_0 + \beta_1 * \text{prop} + \beta_2 * \text{prop}^2 + \beta_3 * \text{disp} + \beta_4 * \text{prop} * \text{disp} + \beta_5 * \\ & \text{disp}^2 + \beta_6 * \text{R\_efficiency} + \beta_7 * \text{prop} * \text{R\_efficiency} + \beta_8 * \text{disp} * \text{R\_efficiency} + \beta_9 * \\ & \text{R\_efficiency}^2 + \beta_{10} * \text{fitness\_cost} + \beta_{11} * \text{prop} * \text{fitness\_cost} + \beta_{12} * \text{disp} * \text{fitness\_cost} + \\ & \beta_{13} * \text{R\_efficiency} * \text{fitness\_cost} + \beta_{14} * \text{fitness\_cost}^2 + \varepsilon \end{aligned}$$

Pour chaque modèle, la proportion de déviance expliquée par le modèle ainsi que les indices de sensibilité sont calculés. La méthode utilisée pour calculer les indices de sensibilité est basée sur la décomposition de la variance du modèle, qui est unique (Saltelli et al., 2000). Les indices sont obtenus pour chaque variable du modèle en effectuant le ratio entre la variance expliquée par la variable (terme d'effet principal et interactions) et la variance globale. Les indices sont compris dans l'intervalle  $[0, 1]$ .

Il est aussi possible de faire de la prédiction à partir de l'estimation des paramètres du modèle en utilisant la fonction `predict.glm`. On prend les intervalles de départ pour la dispersion et la proportion de *V2* et les valeurs les plus favorables pour *P3* pour les autres variables (efficacité maximale et coût en fitness minimal). La visualisation des prédictions des variables réponses peut se faire dans un graphique de contours dans le plan (dispersion, proportion de *V2*). Ainsi, les situations les plus favorables pour l'établissement de *P3* peuvent être évaluées, selon la dispersion, la proportion et l'agrégation et les patterns pour la fréquence moyenne et maximale de *P3* comparés.

## 2.5.2 Cas d'étude 2

Pour le second cas d'étude, la coexistence des 3 pathotypes est évaluée en calculant la fréquence d'hôtes infectieux (compartiment *I*) pour chaque pathotype dans chaque champ

moyennée sur les 100 derniers pas de temps, lorsque l'état d'équilibre du système est atteint. La fréquence moyenne sur le paysage est calculée pour chaque pathotype afin d'avoir une seule valeur représentative qu'on note `mean_prop_P1`, `mean_prop_P2` et `mean_prop_P3`. Si la fréquence est inférieure à un seuil (fixé à 1%), le pathotype est considéré éteint.

Les différentes sorties possibles sont : P1, P2, P3, P1+P2, P1+P3, P2+P3, P1+P2+P3. On considère que les situations P1 et P2 sont équivalentes ainsi que P1+P3 et P2+P3 puisque les deux spécialistes possèdent les mêmes propriétés et capacités d'infections mais sur des variétés différentes (qui sont équivalentes). On regroupe donc le cas P1 et P2 en la sortie S (pour spécialiste) et P1+P3 et P2+P3 en P3+S. Les sorties des simulations peuvent alors se diviser en 5 catégories selon l'équilibre atteint (voir la Figure 7 pour un exemple).

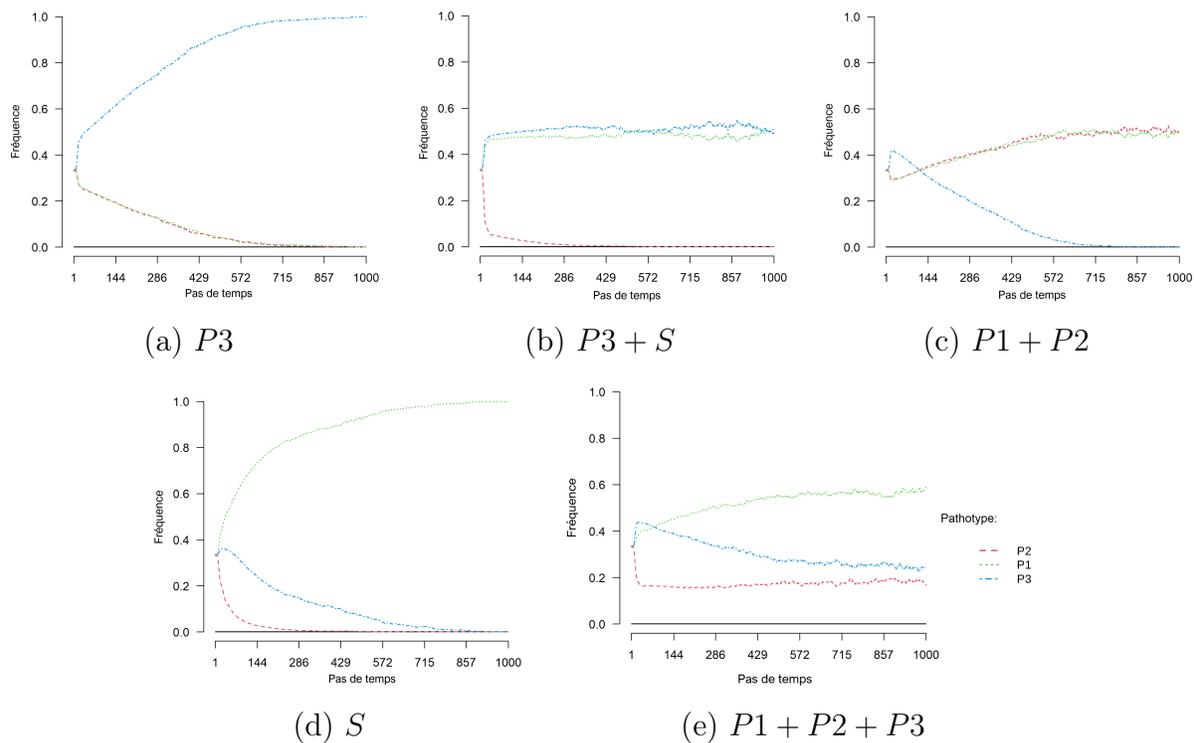


FIGURE 7 – Fréquences des 3 pathotypes P1, P2 (spécialistes) et P3 (généraliste) sur 5 exemples de simulations avec *landsepi* sur 1000 pas de temps et 2 variétés résistantes V1 et V2. Les 5 catégories de sorties possibles des simulations sont présentées ici (on considère les 100 derniers pas de temps). S désigne un des spécialistes P1 ou P2. On considère qu'un pathotype est éteint sous le seuil de 1%. On a soit coexistence entre plusieurs pathotypes (b,c,e), soit la persistance d'un unique pathotype dans le paysage (a,d).

Une variable catégorielle est construite (prenant les valeurs correspondant aux 5 sorties indiquées ci-dessus) et un modèle de régression multinomiale est utilisé pour analyser les

sorties des simulations (package `nnet`). Cette régression est basée sur un algorithme de classification. Le modèle de régression multinomiale est construit pour chaque trait d'histoire de vie (`target_trait`). Les variables explicatives sont l'agrégation spatiale `aggreg`, la dispersion `disp`, la proportion de la variété `V2 prop`, l'efficacité de la résistance `R_efficiency` et le coût en fitness `fitness_cost`. La régression est ensuite utilisée pour faire de la prédiction de la classe (et donc du type de coexistence) en fonction de la dispersion et de la proportion de `V2` en fixant l'efficacité et le coût en fitness, pour chaque niveau d'agrégation spatiale. Les classes obtenues sont tracées dans le plan (dispersion, proportion de `V2`).

## 3 Résultats

### 3.1 Cas d'étude 1

#### 3.1.1 Analyse de sensibilité

Les résultats sont présentés pour le trait d'histoire de vie IR (résistance ciblant le taux d'infection) et le taux de mutation le plus élevé ( $1.10^{-3}$ ), qu'on considère comme référence par rapport aux autres traits d'histoire de vie et l'autre taux de mutation.

Les estimations des coefficients des modèles sont présents en *Annexes*, nous présentons ici les résultats de l'analyse de sensibilité. La déviance expliquée par le modèle (*Annexes*) est de plus en plus faible quand l'agrégation spatiale augmente, que ce soit pour la fréquence moyenne ou maximale de `P3`, comme l'atteste aussi la part de variance résiduelle dans les diagrammes en barres (Figure 8), signifiant que le modèle explique de moins en moins bien les données. Les indices de sensibilité pour la fréquence moyenne (Figure 8) indiquent que l'efficacité est la variable qui contribue le plus aux variations de la fréquence moyenne de `P3` pour une agrégation faible et moyenne alors que c'est le coût en fitness pour une forte agrégation. La contribution de certaines variables augmente lorsque le paysage devient de plus en plus agrégé (coût en fitness) tandis que d'autres diminuent (efficacité, dispersion). Concernant la contribution des variables pour la fréquence maximale (Figure 8), la variable qui contribue le plus pour une agrégation faible et moyenne est l'efficacité, comme pour la fréquence moyenne. Cependant, pour une agrégation forte, c'est la dispersion. On peut

noter que l'importance de la contribution de la dispersion varie dans le sens opposé selon l'agrégation comparé à la fréquence moyenne : ici la contribution augmente avec l'agrégation. La contribution de la variable proportion de  $V2$ , elle, diminue avec l'agrégation.

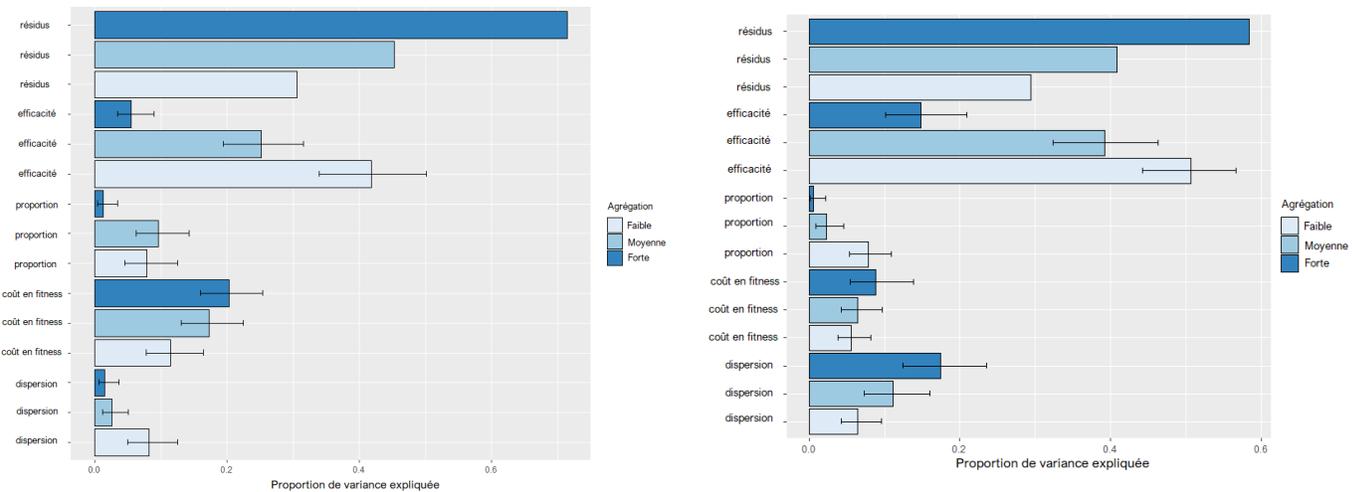


FIGURE 8 – Diagramme en barres des contributions de chaque variable à la variance du GLM de la fréquence moyenne (gauche) et maximale de  $P3$  par champ (droite), pour les 3 niveaux d'agrégation spatiale (trait=IR et taux de mutation= $1.10^{-3}$ )

### 3.1.2 Prédiction de la structure de la population de $P3$

Sur les graphiques de contours (Figure 9), les patterns pour la fréquence moyenne et maximale de  $P3$  et donc les situations les plus favorables diffèrent selon la proportion de la variété  $V2$  et la capacité de dispersion de l'agent pathogène. Pour la fréquence moyenne, d'après les graphiques de contours, la situation la plus favorable pour une agrégation faible et moyenne semble être favorisée par une dispersion élevée et une proportion de  $V2$  entre 30 à 50%. Même si une dispersion faible semble être favorable pour une agrégation forte, les valeurs sont globalement faibles (moins de 3%) et l'intervalle de valeurs peu étendu (variation de 1% entre le min et le max). Pour la fréquence maximale, pour une agrégation moyenne et forte, une dispersion faible est favorable et la proportion a relativement peu d'impact. En revanche, une proportion faible semble être favorable pour une agrégation faible mais la dispersion a peu d'effets.

