THÈSE

Pour obtenir le grade de

Docteur

Délivré par Montpellier SupAgro

Préparée au sein de l'école doctorale GAIA
Et de l'unité de recherche BGPI

Spécialité : Biologie des Interactions

Présentée par Coralie Picard

Modélisation et optimisation de la gestion d'une épidémie : quel impact du paysage ?

Soutenance le 16/10/2018 devant le jury composé de

Elisabeta Vergu, DR, INRA - UMR MaIAGE	Rapporteur
Ivan Sache, PR, AgroParisTech – UMR BIOGER	Rapporteur
Vianney Houlès, Chercheur, ITK	Examinateur
François Brun, Ingénieur, ACTA	Examinateur
Claire Neema, PR, Montpellier SupAgro – UMR BGPI	Présidente du jury
Gaël Thébaud, CR, INRA – UMR BGPI	Directeur de thèse
Samuel Soubeyrand, DR, INRA – Unité BioSP	Co–encadrant
Pascal Hendrikx, Chercheur, ANSES	Co–encadrant

Remerciements

Cette thèse a été une expérience inestimable aussi bien au niveau professionnel qu'au niveau personnel. J'ai donc énormément de personnes à remercier, je tenterai d'être brève sans oublier trop de monde.

Je commence tout naturellement par remercier **Gaël Thébaud**, qui m'a encadrée tout au long de ces 3 années. Je te remercie d'avoir été patient, d'avoir su m'orienter, de m'avoir transmis tes connaissances scientifiques, de m'avoir conseillée et de m'avoir poussée à donner le meilleur de moi-même. Je sors grandie de cette expérience.

Je remercie également **Samuel Soubeyrand** qui m'a accueillie dans l'unité BioSP à Avignon. Je te remercie particulièrement de ta patience et ta pédagogie ; tu as su m'expliquer de nombreuses formules qui n'avaient aucun sens pour moi.

Un grand merci à **Pascal Hendrikx** et **Pauline de Jerphanion** pour leur gentillesse et leurs précieux conseils. Les discussions que nous avons pu avoir m'ont apportées un autre regard sur la recherche.

Merci à **Emmanuel Jacquot** de m'avoir acceptée en thèse dans son équipe.

Je remercie **Victor Picheny** pour l'aide précieuse qu'il m'a apportée sur les techniques d'optimisation qui ont été utilisées dans cette thèse. Je te remercie en particulier pour l'accueil que tu m'as fait dans ton unité de recherche à Toulouse, et pour tout le temps que tu m'as accordé.

Je tiens également à remercier **Elisabeta Vergu**, **Ivan Sache**, **Vianney Houlès**, **François Brun** et **Claire Neema** d'avoir accepté d'évaluer mon travail en participant à mon jury de thèse, ainsi que **Frédéric Fabre**, **Pauline Ezanno** et **Samuel Alizon** pour les précieux conseils qu'ils ont pu me fournir lors des mes comités de thèse.

Un grand merci à **Claire Neema** d'avoir écouté et soutenu l'ensemble des thésards en tant que directrice de l'unité BGPI, d'avoir fait partie de mon comité de thèse ainsi que de mon jury de thèse. Tu es pour moi un exemple.

Merci à l'équipe Epi2V de m'avoir accueillie. Je pense en particulier à **Nicolas**, **Sonia**, **Anaïs** et **Josiane** pour leur soutien. Merci également à **François**. Tu m'as supportée et écoutée même si tu n'étais pas toujours d'accord avec moi, tu es pour moi le grand sage de l'équipe. Merci à **Sylvie** pour tous ses conseils, tant au niveau professionnel que personnel. Et pour finir, un grand merci à **Marlène** et

Véronique pour leur bonne humeur et leur joie de vivre qui m'ont encouragée (sans vous, les réunions d'équipes auraient été beaucoup moins drôles).

J'aimerais également remercier toutes les personnes qui participent à la bonne ambiance de BGPI tous les jours. Les animations du comité des fêtes et les discussions animées sur la passerelle vont me manquer. Je remercie en particulier **Henri**, **Romain**, **Charlotte**, **Katia, Romain, Stella**. Je n'oublierai pas non plus le temps passé dans le bureau d'**Agnès** et **Michel** à écouter les ragots de l'unité.

Au cours de cette thèse, j'ai effectué plusieurs séjours au laboratoire BioSP à Avignon. Ces périodes m'ont à chaque fois permis de me ressourcer et de prendre du recul sur mon travail de thèse. Pour cela, je remercie **Julien**, **Emilie**, **Lucie**, **Souphiane**, **Candy**, **Mariam**, **Marie-Eve**, **Amel**, et bien sûr **Mariem** qui m'a gentiment hébergée.

Un grand merci à tous les copains de thèse, sans qui les journées à BGPI auraient été beaucoup moins animées ! Merci à **Emilie** (merci pour tes 1000 conseils sur « comment survivre en thèse »), **Loup, Diana** (merci de m'avoir fait partager ta merveilleuse vision de la vie), **Béa**, **Matthias**, **Thomas**, **Jérémy** (j'espère que toutes nos conversations de filles n'étaient pas trop fatigantes !), **Kévin**, **Émeline**, **Maud**, **Faustine**, **Raph**, **Pascal**, **Albin** (oui oui, je rentre dans ton jeu, je fais croire que tu travailles à BGPI) et **Jérémy** (j'aurai toujours une petite pensée pour toi lorsque je porterai ma cotte de maille). Merci pour tous ces moments passés sur la passerelle, à la plage, dans les bars, autour d'un barbecue… Je finirai par remercier **Edwige** et **Maëlle**, vous avez été les 2 piliers de ma thèse. Je n'oublierai pas les crises de fous rires et de larmes qui ont eu lieu dans notre bureau. Ces 3 années de thèse n'auraient pas été les mêmes sans vous (Racoon power !). Maëlle, j'espère que nous aurons l'occasion de finir la tournée des bars à vin. Edwige, j'espère que ma prochaine voisine de bureau sera aussi bavarde que toi sinon je vais m'ennuyer ! Merci à toutes les 2 pour toutes les citations qui ont orné le mur de mon bureau. Une citation m'aura particulièrement touchée :

« J'ai décidé d'être heureuse parce que c'est bon pour la santé »

Je remercie les **petites marinières** avec qui j'ai pu participer au concours MT180. Cette expérience fut riche en rencontres et m'a poussé à donner le meilleur de moi-même.

Je remercie également tous ceux qui ont croisé mon chemin lors de mes études et qui ont fait que j'en suis là aujourd'hui. Je pense tout d'abord aux copains APIMET et notamment à **Agathe**, **Margaux**, **Alizée** et **Benjamin**, ainsi qu'aux enseignants qui m'ont donné l'envie de continuer en thèse (je pense en particulier à **Isabel**, **Jacques David** et **Jean-Luc Regnard**). Je remercie également tous les copains de l'INH qui ont été ma famille pendant 5 ans : **Marie**, **Alized**, **Oriane**, **Pauline**, **Loick**,

Valoo, **Clément**, **Jonathan**, **Clouch**, **Alice**, **Adeline**, **Laurène**, **Nina, Pedro**, **Camille, Maëliss, Michel**. Je n'oublierai jamais ces années de travail acharné (ou plutôt d'ADF, de chouilles et de fous rires).

Mille mercis également aux copains de colo **Cha**, **Tintin**, **Fred**, **Elo**, **Cléa**, **Corentin**, **Babush**, **Léane** pour tous les moments que nous avons partagé.

Je ne peux terminer ces remerciements sans parler de mes 2 copines d'enfance, qui ne m'ont jamais quittée. **Marine** et **Flavie**, nous avons commencé à jouer ensemble à la maternelle, et nous ne nous sommes pas séparées jusqu'au bac. Je n'oublierai jamais les cours de philo durant lesquels nous avons refait le monde. Je suis fière de ce que nous sommes devenues.

Bien entendu, je finirai par remercier ma famille sans qui je ne serais jamais arrivée là : **Papa**, **Maman**, **Mélanie**, **Geoffrey**, **Charlotte**, **Papy**, **Mamie**, **Yoshi**, tous les **cousins/oncles/tantes**, et toute la famille **Altar**. Je ne vous le dis sûrement pas assez souvent mais je suis fière et heureuse de faire partie de notre famille. Vous ne vous en êtes peut-être pas rendu compte mais vous avez été d'un soutien infaillible. Chacun à sa façon, vous êtes mes exemples. Je vous dédicace ce mémoire pour que vous voyiez où m'ont mené ces longues études (après ma soutenance, peut-être que vous comprendrez enfin ce sur quoi je travaille).

Enfin, mon **wiwi** je te remercie pour ton soutien sans failles, tes encouragements et tout ton amour.

Sommaire

INTRODUCTION 9

1. Epidémiologie et gestion des maladies 14

1.1. Epidémiologie : définitions et historique 14
1.1.1. Quelques définitions : maladie et épidémie 14
1.1.2. Conséquences des épidémies 15
1.1.3. Définition et historique de l'épidémiologie 15
1.2. Stratégies de gestion des épidémies 16

2. Les modèles en épidémiologie 18

2.1. Qu'est-ce qu'un modèle ? 18
2.2. Quelques caractéristiques des modèles utilisés en épidémiologie 20

3. Les modèles pour optimiser de la gestion des épidémies 21

3.1. Qu'est-ce que l'optimisation ? 21
3.2. Optimisation de la gestion des maladies 22

4. La sharka comme modèle d'étude 23

4.1. La sharka 23
4.2. La gestion de la sharka en France 24
4.3. Optimisation des stratégies de gestion de la sharka 24
4.3.1. Modèle de simulation 24
4.3.2. Optimisation de la gestion de la sharka 26

5. Objectifs de ma these 26

Article 1: PESO: a modelling framework to help improve management strategies for epidemics – application to sharka 27

CHAPITRE 1 : ESTIMATION DE PARAMETRES EPIDEMIOLOGIQUES 37

1. Les modèles pour comprendre la dynamique des épidémies 38

Article 2: Exploiting genetic information to trace plant virus dispersal in landscapes 39

2. Application au virus de la sharka 64

Article 3: Accounting for uninfected hosts in transmission tree reconstruction 65

CHAPITRE 2 : ETUDE DE L'INFLUENCE DU PAYSAGE SUR LES STATEGIES DE GESTION DES MALADIES 85

Article 4: Analyzing the influence of landscape aggregation on disease spread to improve management strategies 89

CHAPITRE 3 : OPTIMISATION DE LA STRATEGIE DE GESTION D'UNE EPIDEMIE 123

1. Optimisation des paramètres de gestion d'une épidémie grâce à l'analyse de sensibilité 124

2. Optimisation des paramètres de gestion d'une épidémie grâce à un algorithme d'optimisation 127

Article 5: Impact of input warping on the Bayesian optimisation of the management of a plant disease using a complex epidemiological model 129

Article 6: *In silico* optimization of a strategy for landscape-wide plant disease management 151

3. Optimisation de la répartition de variétés résistantes dans un paysage 194

Article 7: Optimization of the spatio-temporal deployment of resistant cultivars and disease control options 195

CONCLUSION GENERALE ET PERSPECTIVES 217

1. Conclusion générale 218

2. Perspectives 219

2.1. Amélioration du processus PESO 219

2.1.1. Estimation des paramètres épidémiologiques 219

2.1.2. Simulation de l'épidémie 220

2.1.3. Optimisation de la gestion d'une épidémie 223

2.2. Quelle place pour les modèles épidémiologiques dans la gestion des maladies ? 226

ANNEXE 228

REFERENCES BIBLIOGRAPHIQUES 229

LISTE DES PUBLICATIONS, PRESENTATIONS ET POSTERS 238

ABSTRACT / RESUME 241

Liste des figures (hors articles) et annexe

Figure 1 : Fonctionnement d'un modèle qui traduit en équations le fonctionnement d'un système composé de plantes, du sol et de l'atmosphère, pour obtenir des valeurs de rendement.

Figure 2 : Stratégie de gestion contre le *Plum pox virus* dans les vergers français, définie par l'arrêté national de lutte du 17 mars 2011 (JORF, 2011).

Figure 3 : Méthodes d'optimisation des paramètres de gestion d'une épidémie au cours du temps (exemple avec un pas de temps de 5 ans).

Annexe 1 : Les organismes en charge de l'élaboration et de l'application des stratégies de gestion contre les épidémies.

Introduction

1. Epidémiologie et gestion des maladies

1.1. Epidémiologie : définitions et historique

1.1.1. Quelques définitions : maladie et épidémie

Une **maladie** correspond à l'altération de la santé et des fonctions des êtres vivants (en particulier la nutrition, la croissance et la reproduction). Les maladies peuvent être **transmissibles** ou non. Chez l'homme, les maladies cardio-vasculaires, la majorité des cancers, les maladies auto-immunes, la maladie d'Alzheimer ou encore l'asthme font par exemple partie des maladies non transmissibles entre différents individus. Chez la plante on parle de désordres physiologiques (ou parfois de maladies abiotiques). Ils désignent les perturbations du métabolisme, les retards de croissance ou les anomalies du développement résultant de causes non dépendantes d'êtres vivants. Ils sont souvent dus au climat ou à un défaut d'alimentation de la plante. Les maladies transmissibles d'un individu à un autre concernent essentiellement les anomalies génétiques héréditaires et les maladies infectieuses. On parle de **maladie infectieuse** lorsqu'elle est provoquée par un parasite. Ce parasite est un micro-organisme qui peut être par exemple un virus, une bactérie ou un champignon. Ma thèse s'intéresse seulement aux maladies d'origine infectieuse qui peuvent se propager à travers une population hôte.

La **transmission** d'une maladie infectieuse peut être directe ou indirecte. La transmission est directe lorsqu'elle se fait d'un sujet malade à un sujet sain par contact direct ou par l'air. Ainsi, chez l'homme, le sida se transmet lors de rapports sexuels ou par des échanges sanguins et la gale se transmet lors d'un contact cutané. Chez la plante, le mildiou et le botrytis qui affectent de nombreuses cultures se transmettent aussi par contact. La transmission est indirecte lorsque la maladie est transmise par l'intermédiaire de supports ou vecteurs. Par exemple, la dengue et le chikungunya sont transmis à l'homme par des moustiques, et certains virus de la mosaïque jaune et le virus de la sharka sont transmis à leurs plantes hôtes par des pucerons.

Une maladie peut entrainer une **épidémie** si sa propagation est rapide. Autrement dit, une épidémie désigne l'augmentation rapide de la **prévalence** et de l'**incidence** d'une maladie dans un population en un lieu donné, à un moment donné. Ces deux indicateurs sont les plus utilisés en épidémiologie pour évaluer la fréquence et la vitesse d'apparition d'une maladie. La **prévalence** mesure la part de la population atteinte par une maladie à un moment donné. L'**incidence** d'une maladie mesure l'état de santé d'une population en dénombrant le nombre de nouveau cas sur une période donnée.

De nombreuses épidémies affectent l'homme, les animaux et les cultures. Les plus connues ont entrainé de graves conséquences socio-économiques.

1.1.2. Conséquences des épidémies

Les **épidémies** sont responsables de nombreux décès dans le passé et continuent d'avoir de graves conséquences encore aujourd'hui. Ainsi, la peste noire, causée par la bactérie *Yersinia Pestis,* a tué entre 25 et 50 millions de personnes en Europe entre 1347 et 1352, soit 30 à 50% de la population (Perry et Fetherston, 1997). Quelques années plus tard, la grippe espagnole de 1918 due à la souche H1N1 aurait tué 30 millions de personnes selon l'Institut Pasteur. Egalement, depuis 1981, le sida a entrainé la mort de plus de 25 millions de personnes à travers le monde (UNAIDS, 2015).

Les épidémies peuvent être directement meurtrières chez l'homme, mais elles peuvent aussi avoir de graves conséquences sur l'alimentation et l'économie agricole (Rosegrant et Cline, 2003; Strange et Scott, 2005). Une des épidémies les plus connues, provoquée par le mildiou de la pomme de terre, a frappé l'Europe dans les années 1840 et a entrainé une famine qui a causé le décès de plus d'un million de personnes (Vanhaute et al. 2006). Les épidémies continuent encore aujourd'hui à provoquer des pertes de rendement considérables dans les cultures. Nous pouvons par exemple citer les cas de la rouille, un champignon qui infecte le blé (Galus, 2008), de l'oïdium, qui impacte de nombreuses espèces (tomate, vigne, abricotier etc.) ou encore de la sharka qui altère le rendement et la qualité des fruits chez les *Prunus* (cette maladie sera amplement abordée dans la suite de cette thèse).

Les conséquences socio-économiques des épidémies étant tragiques, une discipline ayant pour but d'étudier les facteurs influant sur la santé et les maladies des populations s'est développée au début du XX^{e} siècle : l'épidémiologie.

1.1.3. Définition et historique de l'épidémiologie

L'**épidémiologie** peut être définie comme l'étude de la **distribution** et des **déterminants** des états ou phénomènes liés à la santé dans une population déterminée et l'application de cette étude à la prévention et au contrôle des problèmes de santé (Last, 2001; MacMahon et al. 1960). Sa finalité est de rechercher des méthodes d'intervention efficaces grâce à une meilleure compréhension des maladies (Jenicek et Clèroux, 1982).

Hippocrate peut être considéré comme le père fondateur de l'épidémiologie, puisque qu'il fut l'un des premiers à chercher des déterminants environnementaux aux maladies (Merrill, 2012). Bien plus tard, les premiers statisticiens ont pris en compte l'environnement dans leurs études en se servant

de méthodes quantitatives pour étudier les maladies dans les populations humaines et éclairer les efforts de prévention et de lutte. Par exemple, John Snow s'est aperçu en 1855 que le risque de choléra à Londres était associé à la consommation de l'eau que distribuait la société Southwark and Vauxhall Water Company et qui était contaminée par les eaux usées (Snow et Richardson, 1936), et (Doll et Hill, 1964) ont étudié les rapports entre tabagisme et cancer du poumon et ont pu établir des relations entre les deux. Au début du XX^{e} siècle, des méthodes mathématiques ont été développées pour être appliquées en épidémiologie (Kermack et Mckendrick, 1927). Bernoulli est l'un des premiers à avoir contribué à l'épidémiologie mathématique en utilisant la modélisation pour évaluer une action de santé publique qui ne peut pas être observée directement : avec un modèle sur la variole, il calcule le gain en terme de vie moyenne si la variolation (inoculation volontaire de la variole) est appliquée.