En évaluant la variance de la fréquence de  $P3$  par champs (voir Annexes) (ou en comparant la fréquence moyenne et maximale de  $P3$  sur les graphiques de contours), il est possible d'évaluer l'établissement local du généraliste. La variance de la fréquence de  $P3$  diminue avec

l'agrégation spatiale ce qui signifie qu'une agrégation spatiale faible est plus favorable à des établissements locaux du généraliste. On peut illustrer un établissement local du généraliste avec la Figure 10 (point noir du graphique de contours correspondant à une agrégation moyenne, une dispersion faible et une proportion des variétés très déséquilibrée). Sur le graphique de contours, la fréquence moyenne de  $P3$  est autour de 3% alors que la fréquence maximale atteint 50%, ce qui peut être observé avec la Figure 10. Une situation d'agrégation moyenne et une proportion des variétés déséquilibrée est propice à la création d'îlots de la variété minoritaire dans le paysage. Dans cet exemple, l'inoculation de  $P2$  se fait dans l'îlot en haut à droite et la conquête de l'îlot en bas à gauche est alors plus difficile étant donné une dispersion faible, ce qui permet au généraliste de se développer et de créer une population locale sur plusieurs champs accolés tout en étant très peu présent sur le reste du paysage.

Les graphiques de contours peuvent aussi être obtenus pour le cas le moins favorable pour  $P3$  (efficacité minimale et coût en fitness maximal) (Annexes). Les fréquences de  $P3$  sont plus faibles que dans le cas précédent, voire négligeables pour la fréquence moyenne (1% ou moins). Pour la fréquence maximale de  $P3$ , les fréquences les plus élevées correspondent à des dispersions faibles pour les 3 agrégations avec un impact marginal de la proportion. On retrouve les patterns observés pour le cas le plus favorable pour  $P3$  des agrégations moyenne et forte.

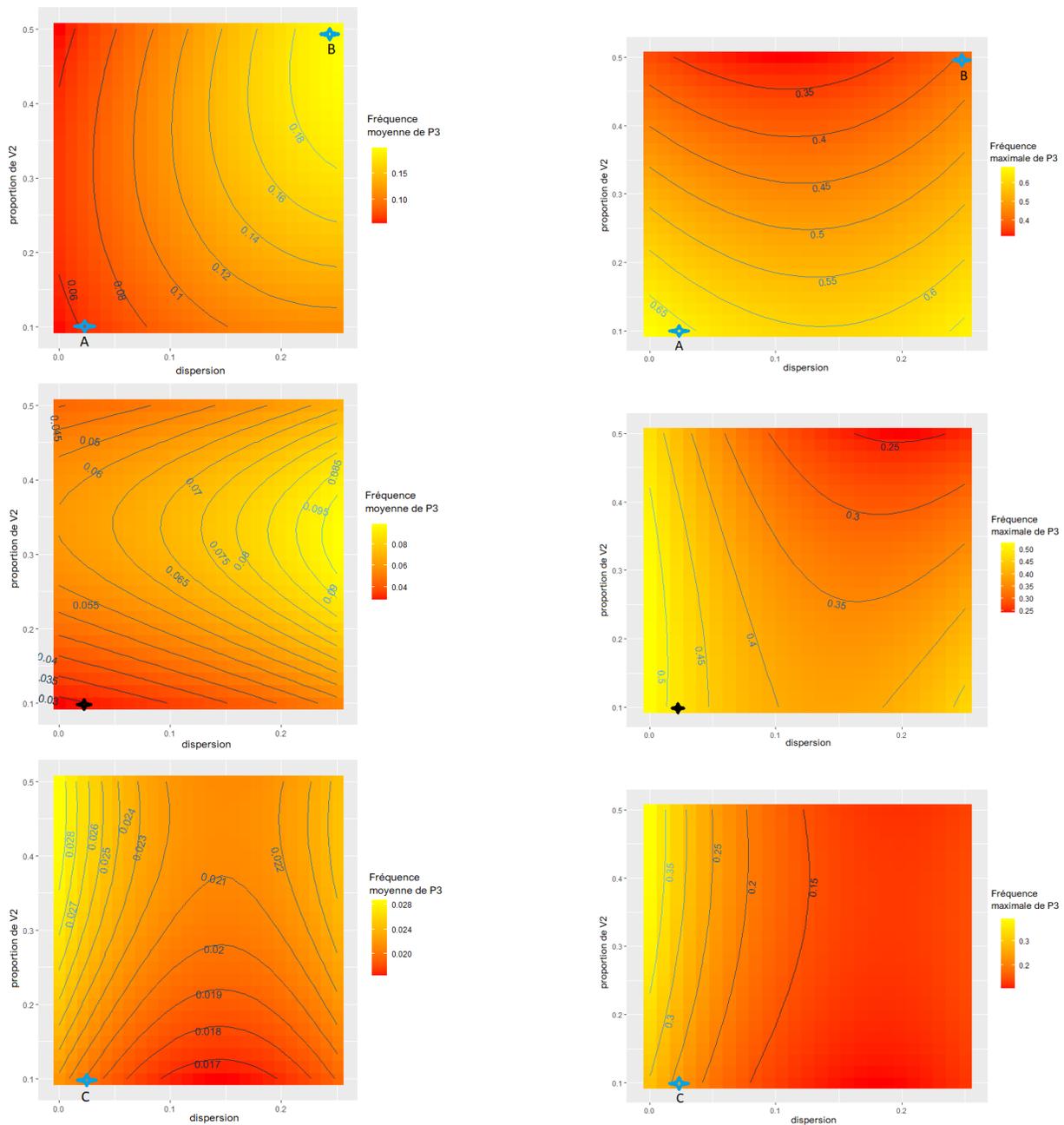


FIGURE 9 – Graphique de contours des prédictions de la fréquence moyenne (à gauche) et maximale (à droite) de  $P_3$  à partir de l'estimation des paramètres du modèle avec l'intervalle de départ pour la dispersion et la proportion de  $V_2$ , l'efficacité au maximum et le coût en fitness au minimum pour les 3 niveaux d'agrégation spatiale (de haut en bas : faible, moyenne et forte) (trait=IR et taux de mutation= $1.10^{-3}$ ). Le point noir est illustré Figure 10. Les points bleus sont illustrés par des exemples de simulations en Annexes.

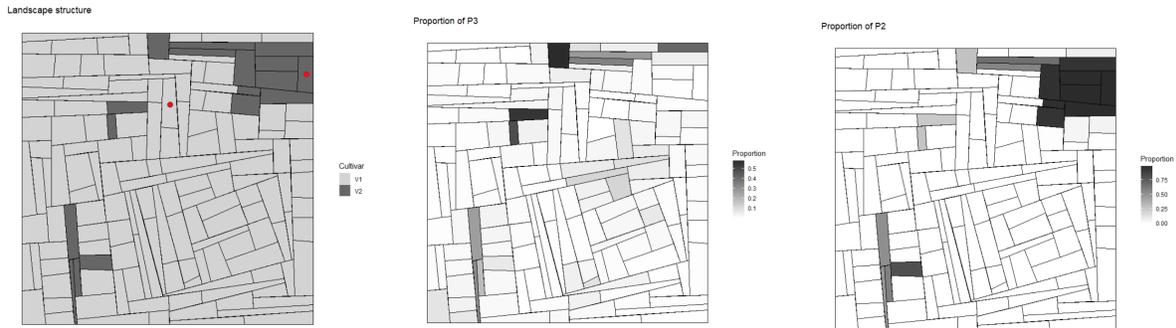


FIGURE 10 – Exemple d’une simulation (point noir sur la Figure 9) avec *landsepi* (structure du paysage (gauche) et fréquences de  $P3$  et  $P2$  (droite)) ( $\text{aggreg}=2$ ,  $\text{prop}=10\%$ ,  $\text{disp}=0.025$ ,  $\text{max}(\text{R\_efficiency})$ ,  $\text{min}(\text{fitness\_cost})$ ). Sur chaque point rouge sont inoculés  $P1$  et  $P2$ . Cet exemple illustre l’effet de la présence d’îlots dans le paysage sur le développement local du généraliste.

### 3.1.3 Effet du trait d’histoire de vie

Les sorties du modèle pour les autres traits d’histoire de vie ( $LAT$ ,  $IP$ ,  $PR$ , voir Figure 5) sont très proches de celles observées avec le trait  $IR$  (voir boxplots en Annexes). L’importance des contributions de chaque variable est comparable, d’après l’analyse de sensibilité et les patterns quasi identiques dans les graphiques de contours.

### 3.1.4 Effet du taux de mutation

Avec un taux de mutation de  $1.10^{-5}$  (voir analyse de sensibilité en Annexes), la variance expliquée par les variables du modèle est quasi nulle, que ce soit pour la fréquence moyenne ou maximale de  $P3$ . Les valeurs atteintes sont en moyenne 100 fois inférieures par rapport au taux de mutation de  $1.10^{-3}$  (voir les graphiques de contours en Annexes). L’apparition de  $P3$  est donc rare et se fait de façon opportuniste, et les valeurs restent trop faibles pour les analyser.

### 3.1.5 Agrégation et fréquence locales

En plus du niveau d’agrégation spatiale, qui est à l’échelle du paysage, on peut évaluer une agrégation plus locale, au niveau des champs, pour chaque paysage différent. Pour cela, on établit, pour chaque champ, la proportion de champs voisins (la taille du voisinage dépend de la dispersion du pathogène) avec la même variété (l’agrégation locale est donc dans l’intervalle

$[0, 1]$ ). Il est alors possible d'étudier la relation entre la fréquence locale (fréquence de  $P3$  par champs) et l'agrégation locale. (Figure 11).

Ces sorties à l'échelle locale ont été analysées à l'aide d'un modèle mixte prenant en compte la dépendance entre parcelles issues d'une même simulation. Ceci s'est fait en introduisant un effet aléatoire structuré par le numéro de simulation en plus des variables explicatives suivantes : le niveau d'agrégation spatiale (`agregg`), la proportion de la variété  $V2$  (`prop`), l'agrégation locale (`loc_aggreg`). Cette analyse (Annexes) permet de montrer que la fréquence locale de  $P3$  baisse significativement avec l'agrégation du paysage (effet non significatif pour une agrégation faible) mais aussi avec l'agrégation locale, comme on pouvait le remarquer sur le graphe.