Dans le domaine végétal, l'épidémiologie a commencé à se développer dans les années 1960 avec Van Der Plank : ce dernier a publié un ouvrage majeur qui offre un cadre théorique pour l'étude de l'épidémiologie des maladies des plantes (Van Der Plank, 1963). Cet ouvrage s'est inspiré de l'expérimentation sur différents systèmes hôtes-pathogènes et a fait rapidement progresser l'épidémiologie végétale, en particulier pour les pathogènes fongiques des feuilles. Depuis, de nombreux modèles mathématique ont été développés sur des pathosystèmes variés (Gilligan, 1985; Jeger, 1989; Kranz, 2012; McLean et al. 1986).

L'ensemble des travaux effectués en épidémiologie permettent de mieux comprendre les épidémies, ce qui va permettre d'élaborer des méthodes de luttes adaptées.

1.2. Stratégies de gestion des épidémies

Les maladies infectieuses résultent de multiples interactions entre un agent infectieux, un mode de transmission, une gamme d'hôtes et l'environnement. Pour lutter contre les maladies, il faut agir sur un ou plusieurs de ces éléments, dont les trois premiers subissent l'influence de l'environnement. Pour agir efficacement, il est indispensable de comprendre comment ces éléments fonctionnent et interagissent entre eux, de façon à mettre au point puis à appliquer et évaluer des mesures de lutte appropriées.

Les stratégies de gestion des maladies sont généralement réfléchies sur le long terme et prennent en compte les aspects économiques, environnementaux et sociétaux. La **gestion d'un organisme nuisible** est définie comme l'évaluation et la sélection des options permettant de réduire le risque

d'introduction et de dissémination d'un organisme nuisible (FAO, 2017). En fonction du contexte, ces options ne vont pas nécessairement être les mêmes. Elles peuvent avoir pour but d'éliminer totalement l'agent pathogène d'une zone (**éradication**, FAO, 2017), de réduire son impact en réduisant la taille de sa population (**suppression**) ou de prévenir sa dissémination dans ou autour d'une zone infestée (**enrayement**, FAO, 2017). La **lutte** contre les organismes nuisibles correspond à l'ensemble des méthodes employées pour arriver à l'éradication, la suppression ou l'enrayement d'une maladie (FAO, 2017). Lorsque la stratégie de lutte consiste à prévenir la propagation des maladies dans les zones non infestées (i.e. d'éviter que les organismes nuisibles ne s'échappent), on parle de **confinement** (FAO, 2017).

Les méthodes de lutte sont diverses et dépendent des caractéristiques de la maladie. Une des pratiques courantes est l'utilisation de produits phytosanitaires (Gubbins et Gilligan, 1999; Hall et al. 2004; Taylor et al. 2016). Par exemple, la maladie bactérienne des agrumes, ou maladie du dragon jaune, est gérée par la pulvérisation d'insecticides pour réduire l'abondance de ses vecteurs : les psylles asiatiques des agrumes (Grafton-Cardwell et al. 2013; Taylor et al. 2016). Néanmoins, dans le contexte actuel de réduction des produits phytosanitaires (Alim'agri, 2018a), la lutte biologique est également de plus en plus employée. Des études se sont par exemple penchées sur des méthodes de lutte biologique contre le champignon *Rhizoctonia solani* qui provoque la fonte des semis chez de nombreuses espèces (Gibson et al. 1999). Afin de réduire l'utilisation de produits phytosanitaires, le développement de variétés résistantes est également en plein essor avec l'arrivée des nouvelles technologies (séquençage haut débit, méthodes de transgénèse, etc.). Des études ont par exemple conduit à la création de variétés de pomme de terre résistantes à la tavelure, l'oïdium ou au feu bactérien (Kellerhals, 2004), de blés résistants à la rouille brune (Fossati et al. 2006), ou encore de variétés d'abricotiers résistants à la sharka (Mariette et al. 2016). Cependant, les résistances développées sont parfois contournées, ces méthodes de lutte ne sont pas toujours disponibles pour certaines maladies, ou ces méthodes ne sont pas assez rentables pour les mettre en œuvre. D'autres stratégies sont alors mises en place comme par exemple l'arrachage des hôtes malades. L'arrachage est couramment utilisé pour les maladies des cultures pérennes, en particulier lorsque cela revient moins cher que l'utilisation de produits chimiques (Gibson et al. 2004) ou pour les agents pathogènes qui ne peuvent pas être contrôlés par ces moyens chimiques (Martins et al. 2012). L'arrachage est par exemple pratiqué contre des pathogènes tels que la mosaïque du manioc (Fauquet et Fargette, 1990), le bunchy top du bananier (Allen, 1978), le swollen shoot du cacao (Dzahini-Obiatey et al. 2006; Thresh et Owusu, 1986), le virus de la tristeza des agrumes (Barnier et al. 2010), le virus du rabougrissement chlorotique de la patate douce (SPCSV, Gibson and Aritua, 2002) ou encore la sharka (Rimbaud et al. 2015b).

Les stratégies de gestion sont conçues avec l'aide de nombreux acteurs de la filière concernée. Par exemple pour la sharka, le Ministère de l'agriculture s'appuie sur des organismes tels que l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail) et l'INRA (Institut national de la recherche agronomique) pour définir la stratégie à mettre en place. Néanmoins, bien que ces stratégies soient largement réfléchies à l'échelle du territoire, il reste possible de les optimiser. En effet, du fait de la complexité des épidémies, il n'est pas intuitif d'identifier les stratégies les plus efficaces pour gérer un pathogène. De nos jours, les essais sur le terrain (ou en laboratoire) régissent la façon dont nous gérons les épidémies. Ces stratégies sont souvent couronnées de succès, mais les risques d'erreurs sont considérables. L'expérimentation permet de tester différents scénarios de gestion ; elle est néanmoins limitée par des contraintes logistiques et financières. En effet, elle ne permet par exemple pas de réaliser des expériences à de larges échelles spatio-temporelles, bien que cet aspect soit important à prendre en compte en épidémiologie. De plus, la variabilité des épidémies ne peut être complètement représentée avec des essais sur le terrain bien qu'elle soit importante à prendre en compte lors de l'élaboration de stratégies de gestion. D'une part, il très couteux de réaliser des essais dans des conditions variées, et d'autre part ces conditions sont difficiles à contrôler : elles proviennent à la fois des facteurs environnementaux (la température fluctue, les conditions du sol diffèrent d'un site à l'autre, la sensibilité des variétés est différente), et des interactions entre les agents pathogènes, leurs hôtes et l'environnement dans lequel ils évoluent (Shaw, 1994).

Les modèles épidémiologiques sont alors une approche intéressante car ils permettent de surmonter certaines limites de l'expérimentation. Ils permettent notamment d'**estimer des paramètres épidémiologiques** concernant certains pathogènes de manière beaucoup plus rapide, et de tester plusieurs scénarios de gestion sur différents cas épidémiques pour **identifier des stratégies optimisées** et ce, à une échelle qui peut varier de la plante à un paysage agricole.

2. Les modèles en épidémiologie

2.1. Qu'est-ce qu'un modèle ?

Un **modèle** est une représentation symbolique de certains aspects d'un objet ou d'un phénomène du **monde réel** (Pavé, 1994). Cette **représentation** peut par exemple se faire par une description mathématique d'un **système** et de son **fonctionnement**. Elle a pour but d'élaborer une théorie plus ou moins précise à partir d'observations et de prévoir ce qu'il se passerait dans certaines conditions.

Par exemple, il existe en agronomie des modèles permettant d'analyser le rendement des parcelles cultivées en fonction des conditions environnementales. Le système peut alors être composé de la plante, du sol et de l'atmosphère, et le modèle peut être basé sur des connaissances théoriques concernant les échanges entre la plante et l'atmosphère, la croissance des plantes en fonction des apports de l'environnement, etc.

Les modèles sont composés de variables d'entrées, de variables de sorties et de paramètres (Figure 1). Les variables d'entrées correspondent aux propriétés connues (ou estimées) du système ; elles sont limitées à celles auxquelles on peut raisonnablement prétendre avoir accès. Les variables de sorties représentent le résultat du modèle. Les paramètres constituent les constantes des équations mathématiques utilisées.

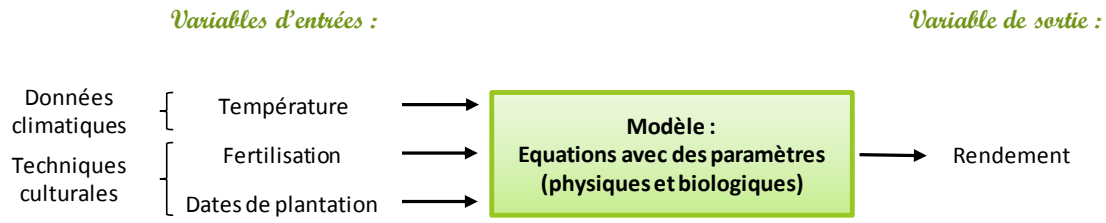


Figure 1 : Fonctionnement d'un modèle qui traduit en équations le fonctionnement d'un système composé de plantes, du sol et de l'atmosphère, pour obtenir des valeurs de rendement.

Dans un modèle, la **représentation** choisie du système est nécessairement **réductrice**. Elle est **simplifiée** à l'aide d'**hypothèses** en fonction de l'objectif. Cependant, malgré le fait que le modèle n'incorpore pas tous les éléments, il doit permettre de comprendre ou de prévoir le comportement du système. De plus, les hypothèses formulées doivent être claires afin que l'utilisateur puisse comprendre aisément le fonctionnement du modèle et juger les limites de son **domaine de validité**, en dehors duquel le modèle n'est plus applicable. En effet, un modèle répond à un objectif bien précis et n'est bien souvent pas généralisable. Par exemple, les modèles épidémiologiques sont fréquemment spécifiques à l'espèce étudiée.

Les modèles peuvent être implémentés sur des systèmes informatiques de plus en plus flexibles et performants grâce aux nombreux langages de programmation. Ils permettent de répondre à des objectifs de plus en plus complexes et précis.

2.2. Quelques caractéristiques des modèles utilisés en épidémiologie

Les premiers **modèles** représentant des épidémies ont tout d'abord été utilisés en épidémiologie médicale pour prédire l'évolution de maladies humaines (Kermack et Mckendrick, 1927; May et Anderson, 1979). Ils n'ont suscité un intérêt en épidémiologie végétale que quelques années plus tard (Van Der Plank, 1963; Zadoks, 1971). Depuis, de nombreux modèles épidémiologiques ont été développés pour décrire et analyser la dynamique spatio-temporelle de maladies des plantes transmises par des insectes vecteurs (Bertschinger, 1997; Gibson, 1997; Gottwald et al. 1999; Holt et Chancellor, 1996; Irwin et al. 1989; Madden et al. 1988; Pethybridge et Madden, 2003; Ruesink, 1986).

La plupart des modèles épidémiologique sont basés sur des équations différentielles liées décrivant les différents stades de la maladie. Ils sont dérivés d'un cadre générique qui permet à la plante (ou une partie de la plante) de passer d'un état sensible (S), à un état infecté (I), puis à un état post-infectieux (R pour « *removed* ») (modèles SIR) (Campbell et Maden, 1990; Chan et Jeger, 1994). Ces modèles peuvent intégrer des options de contrôle, et peuvent être utilisés pour évaluer leur efficacité. Kleczkowski et al. (1996) ont par exemple utilisé un **modèle SIR** pour caractériser les effets de l'inclusion d'un agent de biocontrôle, *Trichoderma viride*, sur la propagation de la maladie de la fonte des semis de radis causée par *Rhizoctonia solani*.

Une grande partie de ces modèles sont simples et **déterministes**, c'est-à-dire sans aléa : pour un état initial, il n'existe qu'un état final, i.e. tout est parfaitement connu. Ces modèles sont en général centrés sur un tissu malade (ou infecté) de la plante (Campbell et Maden, 1990). La plupart peuvent être résolus analytiquement et ont fourni des méthodes très utiles pour comparer les effets des traitements sur la dynamique d'une maladie (Gilligan, 1990). Néanmoins, les modèles déterministes ne permettent pas de prendre en compte la **variabilité** des épidémies due aux interactions entre les agents pathogènes, leur hôte et leur environnement. Des modèles **stochastiques** ont alors été développés. Ceux-ci prennent en compte des phénomènes aléatoires dus à ces interactions : pour un état initial, il existe plusieurs états finaux. Avec plusieurs répétitions de simulations, ils permettent de prédire ce qui se passera dans un grand nombre d'épidémies avec des conditions différentes : « si nous comprenons les interactions entre les agents pathogènes et leur environnement, nous pouvons alors commencer à comprendre la variabilité et peut-être à améliorer la gestion de la maladie » (Gilligan, 2002). Ainsi, suite aux travaux de Kleczkowski et al. (1996), Gibson (1997) a utilisé un modèle stochastique pour analyser l'évolution de la distribution de probabilité des épidémies de la fonte des semis en réalisant plusieurs répétitions pour prédire le risque de maladie.

Plus récemment, des modèles qui tiennent compte de la dynamique de croissance de l'hôte ont été développés. Ici, les tissus sensibles de la plante (voire la plante en entier) changent au fil des saisons et des années : de nouvelles feuilles sont produites, croissent et meurent. Ces dynamiques peuvent être capturées de deux façons principales. La première est un **modèle hôte-parasite** simple qui garde la trace de la biomasse des plantes, de la densité ou du nombre d'hôtes (qui peuvent être des plantes entières, des feuilles ou des racines) et les mesures de la biomasse parasitaire (généralement évaluée comme des lésions, ou spores) (Gilligan et Kleczkowski, 1997; Gubbins et Gilligan, 1999). La seconde, plus largement utilisée, est un **modèle SEIR** dans lequel les organes végétaux passent à travers différents compartiments : ils deviennent sensibles (S), exposés, mais pas encore infectieux (E), infectés et infectieux (I) et post-infectieux (R pour « removed ») via la mort ou l'arrachage de la plante (Gilligan, 1994).

Depuis quelques années, les modèles permettent également de travailler à une **échelle** adaptée à un **paysage agricole**. Cet aspect est important lorsque l'on a pour objectif d'optimiser des stratégies de gestion de pathogènes qui se transmettent à travers les parcelles. Gilligan et Van Den Bosch (2008) affirment d'ailleurs que la gestion d'une maladie ne peut être couronnée de succès que si l'échelle de la gestion correspond à l'échelle de l'épidémie.

Pour résumer, les modèles peuvent être **déterministes** ou **stochastiques**, **spatiaux** ou **non spatiaux**, et modélisent de manière plus ou moins complexe la biologie des hôtes, des vecteurs et des pathogènes. Ils varient en fonction de l'échelle (plante, parcelle, pays, etc.) et de l'objectif visé.

3. Les modèles pour optimiser la gestion des épidémies

3.1. Qu'est-ce que l'optimisation ?

L'optimisation d'une stratégie de gestion consiste à définir les paramètres de cette stratégie conduisant au meilleur résultat. Comme évoqué précédemment, ce résultat est caractérisé en fonction d'un objectif bien précis. Par exemple, dans le cas d'une étude agronomique, l'objectif peut être de maximiser le rendement ; dans le cas d'une étude épidémiologique, l'objectif peut être de minimiser la prévalence dans les parcelles.

Un problème d'optimisation en épidémiologie est défini par :

- un modèle qui permet de reproduire le comportement de la maladie

- un ou plusieurs critères que l'on cherche à minimiser ou maximiser
- un ensemble admissible de paramètres épidémiologiques (et parfois de gestion) qui sont contraints dans leur domaine de définition.

3.2. Optimisation de la gestion des maladies

Lorsque les paramètres épidémiologiques d'une maladie sont disponibles, des modèles permettent d'évaluer différentes stratégies de gestion et d'identifier des stratégies efficaces (Cunniffe et al. 2015a). Ces **stratégies** peuvent être basées sur des **critères** différents. Les études de Cunniffe et al. (2016) et Epanchin-Niell et al. (2012, 2014) sont par exemple focalisées sur l'optimisation d'un critère économique. D'autres travaux s'intéressent à des critères agronomiques tels que le nombre d'arbres arrachés (Parnell et al. 2009 ; Sisterson et Stenger, 2012), le nombre d'arbres asymptomatiques (Cunniffe et al. 2014), le rendement des arbres (Sisterson et Stenger, 2012), l'incidence de la maladie (Holt et al. 1999), la prévalence (Courcoul et al. 2011; Lurette et al. 2009), le taux de propagation de la maladie (Coutts et al. 2010; Filipe et al. 2012), le taux de reproduction de base (R_0) (Chan et Jeger, 1994) ou encore la proportion des plantes infectées à la fin de la simulation (Sisterson et Stenger, 2012).

En fonction de la maladie étudiée, les travaux de modélisation ont cherché à optimiser différentes stratégies de gestion. Ainsi, pour des pathogènes pour lesquels un moyen de lutte chimique existe, des études proposent une optimisation de la pulvérisation des traitements insecticides (Filipe et al. 2012) et notamment de la date du début du traitement (Cunniffe et al. 2016). Des études proposent également d'optimiser le positionnement des plantes hôtes dans les parcelles, et plus particulièrement l'espacement entre ces plantes (Chan et Jeger, 1994; Cunniffe et al. 2014, 2015b; Jeger and Chan, 1995), ainsi que la distance d'isolement des zones contaminées (Chan et Jeger, 1994; Filipe et al. 2012; Jeger et Chan, 1995). D'autres encore proposent d'améliorer la détection d'un pathogène dans les parcelles cultivées en optimisant la fréquence et l'intensité de l'échantillonnage (Parnell et al. 2012, 2014). Pour finir, dans les cas d'épidémies pour lesquelles la méthode de lutte est l'arrachage des plantes, des études ont tenté d'optimiser le nombre d'arbres arrachés, leur localisation dans les parcelles et leur replantation (Cunniffe et al. 2015b; Filipe et al. 2012; Sisterson et Stenger, 2012). Ainsi, de récents travaux de modélisation ont montré comment le contrôle par l'arrachage local des hôtes sensibles peut être efficace contre un pathogène des agrumes en considérant le rayon de la zone d'éradication (i.e. le rayon autour d'un arbre infecté délimitant un cercle dans lequel les arbres doivent être arrachés, Parnell et al. 2009, 2010), ainsi que le calendrier des dates d'arrachage (Cunniffe et al. 2014).

La plupart de ces études réalisent des simulations sur un paysage fixe et ne prennent pas en compte les caractéristiques du paysage. Pourtant, cet aspect peut être essentiel pour étudier les épidémies et les stratégies de gestion à grande échelle. En effet, le paysage peut influencer la dynamique épidémique et, par conséquent, les meilleures stratégies de gestion ne sont pas nécessairement les mêmes selon le paysage. Des approches prometteuses ont été développées pour intégrer les caractéristiques des paysages dans les modèles. Par exemple, il a été démontré qu'il était possible d'optimiser la gestion d'une épidémie du chancre des agrumes en adaptant le rayon de la zone d'arrachage (définie autour d'un arbre contaminé, Parnell et al. 2009). Ce rayon dépend principalement de la structure du paysage : le rayon optimal augmente à la fois avec le niveau d'agrégation des parcelles et la densité de l'hôte dans le paysage (Parnell et al. 2010).

Cependant, bien que ces approches considèrent le paysage, elles sont limitées par une représentation des parcelles symbolisées par des points. En outre, elles sont principalement focalisées sur un seul paramètre de gestion et ne sont pas appliquées pour des stratégies de gestion plus complexes (avec plusieurs paramètres). Peu d'études ont essayé d'optimiser plusieurs paramètres à la fois (Pleydell et al. 2018; Rimbaud et al. 2018a, 2018b).

4. La sharka comme modèle d'étude

4.1. La sharka

La maladie de la sharka est causée par le *Plum pox virus* (PPV) qui se transmet par les pucerons (selon le mode non persistant) ou par l'homme via le transfert de matériel infecté. Elle touche différentes espèces d'arbres du genre *Prunus* (abricotier, pêcher, prunier, cerisier, etc.). La présence de PPV entraine la modification de la chlorophylle dans les feuilles. La photosynthèse est alors affectée ainsi que le métabolisme des sucres, ce qui entraine une augmentation de l'acidité des fruits (Németh, 1986). De plus, les fruits infectés peuvent être déformés par le virus : ils présentent à leur surface des zones irrégulières (taches chlorotiques, Llácer et Cambra, 2006). L'impact du PPV sur les fruits les rend impropres à la consommation voire à la transformation industrielle. Les conséquences économiques sont alors importantes : une étude a estimé l'impact mondial de la sharka à 10 milliards d'euros sur 30 ans (Cambra et al. 2006), avec des pertes de rendement allant jusqu'à 100% pour les cultivars les plus sensibles.

Depuis peu, l'INRA a identifié des sources de résistance chez l'amandier et l'abricotier, et a contribué au développement de quelques variétés d'abricotiers résistants (Mariette et al. 2016). Néanmoins, il n'existe actuellement que des variétés de pêchers et de pruniers cultivés sensibles ou tolérantes au

PPV, hormis une variété transgénique chez le prunier (Scorza et al. 1994). Les chercheurs travaillent donc sur des programmes visant à développer des variétés résistantes au PPV pour ces espèces. En attendant, la gestion de la sharka en France est définie par un arrêté national de lutte, présenté dans le paragraphe suivant.

4.2. La gestion de la sharka en France

En France, la gestion de la sharka est définie par un arrêté national de lutte spécifiant une procédure complexe. Elle implique des inspections visuelles fréquentes des pépinières et des vergers, un arrachage des arbres symptomatiques (voire de vergers entiers) ainsi que des restrictions de plantation (JORF, 2011). Les différentes modalités de gestion sont explicitées sur la figure 2.

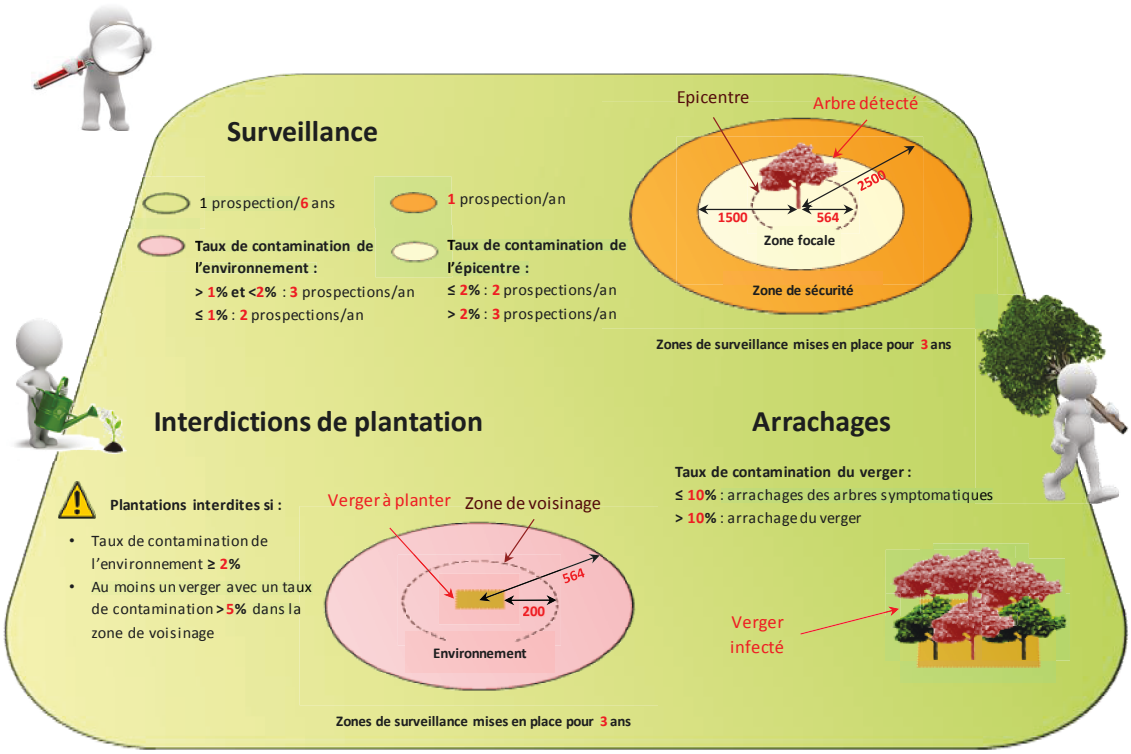


Figure 2 : Stratégie de gestion contre le *Plum pox virus* dans les vergers français, définie par l'arrêté national de lutte du 17 mars 2011 (JORF, 2011).

4.3. Optimisation des stratégies de gestion de la sharka

4.3.1. Modèle de simulation

Un modèle de simulation des épidémies de sharka et de leur gestion a été développé précédemment (Pleydell et al. 2018; Rimbaud et al. 2018a, 2018b). Ce modèle SEIR est stochastique et spatialement

explicite. Il inclut un paysage composé de zones non cultivées et de parcelles sur lesquelles des *Prunus* sont cultivés. Au début de la simulation, les arbres des parcelles cultivées ne sont pas infectés, ils ont le statut « sain ». Le PPV est introduit la première année de la simulation dans une des parcelles cultivées et se disperse ensuite à travers les vergers, provoquant le changement de statut des arbres : ils peuvent être « sains », « infectés » lorsque le virus leur est transmis, « infectieux caché » après une période de latence, « infectieux détecté » lorsque le virus est détecté sur l'arbre grâce aux prospections et « arraché » lorsque l'arbre est enlevé de la parcelle.

Dans le modèle, 6 paramètres caractérisent l'épidémie. Le premier paramètre implémenté dans le modèle contrôle la parcelle de la première introduction du pathogène ($\mathbf{q_\kappa}$). Il correspond au quantile de la connectivité de cette parcelle parmi l'ensemble des parcelles (la connectivité étant définie ici comme la probabilité qu'un puceron d'un verger se déplace sur un autre verger). De nouvelles introductions du pathogène sont contrôlées par un paramètre $\mathbf{\Phi}$, la probabilité d'introduction de la maladie lors de plantations de vergers. La prévalence lors de chaque introduction est tirée dans une distribution favorisant des valeurs plus ou moins élevées en fonction de la probabilité d'avoir une introduction massive, $\mathbf{p_{MI}}$. Le pathogène se disperse ensuite à travers les vergers suivant une fonction de dispersion qui détermine la distance de déplacement du pathogène dans le paysage (cette fonction dépend d'un paramètre $\mathbf{W_{exp}}$). Dans le verger, le nombre d'hôtes qui deviennent infectés dépend du potentiel infectieux exercé par les hôtes infectés. Ce potentiel résulte du nombre et de l'emplacement des arbres infectieux dans le paysage, et d'un coefficient de transmission $\boldsymbol{\beta}$ qui définit l'intensité globale de l'épidémie. Une fois infectés, les hôtes deviennent infectieux et symptomatiques après une période de latence dont la durée moyenne est $\boldsymbol{\theta_{exp}}$. Les 6 paramètres épidémiologiques (q_κ, Φ, p_{MI}, W_{exp}, β, θ_{exp}) peuvent varier entre une borne minimale et une borne maximale, définies pour correspondre aux caractéristiques du virus de la sharka (Rimbaud et al. 2018b).

De plus, une stratégie généralisant la gestion française de la sharka dans les vergers de *Prunus* est implémentée dans le modèle. Elle comprend 23 paramètres, détaillés dans l'article « PESO: a modelling framework to help improve management strategies for epidemics – application to sharka », présenté dans la partie 5 de l'introduction.

Une des sorties du modèle est un critère économique : la valeur actuelle nette (VAN). Elle correspond à la somme de la marge brute (MB) calculée chaque année sur laquelle est appliqué un taux d'actualisation. La marge brute représente la différence entre les produits générés par la culture des arbres productifs et les coûts induits par la production et la stratégie de gestion appliquée (observations, arrachages et replantations, Rimbaud et al. 2018a).

4.3.2. Optimisation de la gestion de la sharka

Les travaux de Rimbaud et al. (2018a) sont basés sur un paysage réel composé de parcelles cultivées en vergers de pêchers généré à partir d'une base de données recueillies dans une zone de production du Gard. Elle contient les coordonnées géographiques de 553 parcelles (0,9 ha par parcelle en moyenne). Grâce à une analyse de sensibilité, des paramètres clés de l'épidémie ont été identifiés (Rimbaud et al. 2018b) et une stratégie efficace a été identifiée dans ce paysage (Rimbaud et al. 2018a). Les paramètres de gestion de cette stratégie correspondent à une surveillance de tous les vergers du paysage de manière plus fréquente et homogène que ce que prévoit la stratégie de gestion française. De plus, quel que soit le degré de contamination local, seuls les arbres symptomatiques sont arrachés (aucun verger n'est arraché entièrement). Ce travail de modélisation a ainsi démontré de façon théorique qu'il existe des stratégies presque aussi efficaces mais moins coûteuses que la stratégie employée actuellement en France, même si cette dernière reste plus rentable que l'absence de lutte.

5. Objectifs de ma thèse

Ma thèse est intégrée dans une démarche de modélisation ayant pour but d'optimiser les stratégies de gestion d'une épidémie. L'article suivant détaille cette démarche en précisant ce qui a déjà été fait, et les travaux effectués dans le cadre de cette thèse.

Pour résumer, ma thèse peut être découpée en 3 volets, qui font partie de la démarche PESO (estimation de paramètres, simulation, optimisation) présentée dans l'article 1 :

- estimer la fonction de dispersion de la sharka à partir de données épidémiologiques et génétiques ;
- intégrer les caractéristiques paysagères dans un modèle de simulation de la maladie pour étudier l'influence du paysage sur les stratégies de gestion ;
- optimiser les stratégies de gestion d'une épidémie en prenant en compte le paysage et la possibilité d'introduction de variétés résistantes.

ARTICLE 1

PESO: a modelling framework to help improve management strategies for epidemics – application to sharka

Coralie Picard, Loup Rimbaud, Pascal Hendrikx, Samuel Soubeyrand, Emmanuel Jacquot and Gaël Thébaud

EPPO Bulletin

2017, Volume 47(2), Pages 231-236.

https://onlinelibrary.wiley.com/doi/abs/10.1111/epp.12375

Bulletin OEPP/EPPO Bulletin (2017) **47** (2), 231–236 ISSN 0250-8052. DOI: 10.1111/epp.12375

PESO: a modelling framework to help improve management strategies for epidemics – application to sharka*

Coralie Picard[1], Loup Rimbaud[2], Pascal Hendrikx[3], Samuel Soubeyrand[4], Emmanuel Jacquot[1] and Gaël Thébaud[1]

[1]*UMR BGPI, INRA, Montpellier SupAgro, Cirad, 34398, Montpellier Cedex 5 (France); e-mail: gael.thebaud@inra.fr*
[2]*CSIRO Agriculture and Food, Crace (Australia)*
[3]*ANSES, Direction des laboratoires, Unité de coordination et d'appui à la surveillance, Lyon (France)*
[4]*BioSP, INRA, Avignon (France)*

The optimization of management strategies for plant diseases is a difficult task because of the complexity and variability of epidemic dynamics. Thanks to their ability to numerically simulate many scenarios, models can be used to estimate epidemiological parameters, assess the effectiveness of different management strategies and optimize them. This article presents the PESO (parameter estimation–simulation–optimization) modelling framework to help improve plant disease management strategies. This framework is based on (i) the characterization of the epidemic dynamics to estimate key epidemiological parameters, (ii) the use of spatially explicit models to simulate epidemic dynamics and disease management, and (iii) the use of numerical optimization methods to identify better management strategies. This approach is generic and can be applied to many diseases. The work presented here focuses on sharka (caused by *Plum pox virus*), which has a worldwide impact on the *Prunus* industry, and is associated with huge disease management costs in many countries, especially in France.

Introduction

The spread of an epidemic results from complex interactions between many biological processes and, potentially, human interventions. The biological processes include for example host infection, latent and infectious periods, pathogen reproduction and dispersal, which may depend on vectors and their biology (Lepoivre, 2003). These processes may be considerably affected by other factors. For instance, the layout of plots in the landscape and their characteristics may affect pathogen dispersal (Papaïx *et al.*, 2014; Parnell *et al.*, 2010). In addition, human interventions may have a drastic impact on the epidemic. This may happen through the introduction of infected planting material into a cultivated plot, or through actions to control disease (e.g. use of pesticides, removal of infected plants). Because of this complexity, the identification of appropriate management strategies is rarely intuitive or accessible through experiments. Thus, in practice, such strategies are often based on expert opinions rather than formal demonstrations.

Epidemiological simulation models are an interesting alternative for identifying effective strategies or optimizing existing ones, because they allow a quick and thorough test of several management options *in silico*. However, it is crucial to find the best compromise between generality (theoretical models applicable in various contexts but generally not very realistic) and specificity (very realistic models but not easily transferable to other contexts), depending on the purpose of the study. In this article we introduce a generic framework called PESO (parameter estimation–simulation–optimization) that should help improve plant disease management through the use of simulation models.

In this approach, in order to obtain reliable and relevant results, the simulation model should integrate accurate knowledge about the epidemiological parameters (e.g. parameters related to host–pathogen interactions). Such understanding of the disease can be gained through experiments, field data or the subsequent estimation of epidemiological parameters, as presented in the first part of this article (Fig. 1, step A). Next, the article describes the different components of the proposed simulation approach (Fig. 1, step B) and presents different methods for the numerical optimization of management strategies (Fig. 1, step C). Here, the framework is presented as a process with successive steps. However, when applying the proposed approach to emerging diseases the process would probably be more iterative than sequential, in order to benefit from the dynamic feedback between experiments, surveillance and modelling.

This approach is applied to sharka, one of the most damaging diseases for trees of the genus *Prunus* (e.g. apricot, peach and plum). It is caused by *Plum pox virus* (PPV; genus *Potyvirus*) and spread by aphids (in a non-persistent manner) or by human activities (through the transfer of infected plant material). The associated yield losses can

*Paper based on work presented at the Joint EFSA-EPPO Workshop: Modelling in Plant Health – how can models support risk assessment of plant pests and decision-making? 12th–14th December 2016, Parma, Italy.