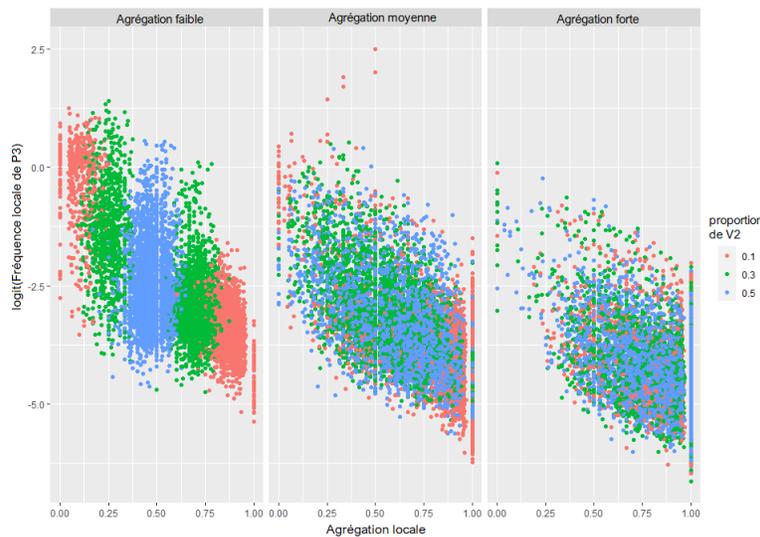


FIGURE 11 – Fréquence locale de  $P3$  (échelle logit) issue des simulations avec `landsepi` en fonction de l'agrégation locale à l'échelle du champ pour les 3 niveaux d'agrégation spatiale et les 3 proportions de  $V2$

## 3.2 Cas d'étude 2

### 3.2.1 Sorties du modèle

On présente les résultats pour le trait d'histoire de vie IR (résistance ciblant le taux d'infection), les résultats pour les autres traits d'histoire de vie étant très proches. Les estimations des coefficients des modèles sont en Annexes, pour les différentes classes.

Les figures sont symétriques par rapport à l'axe `prop=0.5` puisque les variétés  $V1$  et  $V2$

sont interchangeable (caractéristiques identiques). Sur les figures, les zones des classes sont bien délimitées (en raison du seuil) mais en réalité, les fréquences des différents pathotypes augmentent ou diminuent de manière graduelle. Il existe aussi des hétérogénéités à l'intérieur des classes puisqu'une coexistence des 3 pathotypes peut englober à la fois des situations où les 3 pathotypes ont chacun une fréquence d'environ 1/3 mais aussi des situations où le généraliste est à 98% et les 2 spécialistes à 1%.

### Cas le plus favorable à P3

Le cas le plus favorable à  $P3$  est obtenu en fixant l'efficacité au maximum et le coût en fitness au minimum (Figure 12 et 13). Seulement trois catégories sont alors prédites :  $P3$ ,  $P3 + S$  et  $P1 + P2 + P3$  (Figure 12). Pour le paysage le moins agrégé, seulement 2 classes sont présentes :  $P3 + S$  et  $P3$  seul, la situation la plus fréquente. Les spécialistes ne se maintiennent que dans les cas où la variété correspondante est très majoritaire dans le paysage. Plus l'agrégation augmente, plus la situation où le généraliste se maintient seul diminue et plus la coexistence des 3 pathotypes est récurrente. Aussi, plus la dispersion est faible, plus la coexistence des 3 pathotypes est favorisée (une proportion des variétés équilibrée semble aussi préférable). En comparant les prédictions pour un seuil de 1% (Figure 18) aux prédictions pour un seuil de 5% (Annexes), les patterns de coexistence sont très proches.

Pour la coexistence du généraliste et d'un spécialiste, lorsque le paysage est peu agrégé, la répartition des pathotypes peut néanmoins être assez agrégée (exemple en Figure 14). En effet, le généraliste  $P3$  remplace le spécialiste de la variété la moins présente et occupe les zones du paysages où cette variété est plus abondante. Dans un paysage plus agrégé (exemple en Figure 15),  $P3$  occupe clairement la niche de  $P2$ , en débordant même à la frontière avec l'autre variété. Une zone particulièrement occupée par le généraliste est située en bas à gauche du paysage où la structure spatiale agit comme un obstacle pour le spécialiste  $P1$ , au profit du généraliste. En cas de coexistence des 3 pathotypes (exemple en Figure 16), si le paysage est agrégé, chacun des spécialistes est concentré sur sa variété et le généraliste est présent à la frontière des 2 territoires, là où  $V1$  est en proximité avec  $V2$ . Dans un paysage un peu moins agrégé (exemple en Figure 17), le généraliste occupe une place centrale dans le paysage, là où les 2 variétés sont assez mélangées alors que les spécialistes  $P1$  et  $P2$  occupent les zones

là où le paysage est localement homogène.

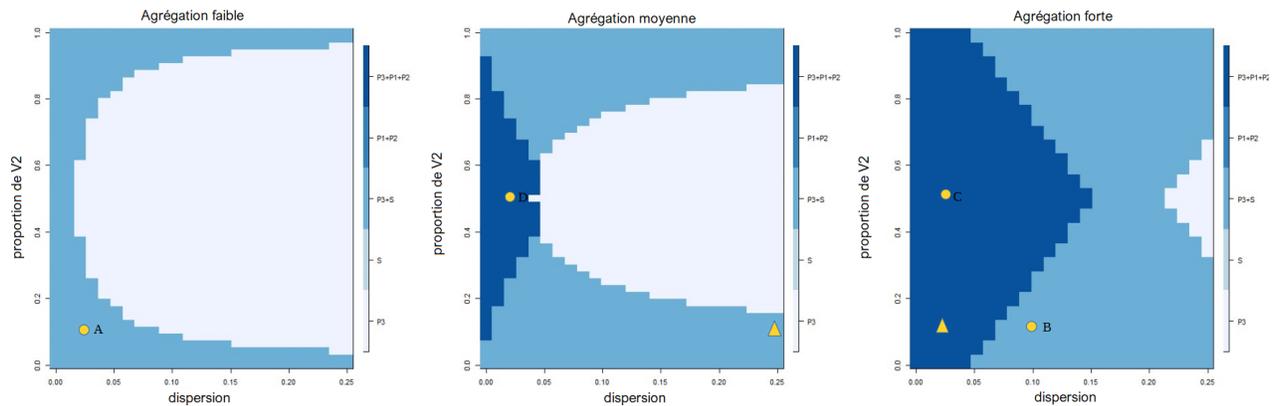


FIGURE 12 – Coexistence entre pathotypes obtenue par prédiction avec les estimations des paramètres du modèle de régression multinomiale selon les 3 niveaux d'agrégation spatiale, les intervalles de départ pour la dispersion et la proportion de V2. L'efficacité est au maximum et le coût en fitness au minimum. Les situations correspondant aux points jaunes sont des figures ci-dessous et les triangles sont illustrés en Annexes. La variable catégorielle de classe (variable réponse du modèle) a été construite avec un seuil de 1%, signifiant qu'un pathotype est considéré éteint sous ce seuil.

Sortie	P3	S	P3+S	P1+P2	P1+P2+P3
Pourcentage de simulations	45.7%	0%	32.2%	0%	22.0%

FIGURE 13 – Pourcentages de simulations appartenant à chaque catégorie de sortie. L'efficacité est au maximum et le coût en fitness au minimum. Un pathotype est considéré éteint sous le seuil de 1% et sa fréquence est calculée sur les 100 derniers pas de temps

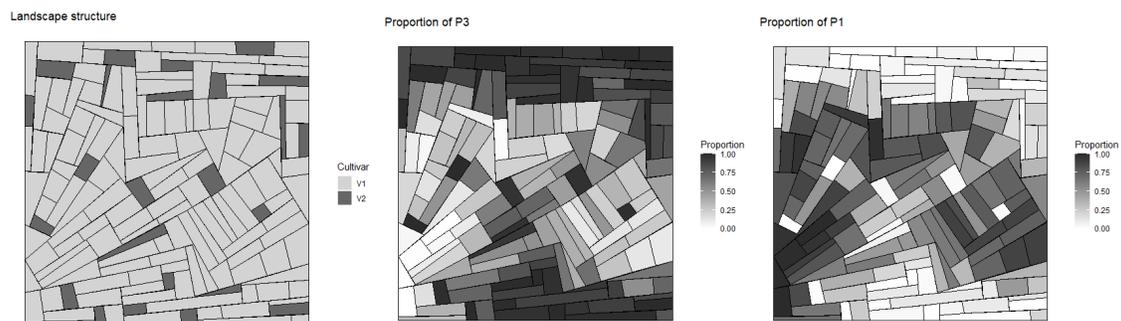


FIGURE 14 – Exemple de simulation (cf point A sur la Figure 12) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P3 et P1), résultant en la coexistence du généraliste et un des spécialistes (aggreg=1, prop=10%, disp =0.025, max(R\_efficiency), min(fitness\_cost)).

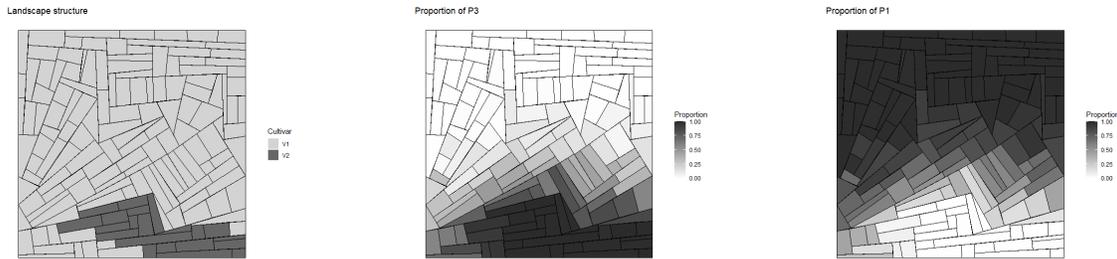


FIGURE 15 – Exemple de simulation (cf point B sur la Figure 12) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P3 et P1) résultant en la coexistence du généraliste et un des spécialistes ( $\text{aggreg}=3$ ,  $\text{prop}=10\%$ ,  $\text{disp}=0.1$ ,  $\text{max}(\text{R\_efficacy})$ ,  $\text{min}(\text{fitness\_cost})$ )

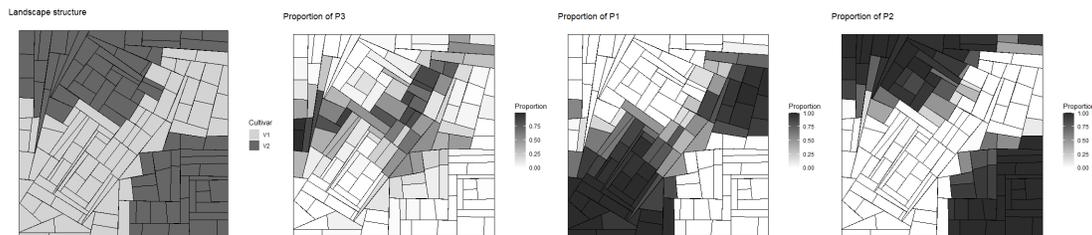


FIGURE 16 – Exemple de simulation (cf point C sur la Figure 12) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P3, P1 et P2) résultant en la coexistence des 3 pathotypes ( $\text{aggreg}=3$ ,  $\text{prop}=50\%$ ,  $\text{disp}=0.025$ ,  $\text{max}(\text{R\_efficacy})$ ,  $\text{min}(\text{fitness\_cost})$ ).

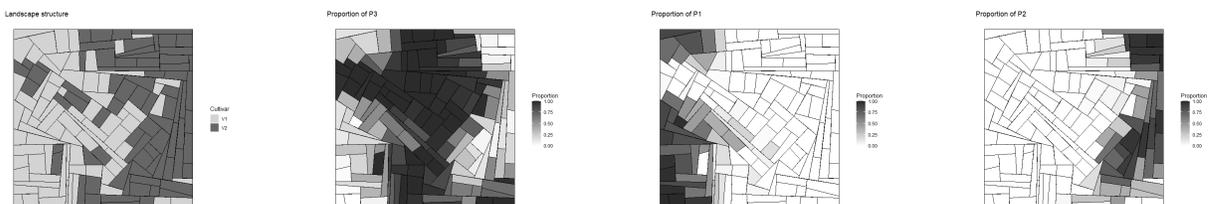


FIGURE 17 – Exemple de simulation (cf point D sur la Figure 12) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P3, P1 et P2) résultant en la coexistence des 3 pathotypes ( $\text{aggreg}=2$ ,  $\text{prop}=50\%$ ,  $\text{disp}=0.025$ ,  $\text{max}(\text{R\_efficacy})$ ,  $\text{min}(\text{fitness\_cost})$ )

### Cas le moins favorable à P3

Si les prédictions sont établies avec d'autres valeurs d'efficacité et coût en fitness (valeurs qui favorisent les spécialistes), les profils sont très différents pour la coexistence (Figure 18).

Pour le paysage le moins agrégé, les 5 classes peuvent apparaître selon les valeurs de dispersion et proportion de V2. En comparant les prédictions pour un seuil de 1% (Figure 18) aux prédictions pour un seuil de 5% (Annexes), la différence notable est que la coexistence des 3 pathotypes est absente pour une agrégation faible : cela signifie que le généraliste apparaît mais à une fréquence marginale (moins de 5%).

Plus l'agrégation augmente, plus la situation où les 2 spécialistes coexistent seuls est majoritaire jusqu'à être l'unique sortie prédite dans le cas le plus agrégé. Le fait que le généraliste soit moins présent pour ces valeurs d'efficacité et coût en fitness est attendu puisqu'elles favorisent les spécialistes et dévalorisent le généraliste (aussi visible dans les pourcentages des classes obtenues, Figure 19).

Lorsque la coexistence se fait entre les 2 spécialistes, chaque spécialiste infecte sa variété propre. Dans le cas de la Figure 20, cela ne se fait pas strictement puisque l'efficacité étant minimale, des infections croisées sont possibles : les spécialistes peuvent infecter plus facilement l'autre variété ( $P2$  sur  $V1$  et  $P1$  sur  $V2$ ). Par exemple, le champ  $V1$  en bas à droite, entouré de champs  $V2$  permet à  $P2$  de s'établir de manière accrue localement.

Pour la coexistence des 3 pathotypes, le généraliste peut être présent de façon beaucoup plus diffuse dans le paysage comparé au cas le plus favorable pour le généraliste, et distribué de façon assez homogène (Figure 21).

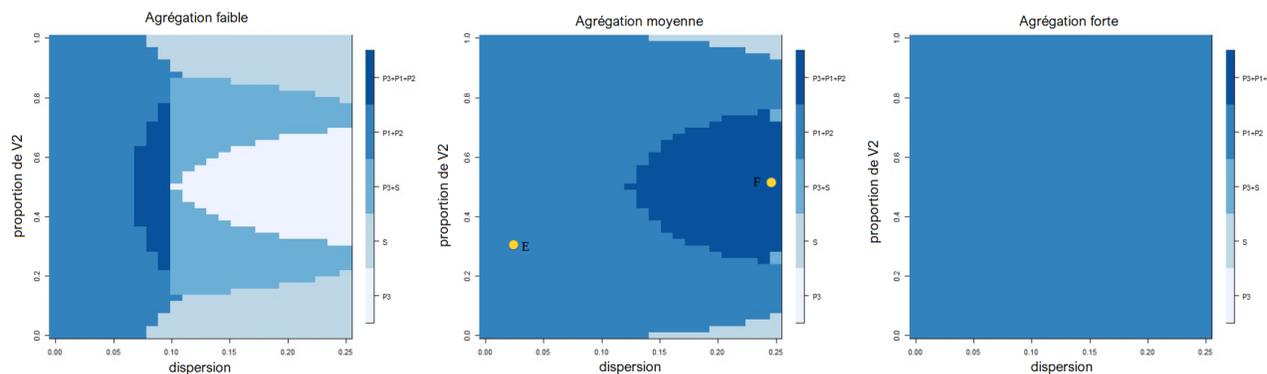


FIGURE 18 – Coexistence entre pathotypes obtenue par prédiction avec les estimations des paramètres du modèle de régression multinomiale selon les 3 niveaux d'agrégation spatiale, les intervalles de départ pour la dispersion et la proportion de  $V2$ . L'efficacité est au minimum et le coût en fitness au maximum. Les situations correspondant aux points jaunes sont des figures ci-dessous. La variable catégorielle de classe (variable réponse du modèle) a été construite avec un seuil de 1%, signifiant qu'un pathotype est considéré éteint sous ce seuil.

Sortie	P3	S	P3+S	P1+P2	P1+P2+P3
Pourcentage de simulations	5%	6.8%	11.1%	69.4%	7.6%

FIGURE 19 – Pourcentages de simulations appartenant à chaque catégorie de sortie. L'efficacité  $R\_efficiency$  est au minimum et le coût en fitness  $fitness\_cost$  au maximum. Un pathotype est considéré éteint sous le seuil de 1% et sa fréquence est calculée sur les 100 derniers pas de temps

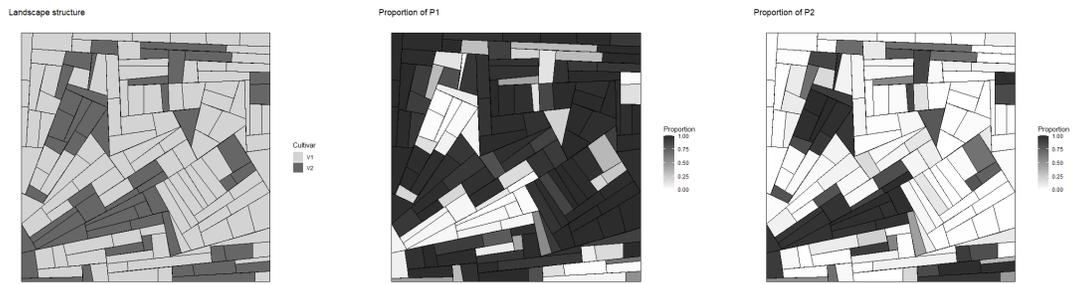


FIGURE 20 – Exemple de simulation (cf point E sur la Figure 18) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P1 et P2) résultant en la coexistence des 2 spécialistes (aggreg=2, prop=30%, disp =0.025, min(R\_efficiency), max(fitness\_cost)).

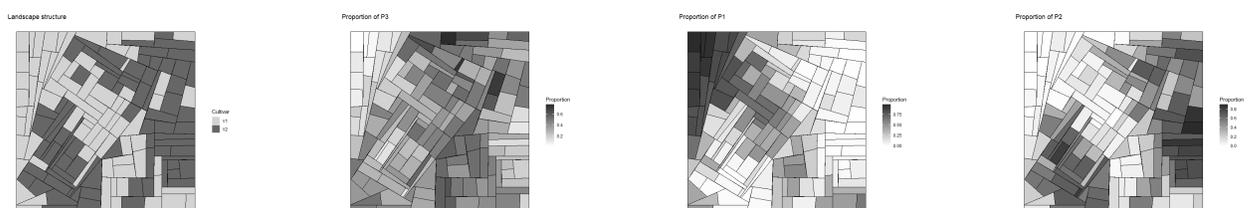


FIGURE 21 – Exemple de simulation (cf point F sur la Figure 18) avec landsepi (structure du paysage (gauche) et fréquences des pathotypes P3, P1 et P2) résultant en la coexistence des 3 pathotypes (aggreg=2, prop=50%, , disp =0.25, min(R\_efficiency), max(fitness\_cost))

### 3.2.2 Agrégation et fréquence locales

De la même façon que pour le cas d'étude 1, la relation entre la fréquence locale (fréquence des 3 pathotypes par champ) et l'agrégation locale peut être évaluée (Figure 22). Plus l'agrégation locale est élevée, plus la fréquence locale des spécialistes sur leurs champs associés peut atteindre des valeurs importantes. Pour le généraliste, en revanche, une agrégation locale faible semble garantir une fréquence locale élevée, comme pour le cas d'étude 1. La différence de distribution des fréquences associées aux champs V1 et V2 est dû au fait que les champs V2 sont minoritaires dans les situations où la proportion de V2 est 30% et 10% ce qui favorise la présence du généraliste dans les champs V2.

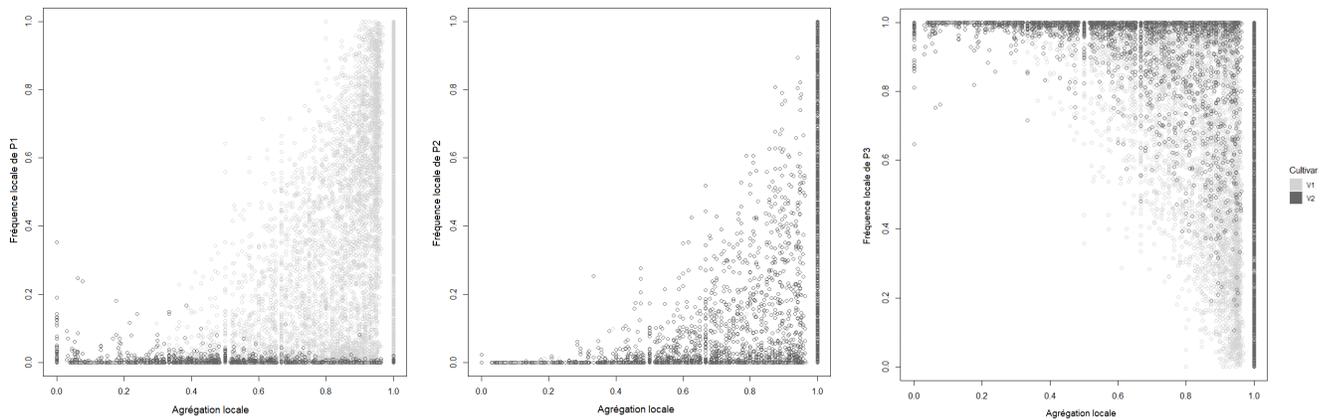


FIGURE 22 – *Fréquences locales des 3 pathotypes en fonction de l'agrégation locale issues des simulations avec landsepi ( $\max(\mathbf{R\_efficiency})$ ,  $\min(\mathbf{fitness\_cost})$ ,  $\mathbf{disp}=0.1$ ) selon les champs cultivés avec V1 ou V2*

## 4 Discussion

D'après les sorties des simulations, l'établissement du pathotype généraliste et sa coexistence avec les autres pathotypes est variable selon la structure du paysage.

Au niveau de l'introduction et du succès d'établissement (cas d'étude 1), moins le paysage est agrégé, plus le généraliste arrive à s'établir, que ce soit sur tout le paysage ou localement. Une forte dispersion paraît aussi faciliter le succès de son établissement à l'échelle du paysage, ainsi qu'un équilibre entre les 2 variétés. Les paramètres favorables à l'établissement local sont différents selon l'agrégation du paysage. Pour les paysages moins agrégés, le succès d'établissement local est favorisé par une proportion très déséquilibrée entre les variétés qui va freiner la propagation du pathotype spécialiste dans les champs, isolés, de la variété minoritaire. Pour des paysages plus agrégés c'est surtout une dispersion faible de l'agent pathogène qui favorise l'établissement local, rendant mécaniquement la conquête d'îlots plus difficile pour le spécialiste de la variété minoritaire et favorise alors l'établissement de populations locales du généraliste (Thrall et al., 2003).

Concernant la coexistence des pathotypes (cas d'étude 2), un paysage agrégé ainsi qu'une dispersion pas trop importante favorisent une forte diversité. Lorsque les 2 spécialistes coexistent, chacun occupe son habitat propre selon le concept de partitionnement de niche (Chesson, 2000). Cependant lorsque des infections croisées sont facilitées (efficacité faible),

les structures spatiales qui isolent des champs d'une variété permettent d'obtenir des fréquences locales plus élevées du pathotype non adapté à la variété sur les champs en question (Figure 20, champ en bas à droite).

Le partitionnement de niche est aussi observé pour la coexistence du généraliste et d'un des spécialistes, où le généraliste occupe la niche du spécialiste éteint. Les situations où le généraliste supplante un des spécialistes sont facilitées par une efficacité forte et un coût en fitness faible qui favorisent le généraliste. Pour des valeurs d'efficacité et coût en fitness moins favorables, il est aussi possible de constater cette situation mais la dispersion doit être suffisamment élevée, la proportion entre variétés déséquilibrée et l'agrégation spatiale faible. Ces coexistences respectent la notion que  $n$  espèces peuvent cohabiter sur  $n$  habitats (Karlin et al., 1972; Strobeck, 1979).