© 2017 The Authors. Journal compilation © 2017 OEPP/EPPO,

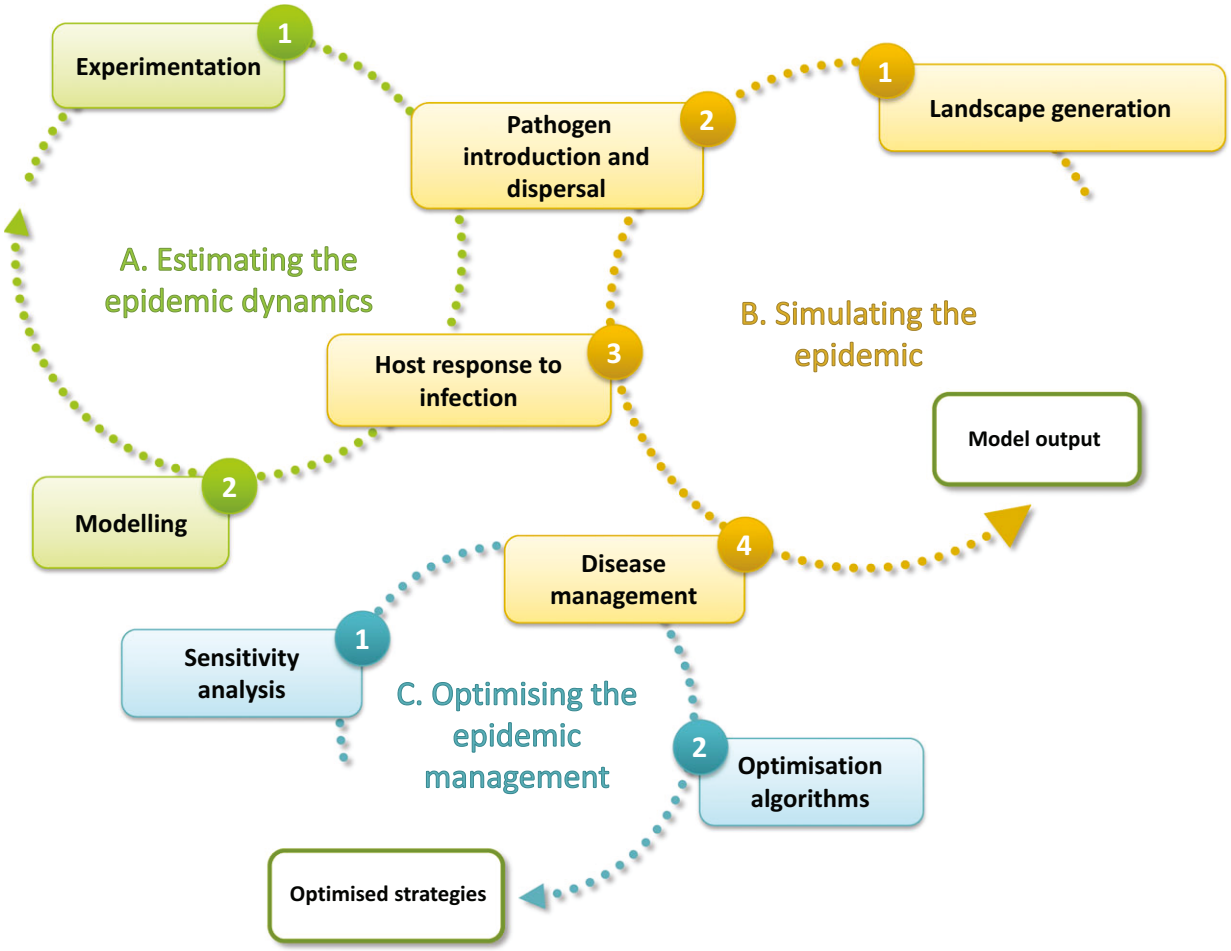


Fig. 1 The PESO modelling framework (parameter estimation–simulation–optimization) to help improve management strategies for epidemics. A: The first step consists of characterizing the epidemic dynamics (i.e. estimating key epidemiological parameters). B: The second step is the development of a simulation model including the generation of realistic landscapes, the introduction and dispersal of the pathogen and the host response to the infection. The model integrates parameter estimates from A. In addition, management strategies can be implemented in the model and assessed using the model output as a performance criterion (e.g. epidemiological or economic measurements of the strategy efficiency). C: Then, the relative influence of management parameters can be assessed using sensitivity analysis, and these parameters can be optimized directly or within optimization algorithms. [Colour figure can be viewed at wileyonlinelibrary.com].

reach 100% for the most susceptible *Prunus* cultivars (Németh, 1986). In addition, fruits of PPV-infected trees can be unsuitable for consumption or industrial processing. This damage to both quantity and quality of fruit causes important economic losses worldwide (Cambra *et al.*, 2006). Consequently, various strategies have been deployed by affected countries (Rimbaud *et al.*, 2015b). In France, sharka management is mandatory and stringent, but also very expensive. Briefly, it is based on quarantine procedures, plantation restrictions, visual inspections of *Prunus* nurseries and orchards, removal of symptomatic trees, and of whole orchards under certain conditions (JORF, 2011; Speich, 2006).

Estimating the epidemic dynamics

The first step of the approach consists of gaining insights into the processes and parameters underlying epidemic dynamics.

These processes can be studied experimentally, showing for example that PPV is not seed-transmitted (Pasquini and Barba, 2006), that latent periods can last several years, especially in apricot (Quiot *et al.*, 1995), or that the latent and incubation periods are almost synchronized for young peach trees (Rimbaud *et al.*, 2015a). Such work provides crucial information about the pathogen; however, some epidemiological parameters cannot be estimated in the laboratory. Furthermore, region-wide experiments (e.g. to test management strategies) are considerably restricted for obvious practical and ethical reasons. Modelling studies can thus provide complementary insights, especially at a scale matching that of typical epidemics.

Models fed with field surveillance data can provide information on epidemiological parameters. For instance, empirical models were used to describe nearest-neighbour distance distributions between infected trees (Gottwald

et al., 2013). More recently, some epidemiological parameters such as the dispersal kernel, latent period and transmission rate were estimated from sharka data collected in more than 600 French orchards over 15 years using a model at the orchard level (Pleydell *et al.*, 2017). Improved estimates may be obtained using such process-based models with tree-level epidemiological data.

An innovative approach consists of estimating more precisely epidemiological parameters to infer 'who infected whom' in outbreaks of infectious diseases, by jointly using spatiotemporal and genetic data (Mollentze *et al.*, 2014). For this purpose, virus sequences should be obtained from most infected hosts within a spatial window matching the scale of disease dispersal. Genetic variations resulting from the high mutation rate of viruses (Drummond *et al.*, 2003) provide information on the genealogical relationships between virus genome sequences, and thus between host infections. Models can account for the probability that any case A infected any case B given the pathogen sequences and the spatiotemporal locations of the infections. This probability is broken down into: the probability that case A was infectious and case B was infected during the same time window; the probability that cases A and B have been in contact (either directly, or indirectly via wind or vectors) given the locations at which A and B were observed; and the probability that the genetic sequence of the pathogen sampled from case A mutated into the sequence sampled from case B given the evolutionary time separating the two sequences. This enables the joint inference of epidemiological parameters (infection date, latent and infectious periods, dispersal function) and pathogen transmission chains. Despite its potential, this approach still has to be applied to plant pathogens (Picard *et al.*, 2017).

Simulating the epidemic

Once epidemic dynamics have been characterized (up to uncertainties that have to be taken into account), simulation models can be exploited to test different epidemic or management scenarios. Such models usually need an explicit and realistic landscape, a scenario for pathogen introduction and dispersal, equations describing the changes in host status and, possibly, management actions that affect disease spread. Finally, different management strategies can be compared based on output variables.

Modelling landscape features

The landscape (characterized by the spatial arrangement and the shape of the different plots) is not often taken into account in epidemiological modelling studies, although it may have a strong impact (Ostfeld *et al.*, 2005; Papaïx *et al.*, 2014). In landscapes where plot dimensions are of the same order of magnitude as inter-plot distances, accurate models of disease spread and management require a realistic representation of the landscape. To achieve this, two options are available: either using the coordinates of elements of a real landscape (Rimbaud, 2015) or developing an algorithm simulating realistic landscapes. The first option has the advantage of realism if a decision has to be made for that specific landscape, or similar ones. The second option enables the main landscape features to be defined, such as the number of plots or their spatial aggregation (i.e. the proportion of the neighbouring area of a plot filled with other plots).

Modelling pathogen introduction and dispersal

In order to simulate an epidemic, the first step is to model pathogen introduction by selecting the first infected host(s). The outbreak can be initiated by one or several infected hosts, depending on whether this outbreak is due to the introduction of the pathogen from an infected host located outside the study area or to the planting of infected individuals in the plot. Each infectious host can be a source of infection for close or distant susceptible hosts, depending on the transmission processes. In epidemiological models, dispersal distances are characterized by a function which represents the statistical distribution of the location of the inoculum after dispersal from a source point. In the work presented in this paper, a stochastic, spatially explicit model first developed on sharka is used (Pleydell *et al.*, 2017; Rimbaud, 2015). In this orchard-based model, PPV is introduced at the beginning of the simulation and new introductions can occur each time a plot is replanted. Dispersal of PPV in the landscape is modelled using flexible mixtures of exponential functions that simulate the first flight of the aphids away from an infected host.

Modelling host–pathogen interactions

Many epidemiological models rely on the SEIR (susceptible–exposed–infectious–removed) architecture. In these models, hosts are characterized by different states. At the beginning of the simulation, hosts are 'susceptible' (i.e. healthy). After disease introduction and dispersal, the host status can change: some hosts become 'exposed' (i.e. infected but not yet infectious), and next 'infectious' after a latent period. Finally, their infectious period stops when they become 'removed' (i.e. epidemiologically inactive or dead due to the disease or to human intervention). In the case of sharka, present knowledge indicates that the infectious hosts show symptoms and that symptomatic hosts are infectious (Rimbaud *et al.*, 2015a). To account for disease detection through symptom-based surveillance, the 'infectious' compartment can be split into 'infectious hidden' (i.e. not yet detected) and 'infectious detected'.

Modelling disease management

A management strategy based on plot surveillance, host removals and plantation restriction is implemented in the

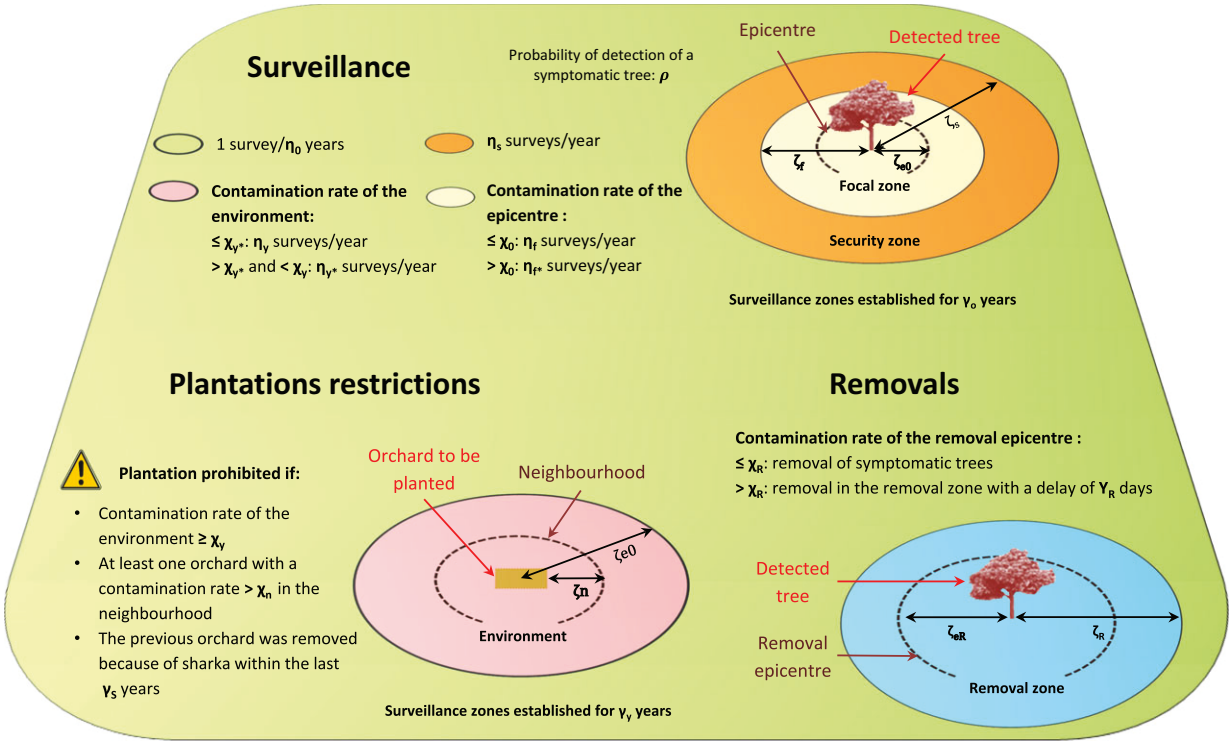


Fig. 2 Summary of sharka management strategies implemented in the model. [Colour figure can be viewed at wileyonlinelibrary.com].

model. The modalities of this strategy (Fig. 2; see Table 1 for parameter definitions) are mainly based on the French strategy for managing sharka in *Prunus* orchards (JORF, 2011), but have been extended to include some strategies applied in other countries, for example the United States (Gottwald *et al.*, 2013).

Choice of the output variable

The output variable of the simulation model should summarize the epidemic and provide relevant information to assess the performance of management strategies. Some output variables are very simple descriptions of the epidemic (e.g. disease prevalence or incidence) or its management (e.g. number of removed hosts, total cost of disease management). Nevertheless, not all such variables are relevant in the context of management optimization (e.g. the strategy minimizing the cost of management is the absence of control), and it may be necessary to design composite criteria. Two different outputs are designed in the work performed on sharka. The first is the mean number of productive hosts per hectare and per year, which is a good proxy for the epidemiological control provided by a given strategy. The second output is the net present value (NPV), which is an economic criterion balancing costs against benefits (and short-term against long-term value). The benefit is generated by the cultivation of healthy or asymptomatic trees (consequently, this benefit indirectly accounts for yield losses due to the disease). The costs are associated with *Prunus* production (independently of the disease) and management actions to control sharka (including the surveillance, removal and replanting of trees; Rimbaud, 2015).

Optimizing the epidemic management

Optimization of a management strategy consists of identifying, or at least approaching, the parameters of this strategy leading to the best outcome.

A first study of various sharka management strategies was carried out using the spatiotemporal simulation model mentioned above. This model used the real landscape of a peach-growing region of Southeastern France, consisting of a mosaic of host (553 plots) and non-host areas. It demonstrated the epidemiological efficiency of the current French management strategy, and proposed ways to improve its economic profitability (Rimbaud, 2015). Briefly, optimized strategies are based on more regular and homogeneous surveillance of the whole landscape than the current French management strategy. In addition, parameter values that lead to orchard removal (e.g. when the contamination rate of an orchard is above 10%) are not recommended by this optimization.

Because such strategies are specific to a particular fixed landscape (Rimbaud, 2015), it could be valuable to test the robustness of the French management strategy and of the optimized strategies to changes in the landscape

Table 1. Sharka management parameters used in the model

Symbol	Control parameters
ρ	Probability for a given symptomatic tree to be detected during one inspection
δ	Mean delay before removal of a detected tree (day)
Y_R	(Boolean) After surveillance, whole orchards are removed: 0, after a mean delay of δ; 1, at the end of the year
γ_S	Delay before replanting of a removed orchard (year)
γ_o	Duration of surveillance zones (year)
γ_y	Duration of young orchards (year)
ζ_s	Radius of the security zone (m)
ζ_f	Radius of the focal zone (m)
ζ_{e0}	Radius of the surveillance epicentre (m)
ζ_n	Radius of the neighbourhood (m)
ζ_R	Radius of the removal zone (m)
ζ_{eR}	Radius of the removal epicentre (m)
η_0	Maximal period between two observations (year)
η_s	Surveillance frequency in security zones (year^{-1})
η_f	Surveillance frequency in focal zones (year^{-1})
η_{f*}	Modified surveillance frequency in focal zones (year^{-1})
η_y	Surveillance frequency in young orchards (year^{-1})
η_{y*}	Modified surveillance frequency in young orchards (year^{-1})
χ_o	Contamination threshold in the surveillance epicentre, above which surveillance frequency in the focal zone is modified
χ_y	Contamination threshold in the environment around young orchards, above which orchard planting is forbidden
χ_{y*}	Contamination threshold in the environment, above which surveillance frequency in young orchards is modified
χ_n	Contamination threshold for any orchard in the neighbourhood, above which orchard planting is forbidden
χ_R	Contamination threshold in the removal epicentre, above which all the orchards inside the removal zone are removed

characteristics. Sensitivity analyses should allow the identification of the most influential control parameters for a range of landscapes varying by their degree of spatial aggregation. Then, landscape-generic or landscape-specific optimization of management strategies may be achieved by analysing the output of numerous parameter combinations. It is important to examine whether a management strategy has to be tailored to each landscape (i.e. a unique strategy leads to very different results for various landscapes), or if a unique management strategy could fit all landscapes (i.e. a strategy leads to a good economic outcome for all landscape types).

By exploring a large number of parameter combinations, such work would help approach the best management strategy for various landscapes. However, this exploration is likely to be constrained by the computational time associated with the great number of parameter combinations. In this context, specific optimization algorithms may be necessary to improve parameter space exploration and more quickly or more accurately reach the optimal parameter combination.

Conclusions

The PESO framework is based on the estimation of key epidemiological parameters, the simulation of the epidemic and its management, and the optimization of management parameters. This work has already shown its efficiency by proposing improved strategies for sharka management in a specific context (Rimbaud, 2015), and deserves to be extended. Significant improvements might be brought about by a more precise estimation of the aphid dispersal function thanks to novel estimation models that can jointly use spatial, temporal and genetic data about the pathogen. In addition, an in-depth study of the impact of landscape characteristics on disease spread and management strategies may open new avenues for the practical use of such results in various landscapes. New optimization approaches may also improve the exploration of the vast number of possible combinations of management parameters in order to find the most efficient management strategy. The authors hope that the PESO approach presented here can serve as a generic conceptual framework for research on the computer-assisted design of disease management strategies. Furthermore, it should help promote science-based decision-making through interactions with the organizations responsible for sharka –and more generally plant disease– management.

Acknowledgements

This work has been supported by a DGA/MRIS PhD grant to LR, and by an INRA/ANSES PhD grant to CP.

Un cadre de modélisation pour aider à améliorer les stratégies de gestion des épidémies - application à la sharka

L'optimisation des stratégies de gestion des maladies des plantes est une tâche difficile en raison de la complexité et de la variabilité des dynamiques épidémiques. Grâce à leur capacité à simuler numériquement de nombreux scénarios, les modèles peuvent être utilisés pour estimer des paramètres épidémiologiques, évaluer l'efficacité de différentes stratégies de gestion et les optimiser. Cet article présente le cadre de modélisation PESO (estimation de paramètres–simulation–optimisation) pour aider à améliorer les stratégies de gestion des maladies de plantes. Ce cadre est basé sur (i) la caractérisation des dynamiques épidémiques pour estimer des paramètres épidémiologiques clés, (ii) l'utilisation de modèles spatialement explicites pour simuler les dynamiques épidémiques et la gestion des maladies, et (iii) l'utilisation de méthodes d'optimisation numérique pour identifier de meilleures stratégies de gestion. Cette approche est générique et peut être appliquée à de nombreuses maladies. Le travail présenté ici porte sur la sharka (causée par le *Plum pox virus*), qui a un impact sur l'industrie mondiale des *Prunus* et est associée à d'énormes coûts de gestion dans de nombreux pays, et en particulier en France.

PESO: система моделирования, направленная на совершенствование стратегии борьбы с эпидемиями - применение к болезни шарка

Вследствие сложности и изменчивости динамики эпидемий, оптимизация стратегий борьбы с болезнями растений остается трудной задачей. Благодаря их способности, симулировать различные сценарии в числовом виде, модели могут быть использованы для оценки эпидемиологических параметров, определения эффективности различных стратегий борьбы и их оптимизации. В статье представлена система моделирования PESO (оценка параметров - симуляция - оптимизация), позволяющая усовершенствовать стратегии борьбы с болезнями растений. Система моделирования основана на: (1) характеризации динамики эпидемий с целью определения ключевых эпидемиологических параметров; (2) использовании детальных пространственных моделей для симуляции динамики эпидемий и борьбы с болезнями; (3) использовании методов численной оптимизации для выявления наиболее эффективных стратегий борьбы. Этот подход является универсальным и может быть применен ко многим болезням. Представленная работа посвящена болезни шарка (вызванной *Plum pox virus*), которая повсеместно наносит ущерб промышленности выращивания слив и приводит к огромным расходам во многих странах, особенно во Франции.

References

Cambra M, Capote N, Myrta A & Llácer G (2006) *Plum pox virus* and the estimated costs associated with sharka disease. *Bulletin OEPP/EPPO Bulletin* **36**, 202–204.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R & Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology & Evolution* **18**, 481–488.

Gottwald TR, Wierenga E, Luo W & Parnell S (2013) Epidemiology of Plum pox 'D' strain in Canada and the USA. *Canadian Journal of Plant Pathology* **35**, 442–457.

JORF (2011) Arrêté du 17 mars 2011 relatif à la lutte contre le Plum Pox Virus, agent causal de la maladie de la sharka, sur les végétaux sensibles du genre Prunus. JORF n°0067 du 20 mars 2011. République Française. Paris, France. NOR: AGRG1105295A.

Lepoivre P (2003) *Phytopathologie*. De Boeck, Bruxelles (BE).

Mollentze N, Nel LH, Townsend S, le Roux K, Hampson K, Haydon DT *et al.* (2014) A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proceedings of the Royal Society B* **281**, 20133251.

Németh M (1986) Plum pox (Sharka). In: *Virus, Mycoplasma and Rickettsia Diseases of Fruit Trees*. (eds Nijhoff M & Junk W), pp. 463–479. Dordrecht (NL).

Ostfeld RS, Glass GE & Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution* **20**, 328–336.

Papaïx J, Adamczyk-Chauvat K, Bouvier A, Kiêu K, Touzeau S, Lannou C *et al.* (2014) Pathogen population dynamics in agricultural landscapes: The Ddal modelling framework. *Infection, Genetics and Evolution* **27**, 509–520.

Parnell S, Gottwald TR, Gilligan CA, Cunniffe NJ & van den Bosch F (2010) The effect of landscape pattern on the optimal eradication zone of an invading epidemic. *Phytopathology* **100**, 638–644.

Pasquini G & Barba M (2006) The question of seed transmissibility of *Plum pox virus*. *Bulletin OEPP/EPPO Bulletin* **36**, 287–292.

Picard C, Dallot S, Brunker K, Berthier K, Roumagnac P, Soubeyrand S *et al.* (2017) Exploiting genetic information to trace plant virus dispersal in landscapes. *Annual Review of Phytopathology* **55**. (in press).

Pleydell D, Soubeyrand S, Dallot S, Labonne G, Chadœuf J, Jacquot E *et al.* (2017) Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. *BioRxiv* 109561. doi: https://doi.org/10.1101/109561

Quiot JB, Boeglin M, Adamolle C, Candresse T, Labonne G & Renaud LY (1995) Behaviour of two isolates of Plum pox virus inoculated on peach and apricot trees: first results. *Acta Horticulturae* **386**, 290–298.

Rimbaud L (2015) Conception et évaluation assistée par la modélisation de stratégies de gestion d'une épidémie dans un paysage hétérogène. PhD Thesis. Montpellier SupAgro, Montpellier (FR).

Rimbaud L, Dallot S, Delaunay A, Borron S, Soubeyrand S, Thébaud G *et al.* (2015a) Assessing the mismatch between incubation and latent periods for vector-borne diseases: the case of sharka. *Phytopathology* **105**, 1408–1416.

Rimbaud L, Dallot S, Gottwald T, Decroocq V, Jacquot E, Soubeyrand S *et al.* (2015b) Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. *Annual Review of Phytopathology* **53**, 357–378.

Speich P (2006) Control and monitoring: *Plum pox virus* quarantine situation in France. *Bulletin OEPP/EPPO Bulletin* **36**, 307–308.

Résultats clés de l'Article 1

UN PROCESSUS DE MODELISATION POUR AIDER A AMELIORER LES STRATEGIES DE GESTION DES EPIDEMIES : APPLICATION A LA SHARKA

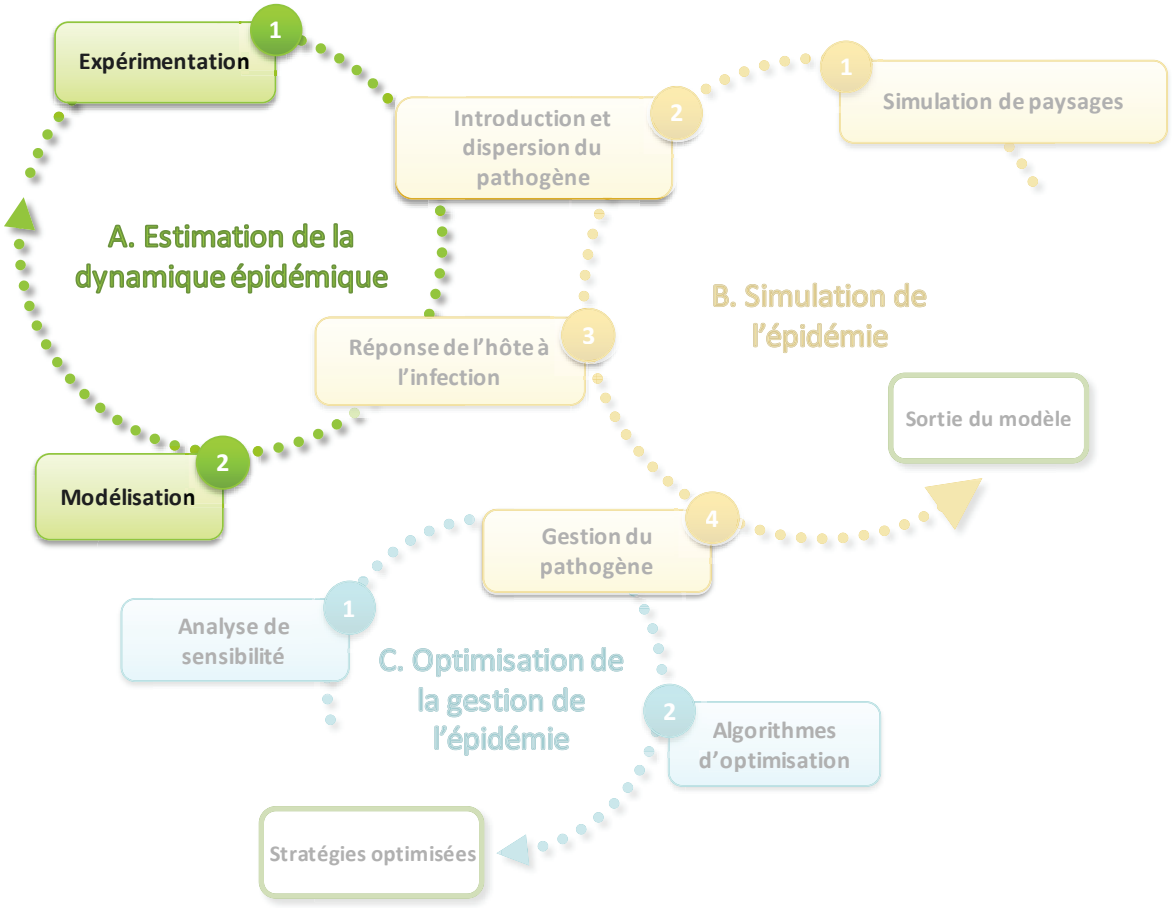
L'optimisation des stratégies de gestion des maladies des plantes peut être compliquée en raison de la complexité des épidémies (qui dépendent à la fois de processus biologiques, des interventions humaines et de l'organisation des parcelles dans le paysage). Grâce à leur capacité à simuler de nombreux scénarios, les modèles peuvent être utilisés pour estimer les paramètres épidémiologiques, ainsi que pour évaluer et optimiser l'efficacité de différentes stratégies de gestion. Cet article présente le processus de modélisation PESO (estimation de paramètres, simulation et optimisation) pour aider à améliorer les stratégies de gestion des maladies des plantes. Ce processus est basé sur :

- la caractérisation de la dynamique épidémique pour estimer les paramètres épidémiologiques clés ;
- l'utilisation de modèles spatialement explicites pour simuler la dynamique épidémique et la gestion de la maladie considérée ;
- l'utilisation de méthodes d'optimisation numériques pour identifier les meilleures stratégies de gestion.

Cette approche est générique et peut être appliquée à de nombreuses maladies. Elle est ici appliquée à la maladie de la sharka (causée par le *Plum pox virus*), qui induit des coûts importants dans de nombreux pays, et notamment en France.

Chapitre 1

ESTIMATION DE PARAMETRES EPIDEMIOLOGIQUES



La première étape du processus de modélisation PESO consiste à estimer les paramètres qui caractérisent une épidémie. Comme cela a été présenté en introduction, les paramètres régissant les épidémies de sharka ont été estimés par Pleydell et al. (2018). Néanmoins, les données utilisées pour réaliser ces estimations ne prennent pas en compte la localisation exacte des arbres infectés, mais la proportion d'arbres infectés par parcelle, ce qui peut réduire la précision de l'estimation de la fonction de dispersion du virus. En effet, la connectivité des parcelles a été calculée à partir de leurs centroïdes, ce qui peut par exemple entrainer un biais dans l'estimation de la fonction de dispersion si un seul côté d'une parcelle comprend des arbres infectés. L'objectif de ce volet de ma thèse consiste à estimer de manière plus précise les paramètres épidémiologiques qui caractérisent la sharka, en utilisant des données épidémiologiques acquises au grain de l'arbre ainsi que des données génétiques.

Pour cela, j'ai tout d'abord réalisé une synthèse bibliographique (présentée dans la première partie de ce chapitre sous forme de revue) qui explique comment des données épidémiques et génétiques peuvent aider à la compréhension des épidémies. Dans une deuxième partie, nous avons tenté d'estimer plusieurs paramètres épidémiologiques de la sharka à l'aide d'un modèle visant à reconstruire les liens de transmission entre les hôtes individuels (en inférant « qui a infecté qui » dans le paysage).

1. Les modèles pour comprendre la dynamique des épidémies

Afin de comprendre la dynamique des épidémies, il est crucial d'identifier comment (voie de transmission), quand (période de transmission et fréquence), et où (hôte, emplacement et distance) ces pathogènes sont transmis. Pour cela, l'épidémiologie moléculaire est de plus en plus utilisée : cette approche exploite l'information sur la variabilité génétique des agents pathogènes pour caractériser leur dispersion et leur évolution. En particulier, des approches permettant d'estimer les paramètres épidémiologiques d'une maladie et d'identifier les voies de transmission de l'agent pathogène responsable entre les hôtes ou les populations hôtes ont été développées depuis une dizaine d'années.

La revue suivante présente certaines de ces approches qui exploitent l'information génétique pour suivre la dispersion d'un virus à travers un paysage. Dans le cadre de ma thèse, j'ai notamment contribué à l'écriture de la 3ème partie qui traite de l'inférence des arbres de transmission des maladies et de l'estimation des paramètres épidémiologiques, ainsi qu'à l'introduction et à la discussion.

ARTICLE 2

Exploiting Genetic Information to Trace Plant Virus Dispersal in Landscapes

Coralie Picard, Sylvie Dallot, Kirstyn Brunker, Karine Berthier, Philippe Roumagnac, Samuel Soubeyrand, Emmanuel Jacquot and Gaël Thébaud

Annual Review of Phytopathology

2017, Volume 55, Pages 139-160.

https://www.annualreviews.org/doi/full/10.1146/annurev-phyto-080516-035616

ANNUAL REVIEWS

Annual Review of Phytopathology

Exploiting Genetic Information to Trace Plant Virus Dispersal in Landscapes

Coralie Picard,[1,*] Sylvie Dallot,[1,*] Kirstyn Brunker,[2] Karine Berthier,[3] Philippe Roumagnac,[1] Samuel Soubeyrand,[4] Emmanuel Jacquot,[1] and Gaël Thébaud[1]

[1]UMR BGPI, INRA, Montpellier SupAgro, CIRAD, 34398, Montpellier Cedex 5, France; email: gael.thebaud@inra.fr

[2]Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, Glasgow, G12 8QQ, United Kingdom

[3]Pathologie Végétale, INRA, 84140, Montfavet, France

[4]BioSP, INRA, 84914, Avignon, France

Annu. Rev. Phytopathol. 2017. 55:139–60

First published as a Review in Advance on May 19, 2017

The *Annual Review of Phytopathology* is online at phyto.annualreviews.org

https://doi.org/10.1146/annurev-phyto-080516-035616

Copyright © 2017 by Annual Reviews.
All rights reserved

*These authors contributed equally to this review.

ANNUAL REVIEWS **Further**

Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Keywords

high-throughput sequencing, host range, invasion pathway, outbreak, transmission tree, vector

Abstract

During the past decade, knowledge of pathogen life history has greatly benefited from the advent and development of molecular epidemiology. This branch of epidemiology uses information on pathogen variation at the molecular level to gain insights into a pathogen's niche and evolution and to characterize pathogen dispersal within and between host populations. Here, we review molecular epidemiology approaches that have been developed to trace plant virus dispersal in landscapes. In particular, we highlight how virus molecular epidemiology, nourished with powerful sequencing technologies, can provide novel insights at the crossroads between the blooming fields of landscape genetics, phylogeography, and evolutionary epidemiology. We present existing approaches and their limitations and contributions to the

INTRODUCTION

bstitution rate: e of fixation of netic changes in a cies

Epidemics caused by the spread of pathogenic agents through host populations can be a high socioeconomic burden (71, 143). In order to support public policy decision-making regarding disease control strategies, scientists need to understand and, ultimately, quantify and predict how pathogens spread within and between host populations. This understanding has been recently improved by attempts to trace pathogen dispersal using molecular epidemiology and novel statistical approaches. Molecular epidemiology uses information on pathogen genetic variation to unravel the niche of a pathogen (including host and vector species) and characterize its dispersal and evolution (129). Such studies focus on the identification of risk factors that affect host exposure or intrinsic susceptibility to pathogens and on the dispersal of these pathogens from infected to susceptible hosts (8). In order to understand and control epidemics, it is indeed crucial to identify how (transmission route), when (transmission period and frequency), and where (host, location, and distance) pathogens are transmitted.

Although, ideally, fully documented epidemiological records would provide a wealth of necessary information, such a detailed level of pathogen-tracing information is not attainable in practice. However, even incomplete and indirect information on pathogen dispersal—such as host range, population connections, and epidemic origin and spread—can be highly valuable. In particular, the quantification of pathogen transmission across various distances, and specifically the characterization of long-distance dispersal events, has major implications for disease management strategies. To address these issues, pathogen tracing relies on indirect approaches that derive epidemiological information from the spatiotemporal structure of pathogen genetic diversity. Viruses are particularly amenable to such studies because their epidemiological and evolutionary dynamics occur at similar short timescales. Moreover, the high number of polymorphisms in their small genomes can be accessed relatively easily, and increasingly in real time, during epidemics (32, 60). As such, viruses are "measurably evolving" pathogens (7, 29).

lemental Material

The number of research articles published on virus molecular epidemiology has increased steadily since the 1990s—and since the 2000s for plant viruses (see **Supplemental Figure 1**). There are a few review articles on the use of plant virus diversity in evolutionary epidemiology (62, 92) or disease emergence studies (36, 40, 60, 105). As a complementary perspective, our purpose here is to specifically review molecular epidemiology approaches for plant viruses and to focus on how the molecular analysis of virus diversity provides insights into the spatiotemporal dynamics of plant virus epidemics. To this end, we explore three questions addressed by scientists in order to trace plant virus dispersal in landscapes: How to find the hosts and access virus diversity? What are the spatiotemporal history and predictors of virus flows in landscapes? How did the virus spread within an outbreak?

HOW TO ACCESS PATHOGEN DIVERSITY IN LANDSCAPES?

The vast majority of plant viruses have single-stranded DNA (ssDNA) or positive-stranded RNA genomes, which have a higher substitution rate (mostly ranging from 10^{-3} to 10^{-5} substitutions/site/year) than other genomes (7, 46, 125, 126). Proofreading-deficient polymerases, short generation times, and frequent bottlenecks on large populations all contribute to the impact of evolutionary forces on virus populations (38, 93) (see sidebar titled Evolutionary Processes Imprint Virus Genomes). Consequently, viral populations often show a high level of genetic diversity both within and between hosts (8, 32) (**Figure 1**). Characterizing the genetic structure and diversity

EVOLUTIONARY PROCESSES IMPRINT VIRUS GENOMES

Five evolutionary forces shape the genomes and genetic diversity of virus populations (60). The resulting patterns provide information on the underlying processes (50):

Mutation: The amount of de novo nucleotide diversity accessible across the spatiotemporal scales impacts the questions that molecular epidemiology approaches can address.

Recombination/reassortment: The generation of novel genetic combinations increases genomic diversity and thus adaptation. Viruses are mostly haploid and clonal, so this process is too infrequent to assume independence between loci.

Migration: The spatial reallocation of genomes increases genetic diversity within (and reduces differentiation between) populations. Estimation of migration rates provides key information on virus flows.