La coexistence entre les 3 pathotypes sur les deux variétés présentes dans le paysage est cependant assez fréquente et est même la situation la plus récurrente par rapport aux autres coexistences. Le généraliste maintient une population à la frontière des différents habitats lorsque le paysage est vraiment agrégé et profite ainsi des zones les plus hétérogènes. Cette cohabitation est cohérente avec le concept de "habitat boundary polymorphism" (Débarre et al., 2011) où la règle des  $n$  espèces pour  $n$  habitats n'est pas respectée. Les conditions les plus favorables à cette situation sont une dispersion limitée et un paysage suffisamment agrégé, tout en gardant une certaine hétérogénéité pour avoir des frontières entre les habitats.

En amont de prises de décision concernant l'agencement d'un paysage agricole, il apparaît crucial de déterminer si les pathotypes de type généraliste sont préférables aux spécialistes dans un contexte donné au niveau de la gravité des épidémies. Dans des paysages suffisamment hétérogènes, des profils spécialistes peuvent être plus souhaitables en se servant de l'effet dilution pour limiter les propagations. Cet effet n'agit que très peu sur le généraliste puisqu'il peut infecter les deux variétés et voir sa dynamique moins contrainte. En revanche, il a été montré dans d'autres cas que les spécialistes avaient des effets plus dévastateurs dans un paysage d'hôtes sensibles comparé à un généraliste (Papaix et al., 2011).

## 5 Perspectives

Que ce soit pour l'établissement des pathotypes ou leur coexistence, la structure spatiale des paysages semble déterminante pour freiner la propagation spatiale d'un pathotype. Par exemple, la présence de champs ou d'îlots isolés agissent comme obstacles pour les spécialistes, au profit du généraliste. Compléter la présente étude par une analyse des impacts des différents pathotypes permettrait d'identifier des organisations spatiales des paysages à même de limiter les épidémies. L'hétérogénéité spatiale dans le paysage et les structures locales dans l'agencement des différentes variétés implique aussi que la localisation des inoculums de départ peut avoir un effet très important sur l'épidémie. En effet, pour l'établissement de *P3*, les propagations des pathotypes sera vraisemblablement différente si les inoculums se trouvent plutôt en périphérie du paysage ou en son centre et si le champ inoculé est isolé ou non. Les modifications que j'ai apportées au modèle permettront néanmoins d'étudier cette question dans de futurs travaux.

Ensuite, il serait intéressant d'étudier si la coexistence entre les pathotypes peut être modifiée en présence d'un goulot d'étranglement causé par l'inter-saison. Pour ce faire il faudrait étudier l'évolution des fréquences des différents pathotypes sur plusieurs saisons de récolte pour se rapprocher de situations réelles avant de passer à des expériences de terrain. De même, ajouter des paramètres supplémentaires dans les simulations tels que des conditions biotiques (température, humidité...) pourrait se révéler utile pour pouvoir appliquer les résultats à des climats différents ou encore implémenter des traitements phytosanitaires (en cours de développement dans *landsepi*) pour trouver des compromis entre les différentes méthodes de lutte. Enfin, il serait judicieux d'incorporer dans l'étude des variétés portant des gènes de résistance quantitatifs. En effet, plusieurs études (comme Brun et al., 2009 ; Pilet-Nayel et al., 2017) montrent la synergie entre résistance qualitative et quantitative lorsque des variétés pyramidées sont utilisées.

## 6 Conclusion

Cette étude a permis de mieux comprendre l'émergence d'un agent pathogène généraliste et les situations où plusieurs pathotypes coexistent, quels sont les concepts écologiques impliqués dans ces phénomènes et les différents paramètres qui les favorisent. Certains paramètres comme l'agrégation spatiale du paysage ou les proportions des variétés sont des paramètres sur lesquels il est possible d'influer assez facilement en agriculture. Ainsi, si un pathotype de type généraliste apparaît moins virulent que les pathotypes spécialistes, un paysage peu agrégé et une proportion des variétés équilibrée paraît plus adapté mais cela implique souvent une population du pathotype plus diffuse dans le paysage. En revanche, si des pathotypes spécialistes sont préférables, un paysage plus agrégé semble plus adapté. Ces résultats dépendent toutefois de nombreux paramètres inhérents aux agents pathogènes comme leur capacité de dispersion ou leur efficacité sur les différentes variétés.

# 7 Annexes

## 7.1 Cas d'étude 1 : Introduction et établissement du généraliste

### 7.1.1 summary() des GLM

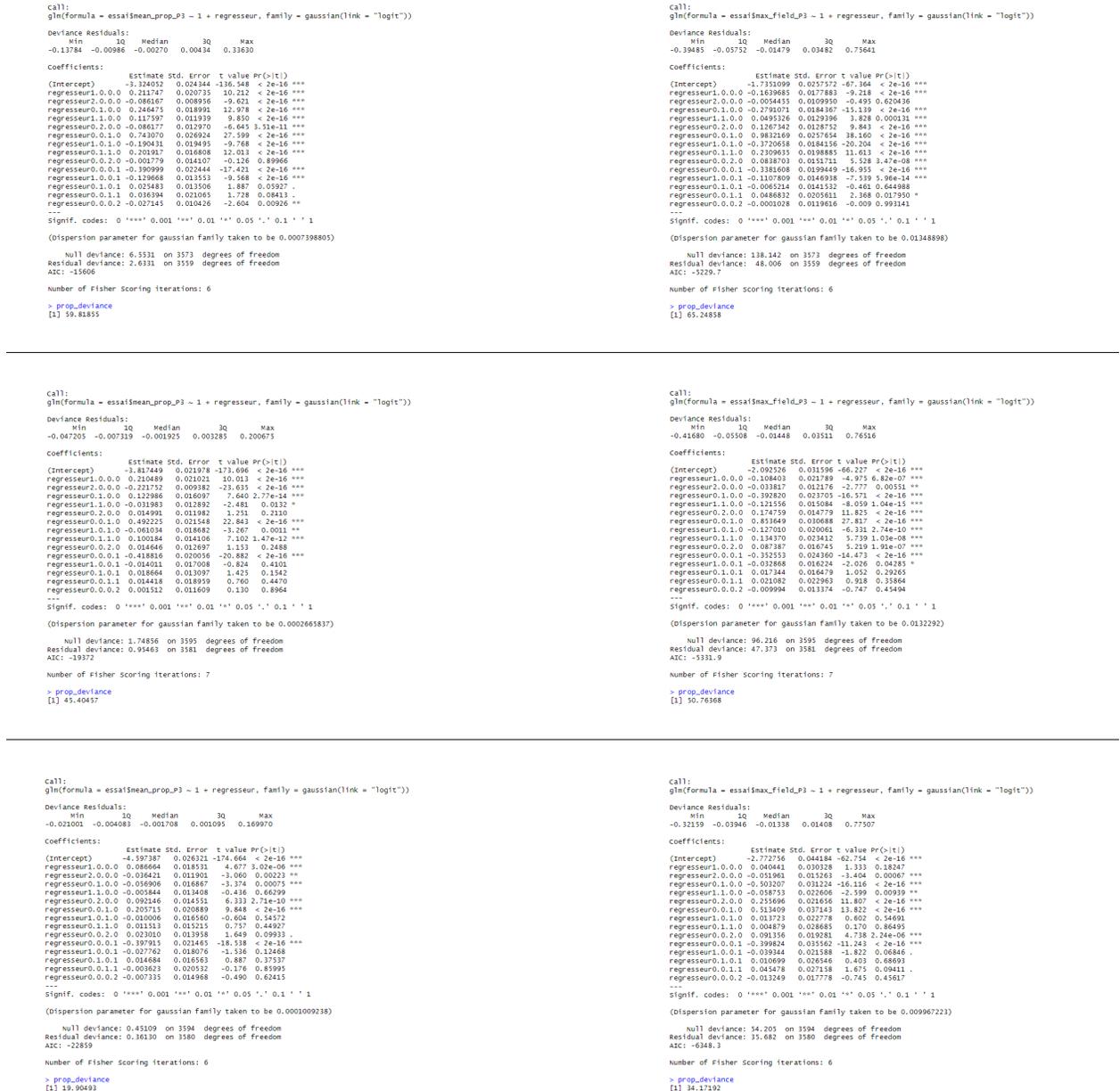


FIGURE 23 – summary() pour les modèles avec la fréquence moyenne de P3 mean\_prop\_P3 (gauche) et fréquence maximale max\_field\_P3 (droite) pour les 3 agrégations (de faible à forte de haut en bas) (target\_trait=IR, mutation\_rate=1.10<sup>-3</sup>). Les coefficients du modèles sont notés regresseur A.B.C.D où A=prop,B=disp,C=R\_efficiency,D= fitness\_cost et 1 et 2 désignent respectivement les termes de degré 1 ou 2. prop\_deviance est la déviance expliquée par le modèle et est calculée avec le ratio :  $\frac{\text{déviance nulle} - \text{déviance résiduelle}}{\text{déviance nulle}}$

### 7.1.2 Boxplot de la variance de la fréquence de P3

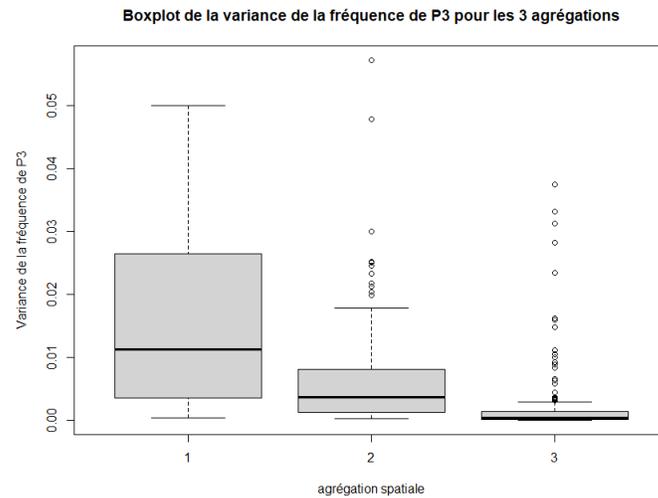


FIGURE 24 – *Boxplot de la variance de la fréquence de P3 au niveau du champ pour les 3 niveaux d'agrégation spatiale (1 : faible, 2 : moyenne, 3 : forte). L'efficacité est au maximum et le coût en fitness au minimum. Une variance élevée signifie des écarts importants dans les fréquences et donc l'établissement de populations locales de P3. La variance diminue avec l'agrégation spatiale, signifiant que l'établissement local est favorisé par une agrégation faible.*

### 7.1.3 Exemples de simulations

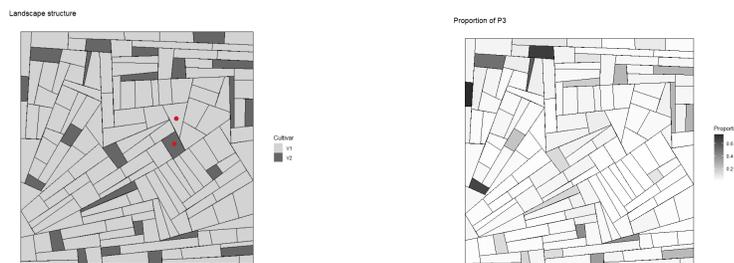


FIGURE 25 – *Exemple de simulation (illustration du point A dans la Figure 9) avec landsepi pour l'établissement du généraliste (structure du paysage (gauche) et fréquence de P3 (droite)) (`aggreg=1`, `prop=10%`, `disp=0.025`, `max(R_efficiency)`, `min(fitness_cost)`). Sur chaque point rouge sont inoculés P1 et P2. La fréquence de P3 est globalement faible sur le paysage mais peut atteindre une fréquence assez haute (jusqu'à 75%) sur quelques champs. Cela illustre bien les observations du graphique de contours avec l'établissement local du généraliste.*



FIGURE 26 – Exemple de simulation (illustration du point B dans la Figure 9) avec landsepi pour l'établissement du généraliste (structure du paysage (gauche) et fréquence de P3 (droite)) ( $\text{aggreg}=1$ ,  $\text{prop}=50\%$ ,  $\text{disp}=0.25$ ,  $\text{max}(\text{R\_efficiency})$ ,  $\text{min}(\text{fitness\_cost})$ ). Sur chaque point rouge sont inoculés P1 et P2. P3 est réparti de façon plutôt homogène dans le paysage : la moyenne est donc assez élevée, comme illustré sur le graphique de contours.