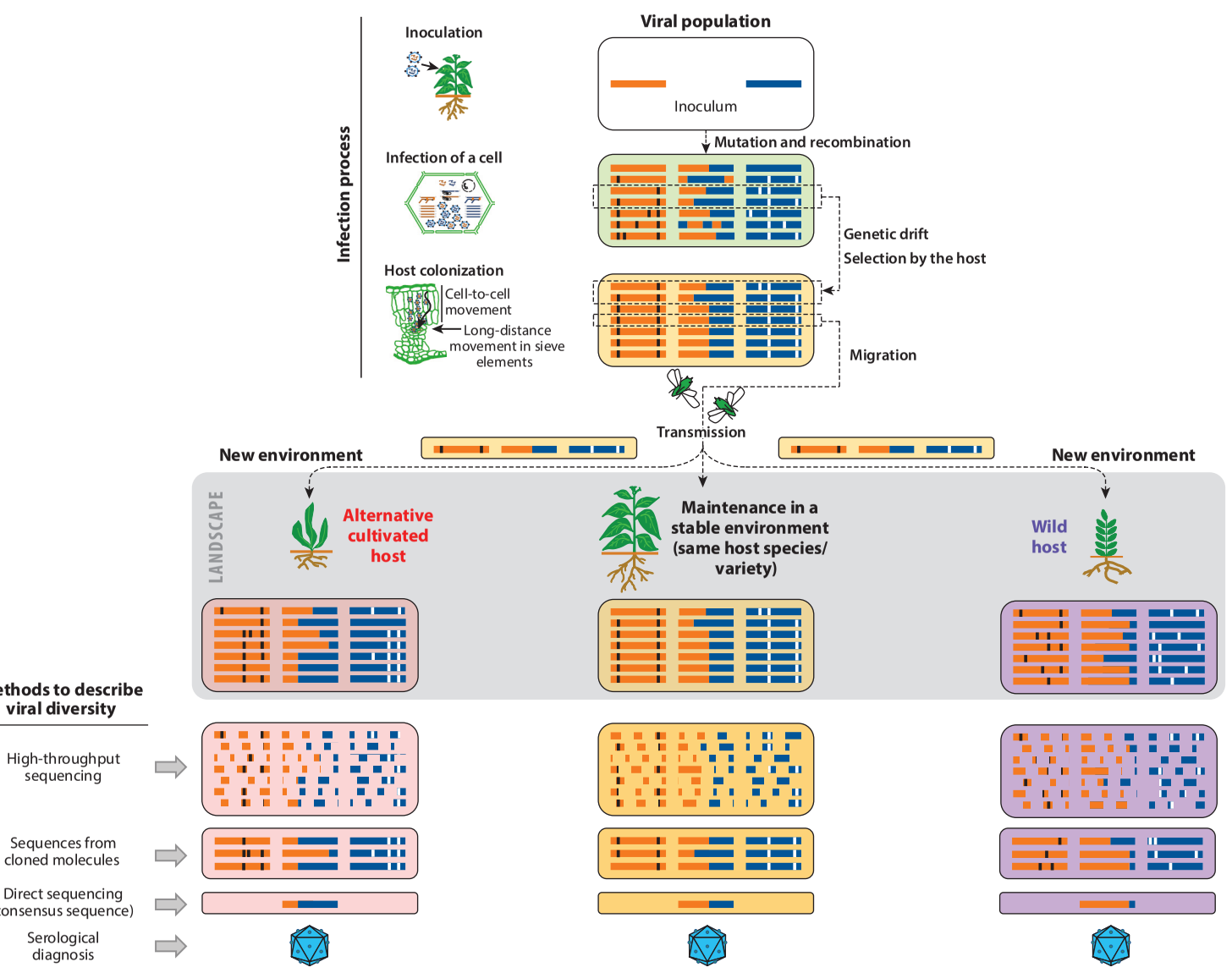
Selection: Fueled by the previous processes, selection is the engine of adaptation and can be stabilizing, directional, or diversifying. Environment-specific selection at some loci can pinpoint the original environment of a genome; however, highly reproducible mutational pathways toward adapted genotypes generate genetic homoplasy (i.e., shared polymorphism absent from the common ancestor) that can be misinterpreted as recombination and thus blur analyses.

Drift: The random sampling of individual genomes founding the next generation changes allele frequencies within a population. Drift promotes fixation of neutral (or slightly deleterious) mutations and thus increases differentiation between populations.

(*b*) the characteristics of the targeted potential hosts (plants/vectors and wild/cultivated and annual/perennial plants), (*c*) the spatiotemporal dynamics of the viral disease, and (*d*) the evolutionary rate of the virus under study. The past four decades have seen a huge evolution in the techniques used to reveal molecular polymorphisms and to sequence genomes.

Characterization of Virus Diversity

In order to propose a classification of viral species and to explore the diversity within virus species, the scientific community initially used biological properties of plant viruses, such as their host range, induced symptoms, and transmission properties, including the range of vectors involved. However, it was later shown that biological approaches are rarely adequate to reveal the structure and diversity of plant virus populations, as most of the polymorphisms of viral genomes have no effect on these biological parameters. In the 1970s, the development of techniques based on the antigenic properties of the capsid protein (18) shed a new light on the variability between and within viral species. Molecular techniques developed in the 1970s–1980s and widely used since the 1990s strongly modified plant virus epidemiology approaches. They allowed the direct characterization of viral genomes through the development of various molecular markers [e.g., restriction fragment length polymorphism (RFLP) (48), single-strand conformation polymorphism (SSCP) (102), ribonuclease protection assay (RPA) (42), and RNase T1 fingerprint (119)] and partial- or whole-genome sequencing using Sanger technology (124) on amplified [e.g., polymerase chain reaction (PCR) or rolling circle amplification (RCA)] products (128) or cloned molecules. Besides providing a quantitative estimate of the viral genetic diversity from within-host to global


Infection process
Inoculation
Infection of a cell
Host colonization
Cell-to-cell movement
Long-distance movement in sieve elements
Viral population
Inoculum
Mutation and recombination
Genetic drift
Selection by the host
Migration
Transmission
New environment
New environment
LANDSCAPE
Alternative cultivated host
Maintenance in a stable environment (same host species/ variety)
Wild host
ethods to describe viral diversity
High-throughput sequencing
Sequences from cloned molecules
Direct sequencing consensus sequence)
Serological diagnosis

Supplemental Material

RECOMBINANT GENOMES: TROUBLE OR TREASURE?

The access to full-length viral genomes has highlighted the major role of recombination in the evolution of RNA and DNA plant viruses (reviewed in 84, 138).

Recombination is known to blur phylogenetic signals; thus, ignoring it when reconstructing the evolutionary histories of viruses will likely lead to misleading inferences (127). It is therefore highly desirable to either exclude recombinant sequences or focus the analysis on nonrecombined genomic regions. Numerous methods and computer programs have been developed for detecting recombination and locating recombination breakpoints (see **Supplemental Table 1**) (85). Their relative performance in terms of power (probability to detect true recombination events) and specificity (avoidance of false positives) has been assessed (107).

However, accounting for recombination may provide extra information to infer transmission trees. Indeed, such evolutionary events occur during multiple infections, which imply that viruses with potentially different geographical origins have simultaneously shared the same host. Nonrandom patterns of sequence exchanges may also provide valuable information about potential geographical or ecological barriers (72). Despite methodological developments such as ancestral recombination graphs (98), computational and theoretical obstacles remain before we can truly integrate recombination in phylodynamic inference (43).

and/or particular recombination breakpoints (12, 120). Moreover, whole-genome sequencing offered unprecedented insights into the infraspecific genetic polymorphism and further evidenced the major role of recombination and reassortment in plant virus evolution (100, 138) (see sidebar titled Recombinant Genomes: Trouble or Treasure?). However, Sanger sequencing approaches have limited throughput, are both resource- and labor-intensive, and depend upon a priori knowledge of virus sequences. Moreover, polymerase-based techniques may be error-prone depending on the enzyme used. Thus, the corresponding data may not always reflect the true viral genetic diversity (94).

With the recent advent of high-throughput sequencing (HTS) technologies, the ability to generate large amounts of sequence data at relatively low cost led to breakthroughs in plant virus discovery and molecular epidemiology. Because they require little a priori knowledge of the targeted virus, metagenomic approaches have enabled the identification of hundreds of unknown viruses (17, 104, 123) as well as the discovery of new variants of known virus species that escaped existing detection procedures (83). Such approaches will undoubtedly improve our understanding of the distribution and dynamics of plant virus diversity in both cultivated and natural areas (123). In addition, HTS technologies can be used to generate consensus genome sequences without an amplification step (79) or a deep characterization of within-host diversity (20, 133).

Reassortment: exchange of full genomic segments resulting in infectious units with new combinations of segments

HTS: high-throughput sequencing

Figure 1

Evolutionary processes leading to the viral diversity observed in a heterogeneous landscape. Different steps of the infection process are presented, from the inoculation of a single cell by two different viral genomes (*orange and blue*) to the systemic infection of the host. Virus replication is an error-prone process that results in the diversification of the viral genetic material through mutation (*white and black bars*) and recombination events (orange/blue chimeric genomes). A population of viral genomes is therefore generated during infection. However, the selection of fitter individuals at the cellular level combined with bottlenecks occurring during host colonization reduces the range of genetic variation within the infected host. Additional bottlenecks during plant-to-plant transmission (generally via vectors) lead to the efficient inoculation of a limited number of virus genomes. The epidemiological processes then shape the viral population according to host features. Viral populations can be characterized using serological [e.g., enzyme-linked immunosorbent

New sequencing technologies, including single-molecule real-time (SMRT) sequencing and other long-read sequencing technologies, should also provide solutions for real-time genomic surveillance of viral outbreaks in the next few years (16, 112).

plotype: unique nbination of rkers on a haploid nome

Sampling Design to Measure Virus Prevalence and Diversity

Most molecular epidemiology approaches require assessing and comparing the genetic diversity of viral populations sampled from different hosts at different spatial (possibly from the host up to the continent) and temporal (often multiple years up to several decades) scales. The diversity of viral populations can be estimated using different criteria according to the research questions, the type of genetic data obtained (i.e., molecular markers targeting one or several genomic regions, partial- or whole-genome sequences), and the analytical method chosen (44). Classical approaches aim at assessing the number and frequency of different haplotypes and the genetic distances between and within populations (92). Besides providing a direct estimation of genetic distances, partial- and whole-genome sequences also enable the quantification of the effects of different evolutionary forces, demographic inference, and the reconstruction of genealogical or phylogenetic relationships.

Biological and environmental variables (e.g., the life cycle of hosts and vectors, host/nonhost crop rotations, landscape structure, dispersal distances of vectors, etc.) that can impact epidemics should be considered when designing appropriate sampling schemes. The presence of symptoms can be used to target infected hosts, but when the study implies assessing relative virus prevalence (e.g., of different strains or in different hosts), plants should be collected regardless of symptom expression to avoid bias due to tolerance or asymptomatic stages of infection. Asymptomatic infections are not predominant in the cultivated compartment (except for tolerant host varieties), but they can represent an important proportion of plants in the wild compartment (122). Moreover, the type of plant material (e.g., leaves or stalks) collected during surveys has to be carefully considered, particularly for samples from the wild compartment, because virus concentration can be low and heterogeneous in infected plants (73). As most plant viral species are transmitted by vectors, sampling of plant material can be advantageously completed by collecting insects from which the virus can be extracted and sequenced (97). Indeed, the comparison between viral lineages found in insects and plants can provide information on the epidemiological cycle and dispersal of the virus. Sampling design and effort should also be adapted to the aim of the study. If intensive sampling of infected hosts is generally required to reconstruct transmission chains (see section How Did the Pathogen Spread within an Outbreak?), less intensive but well-balanced sampling (111) can be sufficient to describe viral diversity, compare population structures, and reconstruct dispersal and introduction events. Rarefaction curves and nonparametric richness estimators can be used to adapt sampling efforts and compare genetic diversities (53, 61). Moreover, hierarchical sampling and hierarchical partitioning of samples among variation factors may allow testing their effect on plant virus genetic differentiation (34). The following sections present the main approaches used to analyze viral sequence data to uncover the spatiotemporal dynamics of plant virus epidemics from the continental scale to the single outbreak.

HOW DO VIRUSES INVADE NEW TERRITORIES AND FURTHER SPREAD IN LANDSCAPES?

During the past few decades, trade globalization and greater human mobility have largely con-

are key to preventing new introductions and improving management strategies (47). Given the complex nature of spatiotemporal interactions across multiple scales, determining and managing the key processes driving pathogen dispersal are challenging. Although an appropriate scale for data collection and analysis should match the scale of the ecological phenomenon under question (87), multiscale information may be necessary to gain a more holistic view of transmission dynamics.

Genetics-based methods to study the spread of pathogens typically stem from the complementary fields of population genetics, landscape genetics, and phylogeography (8, 118, 144). These disciplines generally differ not only in terms of data and analyses commonly used but also by the timescale over which the data are informative. Indeed, population genetics and more recent landscape genetics approaches often use neutral genetic markers to infer population structure and contemporary gene flow at local or regional spatial scales (52). In contrast, phylogeography is mainly based on sequence data and aims to reconstruct long-term population dynamics (usually at an evolutionary timescale) such as dispersal events at continental or global scales (3). However, for measurably evolving pathogens, phylogeographic methods can also reveal patterns at spatial and temporal scales usually investigated using population and landscape genetics approaches (8).

Today, major advances in genetic and spatial data acquisition tools alongside subsequent analytical methods provide new opportunities to infer, within formal statistical frameworks, the processes at the origin of the spatial distributions of viruses and to quantitatively evaluate potential predictors of spread in complex environmental settings (8, 111) (see **Supplemental Table 1**).

Population genetics: study of genetic variation within and between populations

Landscape genetics: study of the geographical and environmental features that structure genetic variation (combines landscape ecology and population genetics)

Phylogeography: study of the spatiotemporal distribution of genetic lineages

Supplemental Material

Exploratory Approaches

Many of the population genetics methods developed to describe spatial genetic structures and estimate migration parameters require neutral and independent markers and/or rely on equilibrium assumptions (Hardy-Weinberg equilibrium, linkage equilibrium) that are rarely met for viruses (49, 51). Up to a decade ago, epidemiological studies focusing on describing and comparing the genetic structure of plant virus populations used molecular markers and/or partial genomic sequences to compute various indices of genetic diversity and measures of differentiation, e.g., mean pairwise nucleotide differences, number of polymorphic sites, pairwise genetic distances, and statistics of differentiation such as F_{ST} and K_{ST} (reviewed in 92). Further developments of versatile software such as Arlequin (37), which implements hierarchical analysis of molecular variance (AMOVA), have provided useful approaches to test for population subdivision according, for example, to geography or host plant species (34, 99, 108). The AMOVA design requires hypotheses on the genetic structure to be tested (e.g., samples are grouped according to geography or host plants). Clustering analyses that do not require such hypotheses on the structuring factors (i.e., a priori characterization of genetic groups) can thus be more appealing to analyze subdivisions in virus populations and identify immigrant genotypes (100, 108, 147). Most model-based clustering methods aim to maximize Hardy-Weinberg and linkage equilibria. Thus, when used on virus data, which are likely to deviate from these assumptions to various degrees, results should be carefully interpreted and completed with some kind of robustness analysis (145). Alternatively, although rarely used on plant viruses, exploratory methods that do not rely on genetic models constitute a valuable first step to assess both spatial and temporal structures of the genetic diversity within virus populations, as well as genotype flow between host populations (65). For example, spatial analysis of molecular variance (SAMOVA; 33) combined with Monmonier's maximum-difference algorithm (90) allowed the identification of both genetic subgroups and major disruptions of geno-

pairwise information (MAPI)], which provides spatial maps of mean genetic differentiation estimated between virus sequences (106). Multivariate analyses, such as the discriminant analysis of principal components (DAPC), can also be used on virus genetic data to analyze population substructure, perform probabilistic assignments (i.e., to detect immigrating genotypes), and identify the most important mutations involved in differentiation between genetic groups (65). De Bruyn et al. (22) combined spatial principal components analyses (sPCAs; 66) and DAPC (65) to study the spatial genetic structure of geminiviruses causing cassava mosaic disease in Madagascar. Such flexible exploratory methods are especially interesting, as they ease the processing of the increasingly large data sets generated using HTS technologies, and they are relatively easy to apply using packages [e.g., adegenet (66) and poppr (69)] of the R statistical software (113).

yesian inference: tistical inference thod in which ves' rule is used to ovide probability tributions of model rameters

RCA: time to the st recent common cestor

terochronous quence: dated netic sequences mpled at different nts in time

Reconstructing Invasion Pathways

Several phylogeographic frameworks are available to infer ancestral locations and spatiotemporal dynamics. These approaches mainly differ in their ability to handle spatial information (separately or simultaneously with the phylogenetic reconstruction) and to account for uncertainty (9). Recent approaches targeting viruses are based on the reconstruction of phylogenies in which temporal and spatial information are explicitly integrated to allow for the simultaneous inference of these processes (28, 77, 78). Moreover, statistical parametric or nonparametric models based on coalescent theory can be used to directly link patterns of genetic diversity to the demographic history of viral populations in a phylodynamic framework (50, 109). The popular programs BEAST and BEAST2 offer an integrative platform to perform these analyses (10, 30). Using a Bayesian inference framework for testing evolutionary hypotheses while accounting for phylogenetic uncertainty, they integrate numerous molecular clock models, discrete and continuous diffusion, and population dynamics. Although being largely validated and used on data sets of human and animal viruses, these methods have been applied only recently to RNA (21, 101, 114, 140, 148), ssDNA (1, 22, 23, 75, 80, 89, 137), and double-stranded (dsDNA) (147) plant viruses. Given their high potential, phylogeographic analyses are likely to keep gaining popularity in plant virus molecular epidemiology studies in the coming years. Here, we describe more precisely the data and methods required to address the questions relative to the geographical origin of a given viral lineage and the reconstruction of invasion pathways.