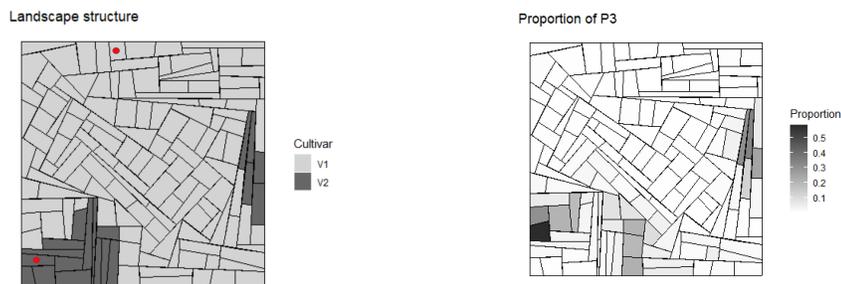


FIGURE 27 – Exemple de simulation (illustration du point C dans la Figure 9) avec landsepi pour l'établissement du généraliste (structure du paysage (gauche) et fréquence de P3 (droite)) ( $\text{aggreg}=3$ ,  $\text{prop}=10\%$ ,  $\text{disp}=0.025$ ,  $\text{max}(\text{R\_efficiency})$ ,  $\text{min}(\text{fitness\_cost})$ ). Sur chaque point rouge sont inoculés P1 et P2. La fréquence de P3 est négligeable sur la quasi-totalité du paysage excepté sur quelques champs où la fréquence peut être importante et où le généraliste construit une population locale, comme on pouvait le voir sur le graphique de contours. L'établissement local de P3 sur l'îlot en bas à gauche est localisé sur des champs V1, à la frontière des deux territoires. Sur l'îlot à droite cependant, P3 occupe des champs V2.

### 7.1.4 Diagrammes en barres

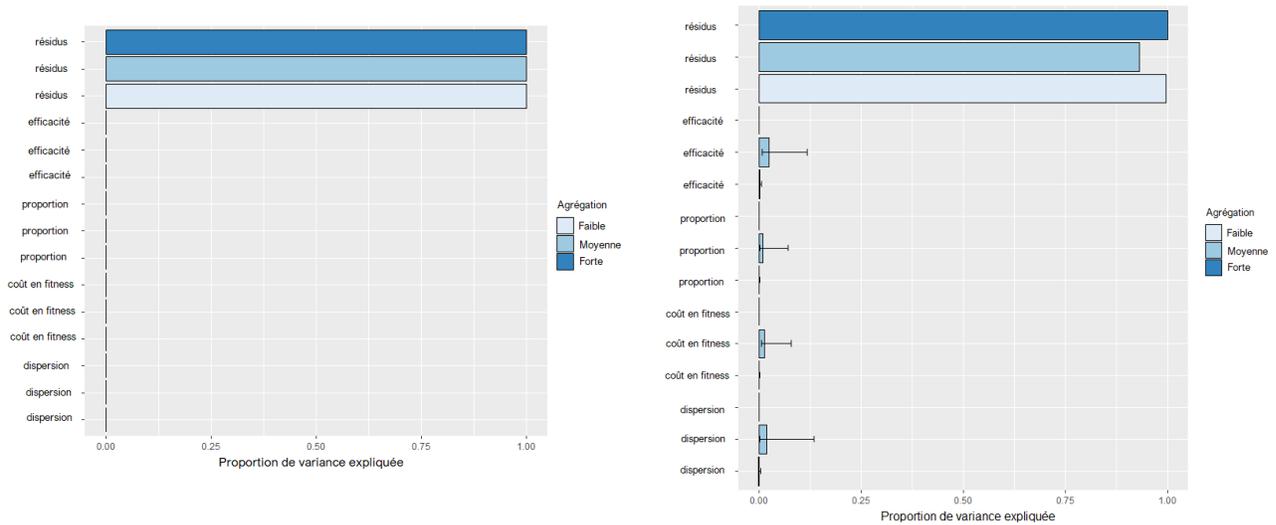


FIGURE 28 – Diagramme en barres des contributions de chaque variable à la variance du GLM de la fréquence moyenne (gauche) et maximale de P3 par champ (droite), pour les 3 niveaux d’agrégation spatiale (trait=IR et  $\text{taux de mutation}=1.10^{-5}$ )

### 7.1.5 Graphiques de contours

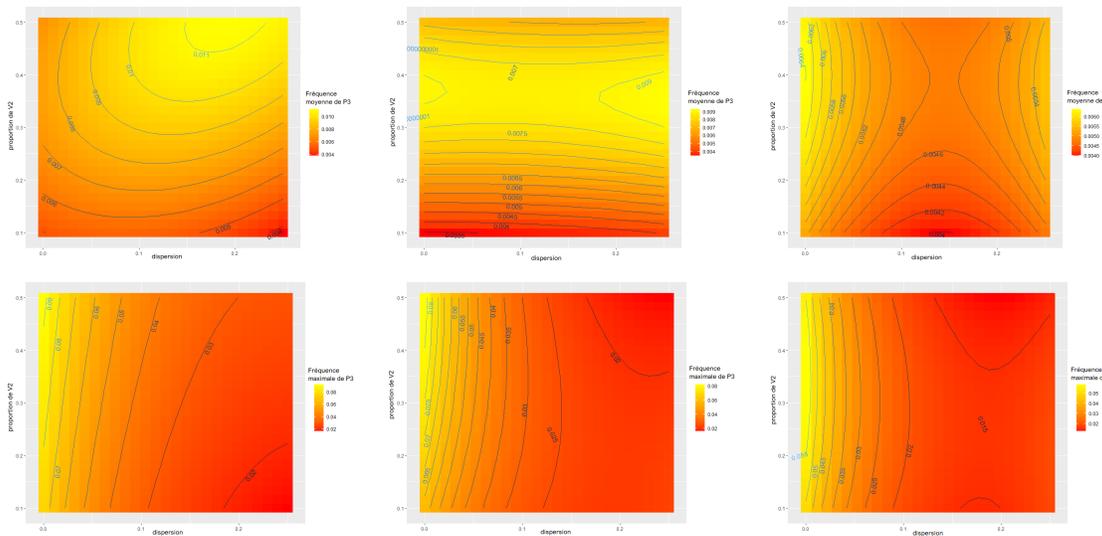


FIGURE 29 – Graphique de contours des prédictions de la fréquence moyenne (en haut) et maximale (en bas) de P3 à partir de l’estimation des paramètres du modèle avec l’intervalle de départ pour la dispersion et la proportion de V2, l’efficacité au **minimum** et le coût en fitness au **maximum** pour les 3 niveaux d’agrégation spatiale (de gauche à droite : faible, moyenne, forte) (trait=IR et  $\text{taux de mutation}=1.10^{-3}$ ).

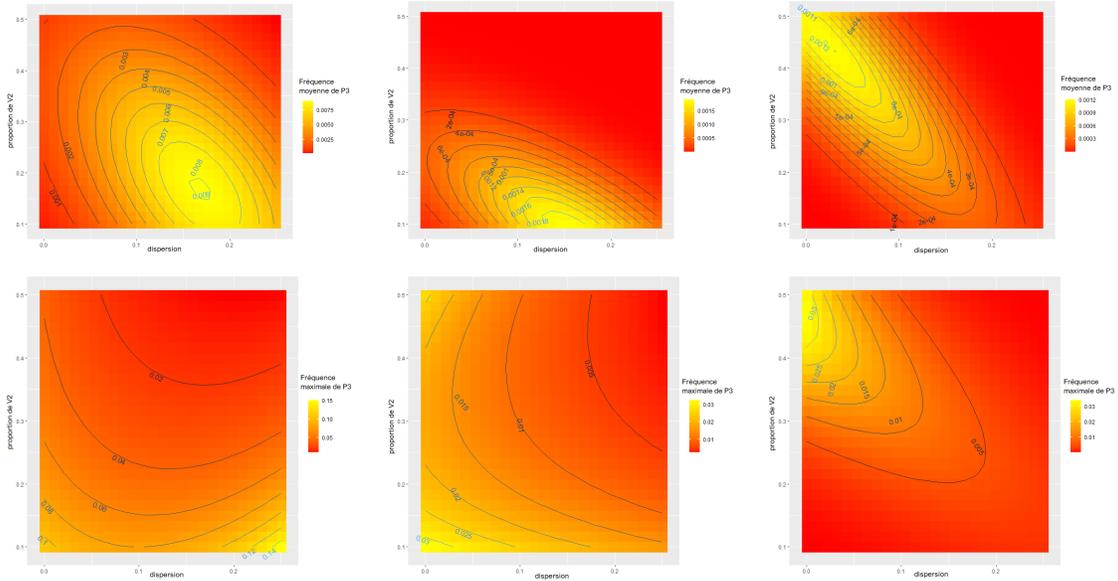


FIGURE 30 – Graphique de contours des prédictions de la fréquence moyenne (en haut) et maximale (en bas) de P3 à partir de l'estimation des paramètres du modèle avec l'intervalle de départ pour la dispersion et la proportion de V2, l'efficacité au maximum et le coût en fitness au minimum pour les 3 niveaux d'agrégation spatiale (de gauche à droite : faible, moyenne, forte) (trait=IR et *taux de mutation* =  $1.10^{-5}$ ).

### 7.1.6 Boxplots comparant les traits d'histoire de vie

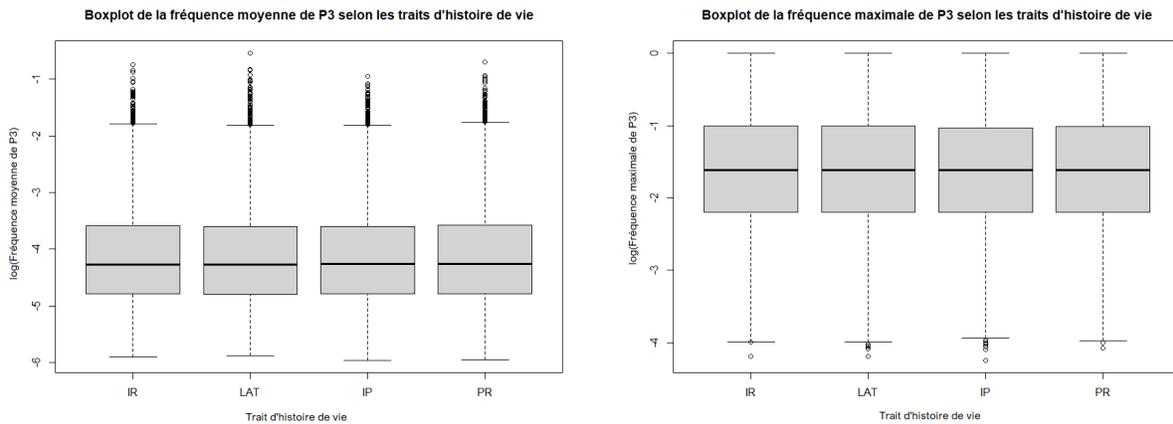


FIGURE 31 – Boxplot de la fréquence moyenne (gauche) et maximale (droite) de P3 (échelle log) selon les différents traits d'histoire de vie (*taux de mutation* =  $1.10^{-3}$ ). Les fréquences sont quasi identiques entre les traits.

### 7.1.7 summary() du modèle mixte

```

Linear mixed-effects model fit by REML
Data: final_tab
      AIC      BIC    loglik
50976.66 51067.18 -25477.33

Random effects:
Formula: ~1 | s1mu1
          (Intercept) Residual
Stddev:   0.4852553  0.5973836

Fixed effects: log(loc_freq/(1 - loc_freq)) ~ -1 + aggreg + aggreg:prop + aggreg:loc_aggreg
              Value Std.Error DF   t-value p-value
aggreg1     -0.044738  0.1339091  174   -0.33409  0.7387
aggreg2     -1.591131  0.1338979  174  -11.88316  0.0000
aggreg3     -2.063567  0.1353876  174  -15.24192  0.0000
aggreg1:prop -1.466816  0.3860521  174   -3.79953  0.0002
aggreg2:prop  0.154792  0.3857991  174    0.40122  0.6887
aggreg3:prop -0.514710  0.3855761  174   -1.33491  0.1836
aggreg1:loc_aggreg -3.613157  0.0306166 27538 -118.01295  0.0000
aggreg2:loc_aggreg -2.909312  0.0287689 27538 -101.12711  0.0000
aggreg3:loc_aggreg -2.584483  0.0341317 27538  -75.72080  0.0000

Correlation:
          aggreg1 aggreg2 aggreg3 aggreg1:prop aggreg2:prop aggreg3:prop aggreg1:loc_aggreg aggreg2:loc_aggreg aggreg3:loc_aggreg
aggreg2    0.000
aggreg3    0.000 0.000
aggreg1:prop -0.872 0.000 0.000
aggreg2:prop  0.000 -0.870 0.000 0.000
aggreg3:prop  0.000 0.000 -0.858 0.000 0.000
aggreg1:loc_aggreg -0.182 0.000 0.000 0.053 0.000 0.000
aggreg2:loc_aggreg 0.000 -0.181 0.000 0.000 0.039 0.000 0.000
aggreg3:loc_aggreg 0.000 0.000 -0.233 0.000 0.000 0.018 0.000 0.000

Standardized Within-Group Residuals:
              Min          Q1          Med          Q3          Max
-4.48099199 -0.64151277 -0.04329061  0.59178972  8.57382497

Number of Observations: 27720
Number of Groups: 180

```

FIGURE 32 – `summary()` du modèle mixte établi avec les résultats des simulations du cas d'étude 1. Le modèle a comme variable réponse la fréquence locale au niveau du champ `loc_freq` (lien logit). Les variables explicatives sont le niveau d'agrégation spatiale `aggreg(1 : faible, 2 : moyenne, 3 : forte)`, la proportion de V2 `prop` et l'agrégation locale `loc_aggreg`, qui sont des effets fixes. La simulation est considérée comme un effet aléatoire.

## 7.2 Cas d'étude 2 : Coexistences entre pathotypes

### 7.2.1 summary() du modèle de régression multinomiale

```

Call:
multinom(formula = COEX ~ aggreg + aggreg:(disp + prop + R_efficiency + fitness_cost) + aggreg:(disp:prop + disp:R_efficiency + disp:fitness_cost + prop:R_efficiency + prop:fitness_cost + R_efficiency:fitness_cost),
data = final_tab, maxit = 1000)

Coefficients:
(Intercept)  aggreg2  aggreg3  aggreg1:disp  aggreg2:disp  aggreg3:disp  aggreg1:prop  aggreg2:prop  aggreg3:prop  aggreg1:R_efficiency  aggreg2:R_efficiency  aggreg3:R_efficiency  aggreg1:fitness_cost  aggreg2:fitness_cost
2    80.492675 -50.95591 -73.675925 -254.76630  46.84773 -140.50873 -382.21235 -72.629437  4.380538 -132.799055 -172.175271 -39.563206  404.017547  256.80811
3    9.376186 -13.55587 -5.444708  75.56841  36.09164 -43.24671 -19.88045  8.682215  3.945679  2.709942  11.046011  6.471258 -9.164987  32.83893
4   -62.857635  53.83225  58.297518  11.71080  18.23466 -92.73178  30.07497  16.031840  17.911341  16.442937  7.332215  10.807449  750.886018  102.31241
5   -38.393507  33.47756  41.362752 -521.49188 -39.27898 -79.57277  54.97130  18.074723  19.296148  6.006195  7.773014  8.010669  653.739397  89.41508

aggreg3:fitness_cost  aggreg1:disp:prop  aggreg2:disp:prop  aggreg3:disp:prop  aggreg1:disp:R_efficiency  aggreg2:disp:R_efficiency  aggreg3:disp:R_efficiency  aggreg1:disp:fitness_cost  aggreg2:disp:fitness_cost
2    181.6763 -301.79722 -47.60123 -49.79713  355.5273 -45.36392 -86.13151  1487.01795  65.30978
3    207.2158 -889.51438 -233.94584 -35.75162 -139.4870 -11.89776 -46.74859  1032.22676 -81.29271
4    355.9693 -878.11092 -372.86161 -43.73106 -1071.0571 -204.68140 -71.78232 -28.20748  413.36176
5    240.7106  55.22336 -165.25556 -34.56733 -114.8288 -34.83866 -68.80648 -104.41808  121.14184

aggreg3:disp:fitness_cost  aggreg1:prop:R_efficiency  aggreg2:prop:R_efficiency  aggreg3:prop:R_efficiency  aggreg1:prop:fitness_cost  aggreg2:prop:fitness_cost  aggreg3:prop:fitness_cost  aggreg1:R_efficiency:fitness_cost
2    9.266347  351.5128673 -14.47946  18.68024 -205.07829  29.50651  58.05776 -648.26927
3   -66.711785  0.9692087 -33.56694 -26.90749  310.55411  359.46000  93.30326 -67.37922
4   149.471968 -60.5141955 -48.47780 -39.46554 -24.71994  390.56263  157.78210 -101.23678
5   147.154670 -13.5134682 -29.58816 -35.37337 -489.35253  252.18468  132.59631 -131.62258

aggreg2:R_efficiency:fitness_cost  aggreg3:R_efficiency:fitness_cost
2    118.05183  127.9283
3   -61.85494  226.9491
4   -29.08912  154.8146
5   -55.18949  225.9205

```

FIGURE 33 – `summary()` de la régression multinomiale établi avec les résultats des simulations du cas d'étude 2. Le modèle a pour variable réponse `COEX`, une variable catégorielle à 5 niveaux correspondant aux 5 sorties possibles des simulations (voir Figure 7). Les variables explicatives sont le niveau d'agrégation spatiale `aggreg(1 : faible, 2 : moyenne, 3 : forte)`, la dispersion `disp`, la proportion de V2 `prop`, l'efficacité `R_efficiency` et le coût en fitness `fitness_cost`.

## 7.2.2 Coexistences

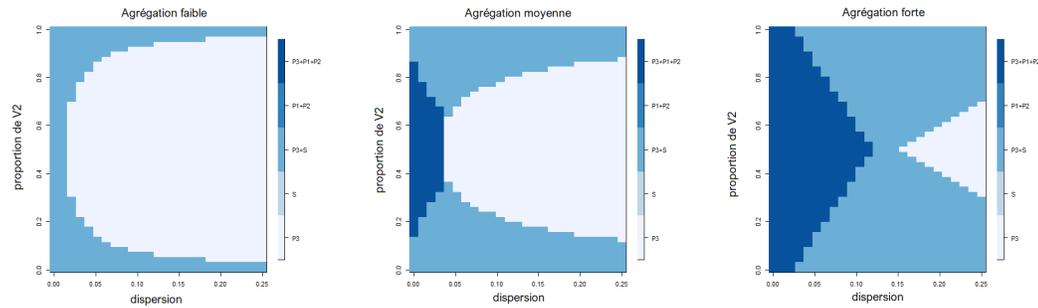


FIGURE 34 – Coexistences entre pathotypes obtenues par prédiction avec les estimations des paramètres du modèle de régression multinomiale selon les 3 niveaux d'agrégation spatiale et les intervalles de départ pour la dispersion et la proportion de V2. L'efficacité est au **maximum** et le coût en fitness au **minimum**. La variable catégorielle de classe (variable réponse du modèle) a été construite avec un seuil de 5%, signifiant qu'un pathotype est considéré présent seulement si sa fréquence moyenne dépasse ce seuil.

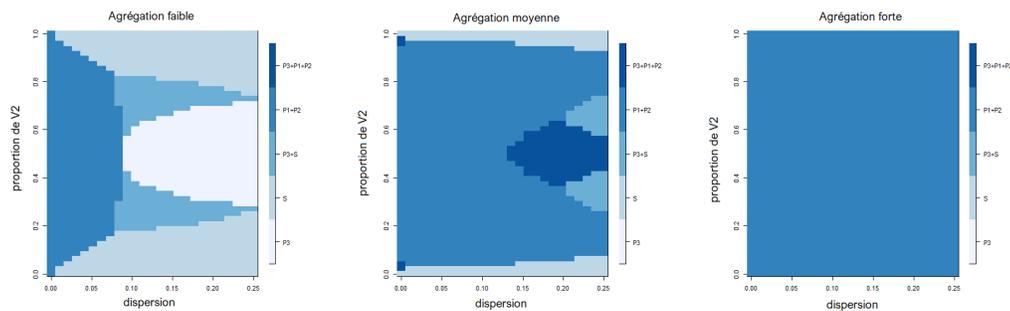


FIGURE 35 – Coexistences entre pathotypes obtenues par prédiction avec les estimations des paramètres du modèle de régression multinomiale selon les 3 niveaux d'agrégation spatiale et les intervalles de départ pour la dispersion et la proportion de V2. L'efficacité est au **minimum** et le coût en fitness au **maximum**. La variable catégorielle de classe (variable réponse du modèle) a été construite avec un seuil de 5%, signifiant qu'un pathotype est considéré présent seulement si sa fréquence moyenne dépasse ce seuil.

### 7.2.3 Exemples de simulations

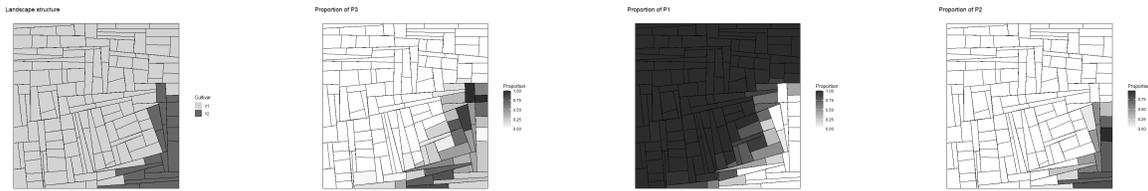


FIGURE 36 – Exemple de simulation avec landsepi pour le cas d'étude 2 (structure du paysage (gauche) et fréquences des pathotypes P3, P1 et P2) résultant en la coexistence des 3 pathotypes ( $\text{aggred}=3$ ,  $\text{prop}=10\%$ ,  $\text{disp}=0.025$ ,  $\text{max}(\text{R\_efficiency})$ ,  $\text{min}(\text{fitness\_cost})$ )

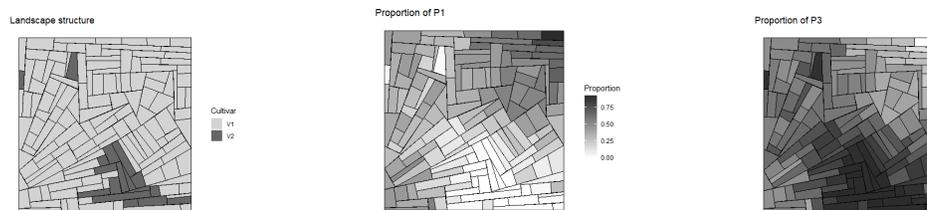


FIGURE 37 – Exemple de simulation avec landsepi pour le cas d'étude 2 (structure du paysage (gauche) et fréquences des pathotypes P1 et P3) résultant en la coexistence du généraliste et un des spécialistes ( $\text{aggred}=2$ ,  $\text{prop}=10\%$ ,  $\text{disp}=0.25$ ,  $\text{max}(\text{R\_efficiency})$ ,  $\text{min}(\text{fitness\_cost})$ )