Phylogeographic analyses commonly use molecular clock models to represent the relationship between genetic distance and calendar time. Consequently, this can be used to estimate the ages of branching events, including the time to the most recent common ancestor (tMRCA) of lineages of interest. Many molecular clock models are available to accommodate for rate heterogeneities (58). Although initial strict clock models assumed a constant rate of molecular evolution throughout the tree, relaxed clocks now allow branch-specific evolutionary rates (27, 58). To calibrate such molecular clocks, studies targeting measurably evolving pathogens such as viruses use heterochronous sequence data. An evolutionary rate can thus be estimated, usually given as a number of nucleotide substitutions per site per year (29). Appropriate temporal sampling allowing the accumulation of genetic variation is recommended to enhance the temporal signal, whereas long-enough genomic sequences are necessary to increase the phylogenetic resolution (111, 130). The use of a herbarium or archeological specimens may allow for a greater temporal depth and thus more precise evolutionary estimates (82, 134). The presence of a temporal signal in the data set should always be tested; different methods and programs can be used (117), including linear regression of phylogenetic root-to-tip distance against sampling date and date-randomization tests (31). However,

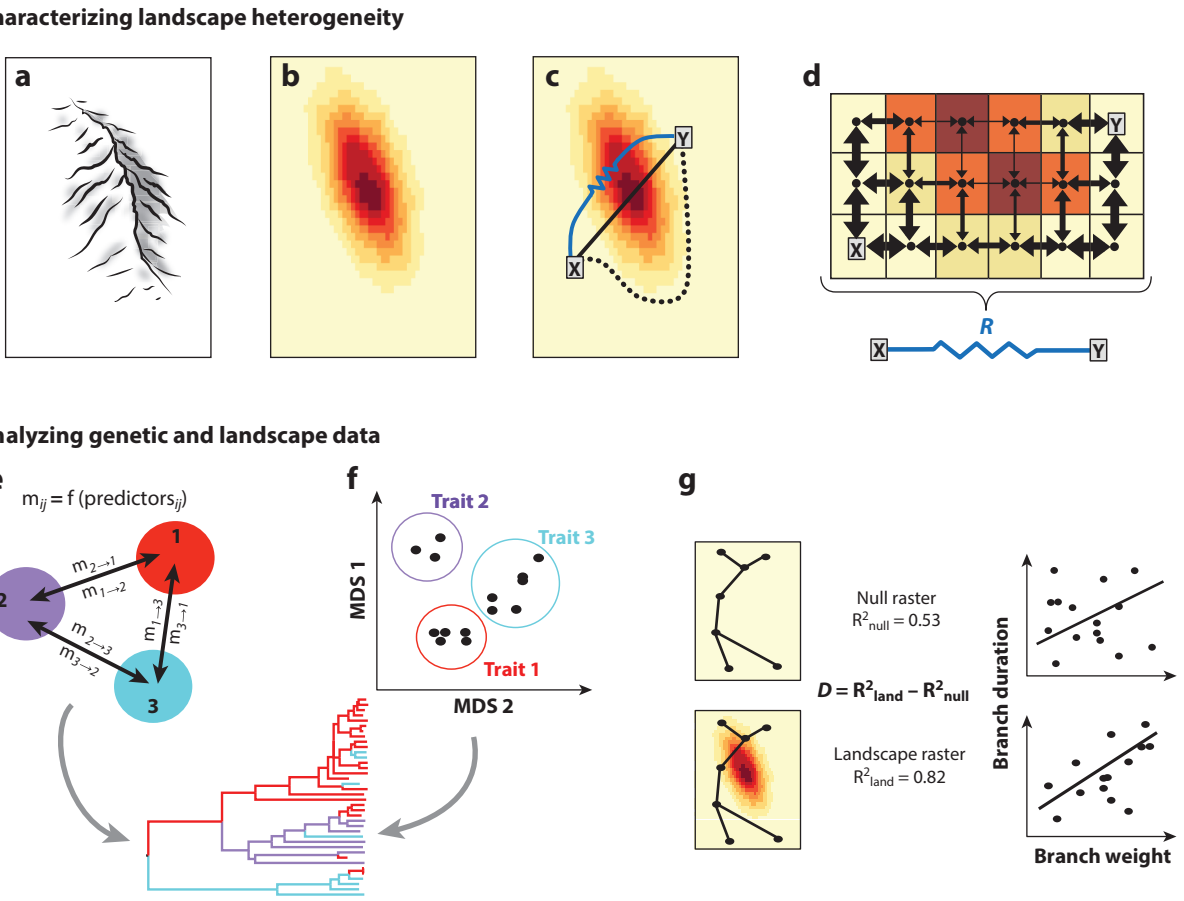
Besides inferring emergence or introduction dates of a given viral lineage in a new location (22, 114, 140), time-calibrated phylogenies can be used to evaluate the efficiency and timeliness of an epidemiological surveillance system by comparing the estimated MRCA ages with the dates of discovery of a given outbreak (114).

Using Bayesian skyline plots and other coalescent-based methods, it is also possible to estimate effective population sizes through time (59) and detect population bottlenecks and subsequent expansion during invasions (1). However, these methods often assume a single well-mixed population, an assumption that is only rarely met as virus populations may be highly structured (spatially and/or by host). Because violation of this assumption can lead to misleading inference (57), a cautious interpretation is required.

Popular phylogeographic methods for reconstructing virus spatial spread from genetic data treat the geographical locations assigned to each sequence as discrete traits (e.g., for viruses sampled in cities or countries) or continuous traits (e.g., samples with latitude-longitude coordinates). Movements are represented as changes in traits along sampled lineages (77, 78). In discrete phylogeography, stochastic diffusion processes are modeled using a continuous-time Markov chain (CTMC), where the transitions between spatial locations in the phylogeny are either symmetrical or asymmetrical to provide a more realistic description of the spatial dynamics. The number of transitions between spatial locations can be inferred, providing valuable information when one is interested in the number and direction of migration events in source-sink dynamics (77). The most significant dispersal pathways can then be identified using Bayesian stochastic search variable selection (BSSVS) (77). The continuous diffusion model relies on relaxed random walk models (Brownian motion process) to explore two-dimensional space and can yield more realistic reconstructions of the dispersal process in a given landscape (78). These models enable the computation of various statistics to quantify the spatial dynamics of an epidemic, such as the diffusion coefficient D that measures spatial velocity (50). Both approaches have been used to reconstruct the invasion pathways of various plant viruses at global and regional scales (75, 137, 140). It is important to emphasize that the accuracy of these methods in estimating the location of ancestors and capturing dispersal patterns is directly linked to the quality of sampling (142). Estimation of ancestral locations might be highly uncertain if an inferred ancestor is only distantly related (spatially) to the sampled cases. Moreover, if samples from key locations or regions are absent or rare, then virus movements will be underestimated and the inferred locations of ancestors may be biased toward over-represented locations. Although these methods are particularly efficient from a computational perspective, a recent study has provided evidence that they may suffer from various biases and statistical inefficiency (81). A new model-based approach, Bayesian structured coalescent approximation (BASTA), has been developed (81) and is implemented in BEAST2. This method is based on the structured coalescent, a statistical model that explicitly accounts for migration effects on the shape and branch lengths of the genealogy.

Phylogeographic approaches have benefited from a rich development of statistical inference tools. The successful application of these methods to plant viruses depends on the assembly of large collections of dated and georeferenced plant virus sequences, as are already available for numerous human/animal viruses. However, despite the huge number of plant virus sequences in molecular databases, temporal and spatial information associated with the submitted sequences are often lacking for the sequences older than the past decade (22).

Integrating Landscape Heterogeneity



gure 2

lating genetic patterns to landscape characteristics. (*Top*) Characterizing landscape heterogeneity using resistance surfaces: (*a*) an mple in which mountains are dispersal barriers; (*b*) rasterized resistance surface in which darker cells indicate higher resistance to persal; (*c*) possible distance metrics between focal points X and Y: Euclidean distance (*solid black line*), least-cost path (*dotted line*), and esistance distance *R* across multiple pathways (*blue line*); (*d*) details of the calculation of *R* between X and Y with heavier arrows icating increasingly facilitated virus flow. (*Bottom*) Methods to incorporate landscape heterogeneity into (*e*, *f*) discrete and continuous phylogeographic analyses: (*e*) generalized linear model (GLM) extension of the diffusion model in which diffusion rates $_{ij}$) between demes (*colored circles*) are a function of a set of explanatory variables tested using Bayesian model averaging (76); pathogen spread relative to a null model, quantified after assignment of phylogeographic traits through landscape-informed stering (here, three clusters) of pathogen locations via multidimensional scaling (14); (*g*) assessment of the effect *D* of an vironmental variable through an increase in the association between lineage movements (branch duration, inferred from a dated ylogeny) and the associated resistance weights, relative to a null model (24).

host distributions, vector movements, and the transfer of infected plant material (87). With the modern capacity to produce and analyze genetic data, new opportunities have arisen to use molecular epidemiology analyses to gain a detailed quantitative understanding of these interactions (**Figure 2**).

ntel test: tistical test of the rrelation between

The effect of landscape heterogeneity on transmission is difficult to quantify given the un-

structure and landscape variables, particularly for barrier effects (65, 145). However, the suitability of such tests to detect landscape effects has come into question (5). Simulation models offer an alternative quantitative approach, an example being the identification of a 50% permeability of rivers to raccoon movements and thus to rabies virus in North America (116). Furthermore, with the accessibility of finely resolved genetic data, the analyses now exploit sophisticated Bayesian phylogeographic frameworks (see section Reconstructing Invasion Pathways) to measure variation in dispersal among landscape components (77, 78). However, it is only recently that a statistical framework, utilizing a generalized linear model (GLM) parameterization, has become available to simultaneously test and quantify the effects of potential predictors on dispersal patterns (41, 76). This has identified the role of human and animal transportation networks on influenza spread (76, 96) and has recently been applied for the first time to a plant virus (140).

An increasingly popular method to account for landscape heterogeneity is to represent variables in terms of their cost or "resistance" to dispersal, based on the "isolation by resistance" (IBR) concept (86). Computer programs such as Circuitscape (131) and the R package gdistance (141) provide various distance-based metrics to measure dispersal potential across different landscape resistance surfaces. Synthesizing landscape information in this way provides a simple input for modeling approaches, as exemplified by the use of a resistance surface based on rice production statistics to quantify the impact of crop intensification on *Rice yellow mottle virus* (RYMV) spread (140). Several other IBR approaches have emerged, including a method to compare phylogenetic reconstructions of dispersal with landscape variation (24). This work provides a framework to extract information from the branches of spatiotemporally referenced phylogenies to perform tests of correlations with landscape characteristics, employing a randomization procedure to determine significance. The framework offers some flexibility in terms of the method and software used to build phylogenies and is less computationally demanding than the GLM approach (76). However, reliance on linear regression to identify correlations may not capture more complex relationships (e.g., quadratic and thresholds) between dispersal and landscape features. Alternatively, Brunker et al. (15) used resistance distances to rescale spatial information and assign phylogenetic traits as a means to directly inform phylogeographic reconstructions, simultaneously providing a means to test the effect of landscape features on epidemics across multiple spatial scales.

These promising ways to integrate landscape heterogeneity are still under development. Potential future improvements include the use of nonlinear multivariate approaches, development of simulation models to assess the relative sensitivity of the various methods to detect barrier effects, and exploration and integration of the temporal dynamics of landscape heterogeneity. More generally, the application of phylodynamic techniques to identify important sources of variation in dispersal is a potentially fruitful endeavor for the next few years (4).

HOW DID THE PATHOGEN SPREAD WITHIN AN OUTBREAK?

Another field which has been developing at an ever-increasing pace during the past decade is the reconstruction of the transmission links within outbreaks. Inferring the history of transmission events within a host population can highlight key drivers of transmission, provide refined estimates of epidemiological parameters and point out risk factors related to vectors, reservoirs, and landscape components (103). Ultimately, such studies can help build epidemiological projections, design control strategies, and deliver scientific advice to governmental agencies. However, inferring "who infected whom" in outbreaks of infectious diseases remains a challenging task.

IR: susceptible, posed, infectious, moved

outbreaks (19, 56, 64, 67, 88, 91). Data can be epidemiological records, such as the spatiotemporal locations of infected hosts, or genetic information on evolutionary relationships between virus genomes sampled from the hosts. In particular, when enough mutations can be observed during an outbreak, the joint analysis of epidemiological and genetic data can provide valuable insights into transmission dynamics. Several approaches currently under development aim to appropriately combine these data. Specifically, we highlight the reconstruction of transmission trees and the estimation of epidemiological parameters. We present below the existing approaches that address this question to determine how measurably evolving pathogens spread within a host population.

Model-Based Inference of Transmission Trees

Transmission trees have been inferred using various modeling approaches. Some of these approaches are intrinsically based on phylogenetic models in which epidemiological information is introduced. Others have started with epidemiological models enriched with genetic information.

The first approach is based on phylogenetic and coalescent models (117). Here, spatial or temporal information is added to the process of phylogenetic reconstruction. Such methods relate the demography of the pathogen to its evolution and may incorporate a diffusion model to account for the movement of the pathogen over geographical space (50, 54, 77, 78, 110, 115, 132) (see section Reconstructing Invasion Pathways). This approach is relatively robust to the intensity of epidemiological sampling. However, because the underlying models do not have an explicit epidemiological formulation (except for some models; 54), the inferred parameters cannot be easily related to the epidemiological processes. Jombart et al. (67) also pointed out that a phylogenetic approach attempts to infer hypothetical common ancestors among the sampled genomes and thus may not be appropriate for a set of genomes containing both ancestors and their descendants. Indeed, phylogenetic methods consider that sampled strains are all tips of an unknown genealogy, making it impossible for a sampled strain to be (directly or indirectly) the ancestor of another sampled strain (121), an issue that is often encountered for densely sampled outbreaks. However, recent works have addressed this issue by developing an algorithm to infer phylogenetic trees in which sampled sequences can be direct ancestors of other sampled sequences (45) or by employing an individual-based disease transmission model and a coalescent process taking place within each host (54).

The second approach uses spatial epidemiological models of transmission and models of genetic drift to directly reconstruct the transmission tree reflecting "who infected whom." This approach is generally based on stochastic and spatiotemporal SEIR (susceptible, exposed, infectious, removed) models explicitly representing successive states of host individuals (64, 88, 91, 135, 149, 150) to recognize the host population structure and epidemiological processes governing host-pathogen interactions. A model for the spread of the pathogen in the population and a model for the accumulation of point mutations over time are often used to calculate the probability that the genetic sequence transmitted from case A to case B could have mutated into the sequences sampled from the two cases in the duration between transmission and sampling (**Figure 3**) (88). More specifically, a first study identified a large set of transmission trees consistent with the available genetic data and then ranked these trees with respect to a likelihood computed from temporal data, revealing the most likely set of transmission trees (19). In later works, the likelihood of the transmission tree J given temporal (T), spatial (X), and genetic (G) data was approximated by the product of three independent likelihoods (150): $L(J|T,X,G) = L1(J|T) \times L2(J|X) \times L3(J|G)$.

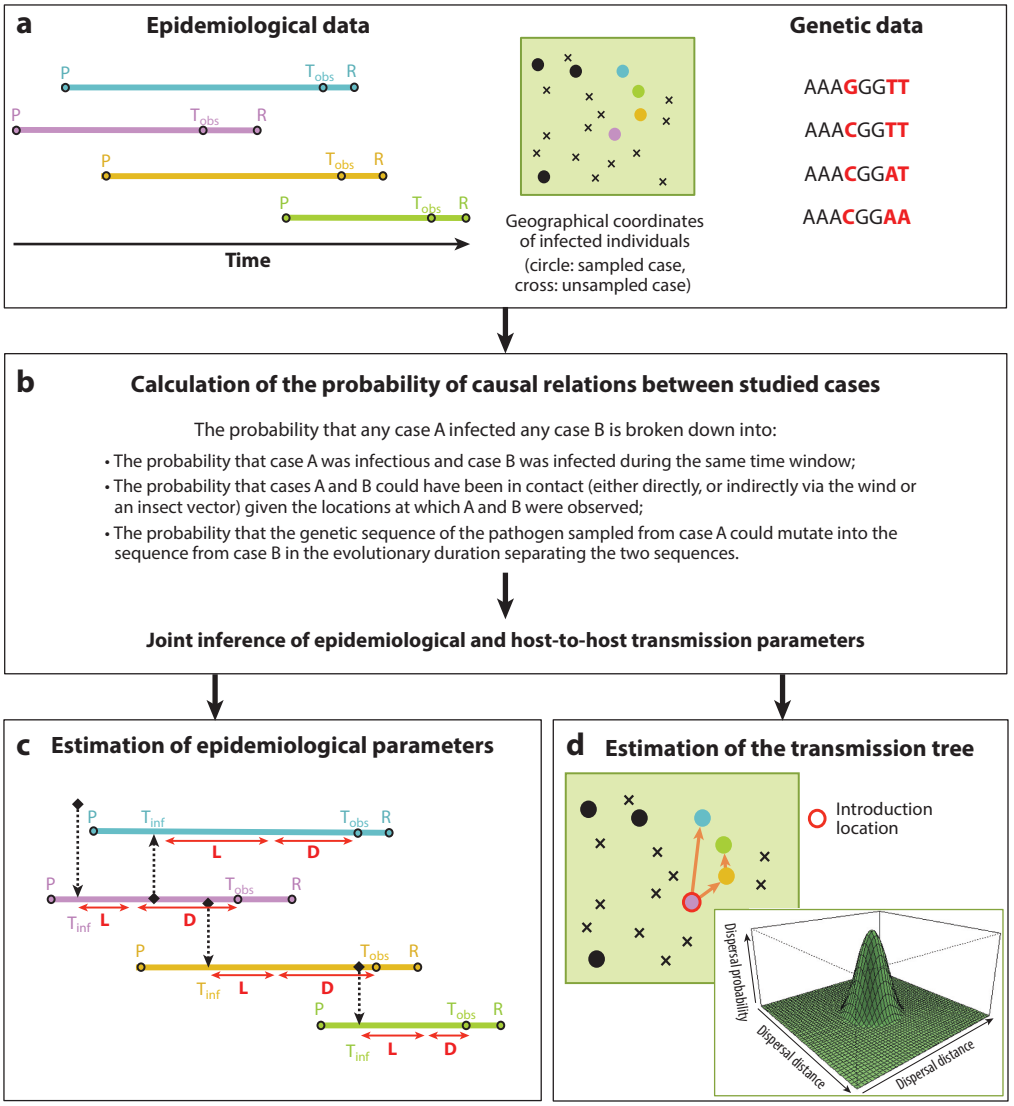


Figure 3

Inference of epidemiological parameters and transmission trees in a landscape. (*a*) Inputs of space-time-genetic SEIR (susceptible, exposed, infectious, removed) models are epidemiological data (e.g., plantation dates, dates of symptom detection, removal dates, locations of infected trees) and genetic data (genetic sequences). (*b*) Based on these inputs and the calculation of probabilities of causal relations between studied cases, the space-time-genetic SEIR models and the accompanying estimation algorithms provide joint estimations of transmission links and epidemiological parameters. (*c*) Estimation of epidemiological parameters defining the duration of the latent period, the duration of the infectious period before detection, and the time of infection. (*d*) Estimation of the transmission tree ("who infected whom") thanks to spatial parameters (introduction location and dispersal function). Abbreviations: P, planting date; T_{obs}, time of symptom detection; R, removal date; L, duration of the latency period; D, duration of the infectious period before detection; T_{inf}, time of infection.

into account the inherent dependence between temporal, spatial, and genetic data and calculated the likelihood of transmission trees (88, 91). These methods have been very valuable in unraveling transmission pathways during outbreaks. However, they either avoid explicit inference of the unobserved pathogen sequences transmitted during infection (64, 88, 91, 149, 150) or use approximate Bayesian inference to account for these sequences (135). Such approximate approaches greatly reduce the computational challenges associated with inferring the unobserved transmitted sequences and facilitate statistical inference, particularly when the transmission tree is of primary interest. Instead of using approximations, Lau et al. (74) considered a Bayesian framework that simultaneously and explicitly infers the transmission tree and the transmitted pathogen sequences. This approach facilitates the use of realistic likelihood functions and allows the systematic joint inference of epidemio-evolutionary processes from partially observed outbreaks.