## 7.3 Section modifiée de la vignette 1 de landsepi

### Setting the inoculum

To set the inoculum, the function `setInoculum()` is used. It can take many inputs depending on how one wants to parametrize the beginning of the simulation. Before setting the inoculum, it can be useful to use the functions `getMatrixGenePatho()`, `getMatrixCultivarPatho()` and `getMatrixCroptypePatho()` to acknowledge which pathotypes can infect which genes, cultivars and croptypes respectively. Each function returns matrices showing possibility of infection from the pathotype (value of 1) or protection from infection (value of 0).

```
getMatrixGenePatho(simul_params)
#>           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
#> MG 1           0  0  0  0  0  0  0  0  0  0  0
#> Lr34           1  1  1  1  1  1  1  1  1  1  1
#> gene 3         1  1  1  1  1  1  1  1  1  1  1
#> nonhost resistance 0  0  0  0  0  0  0  0  0  0  0
#>           [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21]
#> MG 1           0  1  1  1  1  1  1  1  1  1  1
#> Lr34           1  1  1  1  1  1  1  1  1  1  1
#> gene 3         1  1  1  1  1  1  1  1  1  1  1
#> nonhost resistance 0  0  0  0  0  0  0  0  0  0  0
#>           [,22] [,23] [,24]
#> MG 1           1  1  1
#> Lr34           1  1  1
#> gene 3         1  1  1
#> nonhost resistance 0  0  0

getMatrixCultivarPatho(simul_params)
#>           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
#> Susceptible  1  1  1  1  1  1  1  1  1  1  1  1
#> Resistant1   0  0  0  0  0  0  0  0  0  0  0  0
#> Resistant2   1  1  1  1  1  1  1  1  1  1  1  1
#> Resistant3   0  0  0  0  0  0  0  0  0  0  0  0
#> Forest       0  0  0  0  0  0  0  0  0  0  0  0
#>           [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23]
#> Susceptible  1  1  1  1  1  1  1  1  1  1  1  1
#> Resistant1   1  1  1  1  1  1  1  1  1  1  1  1
#> Resistant2   1  1  1  1  1  1  1  1  1  1  1  1
#> Resistant3   0  0  0  0  0  0  0  0  0  0  0  0
#> Forest       0  0  0  0  0  0  0  0  0  0  0  0
#>           [,24]
#> Susceptible  1
#> Resistant1   1
#> Resistant2   1
#> Resistant3   0
#> Forest       0

getMatrixCroptypePatho(simul_params)
#>           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
#> Susceptible crop  1  1  1  1  1  1  1  1  1  1  1
#> Pure resistant crop 0  0  0  0  0  0  0  0  0  0  0
#> Mixture           1  1  1  1  1  1  1  1  1  1  1
#> Other             0  0  0  0  0  0  0  0  0  0  0
#>           [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21]
#> Susceptible crop  1  1  1  1  1  1  1  1  1  1  1
#> Pure resistant crop 0  1  1  1  1  1  1  1  1  1  1
#> Mixture           1  1  1  1  1  1  1  1  1  1  1
#> Other             0  0  0  0  0  0  0  0  0  0  0
#>           [,22] [,23] [,24]
#> Susceptible crop  1  1  1
#> Pure resistant crop 1  1  1
#> Mixture           1  1  1
#> Other             0  0  0
```

The different possibilities to set the inoculum can be separated in two categories : global or local inoculum.

### Global inoculum

If only susceptible cultivars are to be infected at  $t=0$  by the avirulent pathotype (no adaptation to resistance), only the input `pI0_patho` can be entered, to set the probability for an individual to be infectious (i.e. state I) at the beginning of the simulation. The object `simul_params` is updated accordingly.

```
simul_params <- setInoculum(simul_params, pI0_patho = 5e-4)
```

If a more custom inoculum is wanted, it's possible to set the pathotypes (pathotype indexes) that are already present at  $t=0$  (and their respective probabilities to infect cultivars) as well as the cultivars (cultivar names) that are to be infected at  $t=0$ , on the whole landscape. Note that the inoculum set will be adapted so the pathotypes will infect only the cultivars they are able to infect.

```
present_patho<-c(1,2,5,8,23)
pI0_patho<-c(5e-4,4e-4,3e-4,2e-4,1e-4)
infected_cultivar<-c("Susceptible","Resistant1")
simul_params <-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                           infected_cultivar=infected_cultivar)
```

It's also possible to have different inoculations on the cultivars by setting a logical value to `TRUE` and `pI0_patho` as a 2D array.

```
pI0_patho<- array(data = 0, dim = c(length(infected_cultivar),length(present_patho)))
pI0_patho[1,1]=0.04
pI0_patho[1,2]=0.03
pI0_patho[2,1]=0.02
pI0_patho[2,2]=0.01
pI0_patho
```

```
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 0.04 0.03  0   0   0
#> [2,] 0.02 0.01  0   0   0
```

```
simul_params <-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                           infected_cultivar=infected_cultivar,
                           diff_cultivar=TRUE)
```

### Local inoculum

It's also possible to have a more local inoculum by providing the fields that are to be infected and the logical value `local` to `TRUE`. In that case, the inoculum can be the same on all chosen fields or different for each field and the same applies to cultivars. These options are controlled through logical values (`diff_field` and `diff_cultivar`) and by the dimensions of `pI0_patho`.

#### Option 1. Same inoculum on the local fields and the present cultivars :

```
local_field=c(3,4)
pI0_patho<-c(5e-4,4e-4,3e-4,2e-4,1e-4)
simul_params<-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                          infected_cultivar=infected_cultivar,local=TRUE,local_field=local_field)
```

#### Option 2. Different inoculum on the present cultivars :

```
pI0_patho<- array(data = 0, dim = c(length(infected_cultivar),length(present_patho)))
pI0_patho #filled
```

```
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 0.04 0.03  0   0   0
#> [2,] 0.02 0.01  0   0   0
```

```
simul_params<-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                          infected_cultivar=infected_cultivar,local=TRUE,local_field=local_field,
                          diff_cultivar=TRUE)
```

#### Option 3. Different inoculum on the local fields :

```
pI0_patho<- array(data = 0, dim = c(length(local_field),length(present_patho)))
pI0_patho #filled
```

```
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 0.04 0.03  0   0   0
#> [2,] 0.02 0.01  0   0   0
```

```
simul_params<-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                          infected_cultivar=infected_cultivar,local=TRUE,local_field=local_field,
                          diff_field=TRUE)
```

#### Option 4. Different inoculum on the present cultivars and on the local fields :

```
pI0_patho<- array(data = 0, dim = c(length(infected_cultivar),length(present_patho),length(local_field))
pI0_patho #filled
```

```
#> , , 1
#>
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 0.08 0.07 0.06 0.05 0.04
#> [2,] 0.00 0.00 0.00 0.06 0.05
#>
#> , , 2
#>
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] 0.04 0.03  0 0.00 0.00
#> [2,] 0.00 0.00  0 0.02 0.01
```

```
simul_params<-setInoculum(simul_params,present_patho=present_patho,pI0_patho=pI0_patho,
                          infected_cultivar=infected_cultivar,local=TRUE,local_field=local_field,
                          diff_cultivar=TRUE,diff_field=TRUE)
```

#### Custom inoculum using a 3D array

At last, if one wants to run a simulation with a custom inoculum on the whole landscape, the input `pI0_mat` is used with a 3-dimensional array of dimensions `Ncultivars.Npathotypes.Nfields` (matrices with cultivars in rows and pathotypes in columns). It is possible to check the validity of the inoculum with `checkInoculumLandscape()` that will either return `TRUE`, either a corrected version of the inoculum.

```
#pI0_vector is of the form [1:5,1:24,1:155] (5 cultivars, 24 pathotypes, 155 fields)
simul_params <- setInoculum(simul_params, pI0_mat = pI0_vector)
checkInoculumLandscape(simul_params)
```

## Références

- Bonnefoy, 2012. *Pesticides : vers le risque zéro* [en ligne]. [visité le 2023-04-24]. Disp. à l'adr. : <https://www.senat.fr/rap/r12-042-1/r12-042-1.html>.
- Brun, H. et al., 2009. Quantitative Resistance Increases the Durability of Qualitative Resistance to *Leptosphaeria maculans* in *Brassica napus*. *The New phytologist*. T. 185, p. 285-99.
- Chesson, P., 2000. Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*. T. 31, p. 343-366.
- Débarre, F. ; Lenormand, T., 2011. Distance-limited dispersal promotes coexistence at habitat boundaries : reconsidering the competitive exclusion principle. *Ecology Letters*. T. 14, p. 260-266.
- Ehrlich, P. R.; Raven, P. H., 1964. Butterflies and Plants: A Study in Coevolution. *Evolution*. Vol. 18, p. 586-608.
- Flor, H. H., 1956. The Complementary Genic Systems in Flax and Flax Rust. *Advances in Genetics*. Vol. 8, p. 29-54.
- Karlin, S. ; McGregor, J., 1972. Polymorphisms for genetic and ecological systems with weak coupling. *Theoretical Population Biology*. T. 3, p. 210-238.
- Keesing, F. et al., 2006. Effects of species diversity on disease risk. *Ecology Letters*. T. 9, p. 485-498.
- Lannou, C., 2012. Variation and selection of quantitative traits in plant pathogens. *Annual Review of Phytopathology*. T. 50, p. 319-338.
- Lannou, C. et al., 2021. *L'immunité des plantes - Pour des cultures résistantes aux maladies*. Éditions Quæ.
- Leroux, P.; Walker, A.-S., 2011. Multiple mechanisms account for resistance to sterol 14-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. *Pest Management Science*. Jg. 67, p. 44-59.

- Melbourne, B. A. et al., 2007. Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? *Ecology Letters*. Vol. 10, p. 77–94.
- Mundt, C. C. et al., 2009. Long-distance dispersal and accelerating waves of disease : empirical relationships. *The American Naturalist*. T. 173, p. 456-466.
- Omrane, S. et al., 2017. Plasticity of the MFS1 Promoter Leads to Multidrug Resistance in the Wheat Pathogen *Zymoseptoria tritici*. *mSphere*. T. 21, p. 157-197.
- Papaïx, J. et al., 2011. Influence of cultivated landscape composition on variety resistance : an assessment based on wheat leaf rust epidemics. *New Phytologist*. T. 191, p. 1095-1107.
- Pilet-Nayel, M.-L. et al., 2017. Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. *Frontiers in Plant Science*. T. 8.
- Rimbaud, L. et al., 2018. Assessing the durability and efficiency of landscape-based strategies to deploy plant resistance to pathogens. *PLOS Computational Biology*. Vol. 14, e1006067.
- Rimbaud, L. et al., 2021. Models of Plant Resistance Deployment. *Annual Review of Phytopathology*. T. 59, p. 125-152.
- Saltelli, A. et al., 2000. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*. T. 15, p. 377-395.
- Strobeck, C., 1979. Haploid Selection with n Alleles in m Niches. *The American Naturalist*. T. 113, p. 439-444.
- Stukenbrock, E. H. et al., 2007. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*. T. 24, p. 398-411.
- Thrall, P. et al., 2003. Influence of spatial structure on pathogen colonization and extinction : a test using an experimental metapopulation. *Plant Pathology*. T. 52, p. 350-361.
- Wei, W. ; Krone, S. M., 2005. Spatial invasion by a mutant pathogen. *Journal of theoretical biology*. T. 236, p. 335-348.
- With, K. A., 2002. The Landscape Ecology of Invasive Spread. *Conservation Biology*. Vol. 16, p. 1192–1203.

---

Yorke, J. A. et al., 1978. Dynamics and Control of the Transmission of Gonorrhoea. *Sexually Transmitted Diseases*. T. 5, p. 51-56.