Compared with phylogenetic approaches, space-time-genetic SEIR approaches generally require a moderate to high proportion of infected hosts for accurate inference. This is particularly true for early studies assuming that sampled cases were directly related through transmission (19, 67, 91). More recent works accommodate the inherent complexities of polyphyletic and partially sampled outbreaks (64, 74, 88). Thus, space-time-genetic SEIR models and associated estimation algorithms yield increasingly satisfactory reconstructions of transmission trees. Such models can nevertheless result in misleading interpretations of transmission dynamics if they use a single sequence from each infected case in situations of mixed infections (25, 146), although works in progress tend to overcome this problem (25). Phylogenetic and space-time-genetic SEIR approaches have recently begun to merge by combining features of phylogenetic and transmission tree approaches to reconstruct partially observed transmission networks (35, 39, 54, 70). In addition to their ability to infer a transmission tree, most of the approaches presented in this section provide estimates of other important epidemiological parameters.

sic reproduction mber: number of ividuals infected ring the infectious iod of an infected ividual placed into uninfected pulation

tent period: delay ween inoculation of athogen and the ginning of the host ectious period

cubation period: ay between culation of a hogen and the set of the first nptom of the ected host

spersal function: obability tribution of the ival location of an ividual (e.g., an ect vector) persed from a given rting location

Estimation of Epidemiological Parameters

For several decades now, S(E)I(R) models (without genetics) have been fitted to data on the number of cases through time to estimate epidemiological parameters such as the basic reproduction number R_0 or thresholds of vaccination coverage (2, 13, 26). The emergence of epidemio-evolutionary approaches based on S(E)I(R) models should lead to finer estimations by exploiting information brought by genetic data. These approaches often explicitly include (and allow inference about) parameters related to infection strength (and sometimes its heterogeneity among hosts), the latent period, the incubation period, the dispersal function (which partly determines the speed and spatial extent of disease spread), and the substitution rate (e.g., 136). In addition, these approaches allow the calculation of the effective reproduction number over time or the total infected population over a given spatiotemporal window.

The estimation algorithms have generally been developed within a Bayesian framework (25, 54, 88, 91, 135, 150) to incorporate prior knowledge about the parameters and to benefit from techniques allowing the inference of hidden variables, such as infection times and transmitted pathogen sequences when transmission trees and epidemiological parameters are estimated jointly. It is especially interesting to incorporate prior knowledge about influential epidemiological or evolutionary parameters on which the data used for model fitting bring little information. Finally, the output of Bayesian estimation algorithms is a sample of the joint posterior distribution of the parameters. Such samples can be used to provide not only point estimates of parameters but also uncertainties in, and dependencies between, estimates. Such models have not yet been used to

SUMMARY POINTS

1. The high substitution rate of viruses implies that evolutionary and epidemiological processes are observable at the same timescales, and that viral genomes are scattered with imprints that can be used to infer virus dynamics in landscapes through space and time.
2. Continual advances in virus characterization methods have vastly expanded our knowledge of the existing virus species and of their intraspecific diversity.
3. Appropriate sampling schemes are required to prevent bias when studying how the diversity of viral populations is structured by biological and environmental variables.
4. Because viruses are clonal, assumption-free exploratory analyses are more appropriate than classical population genetics approaches to describe the spatial structure of viral diversity.
5. Phylogeographic models enable inference of invasion pathways over large areas based on the geographical coordinates of dated sequences.
6. New approaches combining landscape genetics and phylogeography provide a means to test the impact of landscape configuration and composition on virus spatiotemporal dynamics.
7. Recent phylodynamic and space-time-genetic SEIR models can be used to infer transmission trees and other key epidemiological and evolutionary parameters, based on virus sequences from intensively sampled outbreaks.

FUTURE ISSUES

1. Characterizing plant virus diversity at the ecosystem scale is still needed to better understand the spatiotemporal dynamics of plant viruses in cultivated and natural areas.
2. Molecular epidemiology studies should considerably benefit from advances in real-time, portable genome sequencing and high-throughput sequencing to produce long reads and high-fidelity sequences.
3. In parallel, more powerful estimation approaches will be welcome to exploit the ever-increasing number of sequences representing virus diversity both between and within hosts. Progress could take the form of faster algorithms using robust approximations, more flexible models, and complex models of the various processes underlying within- and between-host dynamics.
4. Molecular epidemiology studies generally focus on the neutral genetic diversity of non-recombinant sequences. Integrating information brought by recombinant sequences and relating genetic changes under selection with epidemiological changes are promising methodological challenges.
5. Better characterizing the various landscape types and host characteristics and estimating their impact are both challenging and important for the understanding of plant virus spread.
6. High-resolution inference of "who infected whom" based on sequencing data is a promis-

7. Development of new frameworks to enable improved integration of data and models may lead to real-time characterization and prediction of outbreaks. This might take the form of streamlined pipelines from sample collection to sequencing, from bioinformatics analysis through updated phylogenies to estimation of parameters feeding disease management models and, finally, feedback procedures toward disease control organizations.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work has been supported by an INRA/ANSES PhD grant (C.P.), the EU FP7-PEOPLE program (PIOF-GA-2013-622571) (P.R.), the ANR-funded SMITID project (ANR-16-CE35-0006), and Agropolis Fondation (E-SPACE project).

LITERATURE CITED

1. Almeida RP, Bennett GM, Anhalt MD, Tsai CW, O'Grady P. 2009. Spread of an introduced vector-borne banana virus in Hawaii. *Mol. Ecol.* 18(1):136–46
2. Anderson RM, May RM. 1992. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford Univ. Press
3. Avise JC. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard Univ. Press
4. Baele G, Suchard MA, Rambaut A, Lemey P. 2017. Emerging concepts of data integration in pathogen phylodynamics. *Syst. Biol.* 66(1):e47–65
5. Balkenhol N, Waits LP, Dezzani RJ. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32(5):818–30
6. Bebber DP, Holmes T, Gurr SJ. 2014. The global spread of crop pests and pathogens. *Glob. Ecol. Biogeogr.* 23(12):1398–407
7. **Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. 2015. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* 30(6):306–13**
8. Biek R, Real LA. 2010. The landscape genetics of infectious disease emergence and spread. *Mol. Ecol.* 19(17):3515–31
9. Bloomquist EW, Lemey P, Suchard MA. 2010. Three roads diverged? Routes to phylogeographic inference. *Trends Ecol. Evol.* 25(11):626–32
10. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* 10(4):e1003537
11. Bousalem M, Dallot S, Fuji S, Natsuaki KT. 2003. Origin, world-wide dispersion, bio-geographical diversification, radiation and recombination: an evolutionary history of *Yam mild mosaic virus* (YMMV). *Infect. Genet. Evol.* 3(3):189–206
12. Bousalem M, Dallot S, Guyader S. 2000. The use of phylogenetic data to develop molecular tools for the detection and genotyping of *Yam mosaic virus*. Potential application in molecular epidemiology. *J. Virol. Methods* 90(1):25–36
13. Britton T, Giardina F. 2016. Introduction to statistical inference for infectious diseases. *J. Société Fr.*

This review highlights landscape genetic principles relevant to infection dynamics.

15. Brunker K, Marston DA, Horton DL, Cleaveland S, Fooks AR, et al. 2015. Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. *Virus Evol*. 1(1):vev011
16. Bull RA, Eltahla AA, Rodrigo C, Koekkoek SM, Walker M, et al. 2016. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC Genom*. 17:247
17. Candresse T, Filloux D, Muhire B, Julian C, Galzi S, et al. 2014. Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLOS ONE* 9(7):e102945
18. Clark MF, Adams AN. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol.* 34(3):475–83
19. **Cottam EM, Thébaud G, Wadsworth J, Gloster J, Mansley L, et al. 2008. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. R. Soc. B* 275(1637):887–95**
20. Cuevas JM, Willemsen A, Hillung J, Zwart MP, Elena SF. 2015. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. *Mol. Biol. Evol.* 32(5):1132–47
21. Davino S, Willemsen A, Panno S, Davino M, Catara A, et al. 2013. Emergence and phylodynamics of *Citrus tristeza virus* in Sicily, Italy. *PLOS ONE* 8(6):e66700
22. De Bruyn A, Harimalala M, Zinga I, Mabvakure BM, Hoareau M, et al. 2016. Divergent evolutionary and epidemiological dynamics of cassava mosaic geminiviruses in Madagascar. *BMC Evol. Biol.* 16:182
23. De Bruyn A, Villemot J, Lefeuvre P, Villar E, Hoareau M, et al. 2012. East African cassava mosaic-like viruses from Africa to Indian Ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evol. Biol.* 12:228
24. Dellicour S, Rose R, Pybus OG. 2016. Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinform*. 17:82
25. Didelot X, Gardy J, Colijn C. 2014. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol. Biol. Evol.* 31(7):1869–79
26. Diekmann O, Heesterbeek H, Britton T. 2012. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton, NJ: Princeton Univ. Press
27. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLOS Biol*. 4(5):e88
28. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161(3):1307–20
29. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003. Measurably evolving populations. *Trends Ecol. Evol.* 18(9):481–88
30. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29(8):1969–73
31. Duchêne S, Duchêne D, Holmes EC, Ho SYW. 2015. The performance of the date-randomisation test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* 32(7):1895–906
32. **Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9(4):267–76**
33. Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11(12):2571–81
34. D'Urso F, Sambade A, Moya A, Guerri J, Moreno P. 2003. Variation of haplotype distributions of two genomic regions of *Citrus tristeza virus* populations from eastern Spain. *Mol. Ecol.* 12(2):517–26
35. Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, et al. 2016. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *eLife* 5:e16644
36. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, et al. 2011. The evolutionary genetics of emerging plant RNA viruses. *Mol. Plant-Microbe Interact.* 24(3):287–93
37. Excoffier L, Laval G, Schneider S. 2007. An integrated software package for population genetics data analysis. *Evol. Bioinform.* 1:47–50
38. Fabre F, Moury B, Johansen EI, Simon V, Jacquemond M, Senoussi R. 2014. Narrow bottlenecks affect

19. This early approach infers transmission trees using epidemiological data and virus genomes.

32. This review clearly explains the factors affecting virus mutation and substitution rates.

39. Famulare M, Hu H. 2015. Extracting transmission networks from phylogeographic data for epidemic and endemic diseases: Ebola virus in Sierra Leone, 2009 H1N1 pandemic influenza and polio in Nigeria. *Int. Health* 7(2):130–38
40. Fargette D, Konaté G, Fauquet C, Muller E, Peterschmitt M, Thresh JM. 2006. Molecular ecology and emergence of tropical plant viruses. *Annu. Rev. Phytopathol.* 44:235–60
41. Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P. 2013. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philos. Trans. R. Soc. B* 368(1614):20120196
42. Fraile A, Malpica JM, Aranda MA, Rodríguez-Cerezo E, García-Arenal F. 1996. Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant *Nicotiana glauca*. *Virology* 223(1):148–55
43. Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. 2015. Eight challenges in phylodynamic inference. *Epidemics* 10:88–92
44. García-Arenal F, Fraile A, Malpica JM. 2001. Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39:157–86
45. Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Comput. Biol.* 10(12):e1003919
46. Gibbs AJ, Fargette D, García-Arenal F, Gibbs MJ. 2010. Time—the emerging dimension of plant virus studies. *J. Gen. Virol.* 91(1):13–22
47. Gilligan CA. 2008. Sustainable agriculture and plant diseases: an epidemiological perspective. *Philos. Trans. R. Soc. B* 363(1492):741–59
48. Glais L, Kerlan C, Tribodet M, Marie-Jeanne Tordo V, Robaglia C, Astier-Manifacier S. 1996. Molecular characterization of potato virus Y^N isolates by PCR-RFLP. *Eur. J. Plant Pathol.* 102(7):655–62
49. Goss EM. 2015. Genome-enabled analysis of plant-pathogen migration. *Annu. Rev. Phytopathol.* 53:121–35
50. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, et al. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303(5656):327–32
51. Grünwald NJ, Goss EM. 2011. Evolution and population genetics of exotic and re-emerging pathogens: novel tools and approaches. *Annu. Rev. Phytopathol.* 49:249–67
52. Guillot G, Leblois R, Coulon A, Frantz AC. 2009. Statistical methods in spatial genetics. *Mol. Ecol.* 18(23):4734–56
53. Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. 2013. Robust estimation of microbial diversity in theory and in practice. *ISME J.* 7(6):1092–101
54. Hall M, Woolhouse M, Rambaut A. 2015. Epidemic reconstruction in a phylogenetics framework: transmission trees as partitions of the node set. *PLOS Comput. Biol.* 11(12):e1004613
55. Hampson K, Dushoff J, Cleaveland S, Haydon DT, Kaare M, et al. 2009. Transmission dynamics and prospects for the elimination of canine rabies. *PLOS Biol.* 7(3):e1000053
56. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, et al. 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. R. Soc. B* 270(1511):121–27
57. Heller R, Chikhi L, Siegismund HR. 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLOS ONE* 8(5):e62992
58. Ho SYW, Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Mol. Ecol.* 23(24):5947–65
59. Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11(3):423–34
60. Holmes EC. 2009. The evolutionary genetics of emerging viruses. *Annu. Rev. Ecol. Evol. Syst.* 40:353–72
61. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67(10):4399–406
62. Jeger MJ, Seal SE, Van den Bosch F. 2006. Evolutionary epidemiology of plant virus disease. *Adv. Virus Res.* 67:163–203

This review sents the principles he phylodynamics nework, which ates phylogenies and demiology.

This review sents the different lecular clock models.

64. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. 2014. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLOS Comput. Biol.* 10(1):e1003457
65. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 11:94
66. Jombart T, Devillard S, Dufour A-B, Pontier D. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101(1):92–103
67. Jombart T, Eggo RM, Dodd PJ, Balloux F. 2011. Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity* 106(2):383–90
68. Jridi C, Martin J-F, Marie-Jeanne V, Labonne G, Blanc S. 2006. Distinct viral populations differentiate and evolve independently in a single perennial host plant. *J. Virol.* 80(5):2349–57
69. Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208
70. Kenah E, Britton T, Halloran ME, Longini IM Jr. 2016. Molecular infectious disease epidemiology: survival analysis and algorithms linking phylogenies to transmission trees. *PLOS Comput. Biol.* 12(4):e1004869
71. Klinkowski M. 1970. Catastrophic plant diseases. *Annu. Rev. Phytopathol.* 8:37–60
72. Kraberger S, Harkins GW, Kumari SG, Thomas JE, Schwinghamer MW, et al. 2013. Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East. *Virology* 444(1–2):282–91
73. Lacroix C, Renner K, Cole E, Seabloom EW, Borer ET, Malmstrom CM. 2016. Methodological guidelines for accurate detection of viruses in wild plant species. *Appl. Environ. Microbiol.* 82(6):1966–75
74. Lau MSY, Marion G, Streftaris G, Gibson G. 2015. A systematic Bayesian integration of epidemiological and genetic data. *PLOS Comput. Biol.* 11(11):e1004633
75. Lefeuvre P, Martin DP, Harkins G, Lemey P, Gray AJA, et al. 2010. The spread of *Tomato yellow leaf curl virus* from the Middle East to the world. *PLOS Pathog*. 6(10):e1001164
76. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, et al. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLOS Pathog*. 10(2):e1003932
77. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* 5(9):e1000520
78. Lemey P, Rambaut A, Welch JJ, Suchard MA. 2010. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* 27(8):1877–85
79. Logan G, Freimanis GL, King DJ, Valdazo-González B, Bachanek-Bankowska K, et al. 2014. A universal protocol to generate consensus level genome sequences for foot-and-mouth disease virus and other positive-sense polyadenylated RNA viruses using the Illumina MiSeq. *BMC Genom*. 15:828
80. Mabvakure B, Martin DP, Kraberger S, Cloete L, van Brunschot S, et al. 2016. Ongoing geographical spread of *Tomato yellow leaf curl virus*. *Virology* 498:257–64
81. Maio ND, Wu CH, O'Reilly KM, Wilson D. 2015. New routes to phylogeography: a Bayesian structured coalescent approximation. *PLOS Genet*. 11(8):e1005421
82. Malmstrom CM, Shu R, Linton EW, Newton LA, Cook MA. 2007. Barley yellow dwarf viruses (BYDVS) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *J. Ecol.* 95(6):1153–66
83. Marais A, Faure C, Couture C, Bergey B, Gentit P, Candresse T. 2014. Characterization by deep sequencing of divergent *Plum bark necrosis stem pitting–associated virus* (PBNSPaV) isolates and development of a broad-spectrum PBNSPaV detection assay. *Phytopathology* 104(6):660–66
84. Martin DP, Biagini P, Lefeuvre P, Golden M, Roumagnac P, Varsani A. 2011. Recombination in eukaryotic single stranded DNA viruses. *Viruses* 3(9):1699–738
85. Martin DP, Lemey P, Posada D. 2011. Analysing recombination in nucleotide sequences. *Mol. Ecol. Resour.* 11(6):943–55
86. McRae BH. 2006. Isolation by resistance. *Evolution* 60(8):1551–61

67. This study presents the first R package enabling to reconstruct transmission graphs from genetic data.

77. This article presents a Bayesian framework for inference, visualization, and hypothesis testing of phylogeographic history.

81. This paper introduces BASTA, a phylogeographic model combining accurate coalescence methods and computational efficiency.

This article scribes a framework reconstructing nsmission trees using netic, temporal, and tial information.

88. **Mollentze N, Nel LH, Townsend S, Roux K, Hampson K, et al. 2014. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. *Proc. R. Soc. B* 281(1782):20133251**
89. Monjane AL, Harkins GW, Martin DP, Lemey P, Lefeuvre P, et al. 2011. Reconstructing the history of Maize streak virus strain A dispersal to reveal diversification hot spots and its origin in southern Africa. *J. Virol.* 85(18):9623–36
90. Monmonier MS. 1973. Maximum-difference barriers: an alternative numerical regionalization method. *Geogr. Anal.* 5(3):245–61
91. Morelli MJ, Thébaud G, Chadœuf J, King DP, Haydon DT, Soubeyrand S. 2012. A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. *PLOS Comput. Biol.* 8(11):e1002768
92. Moury B, Desbiez C, Jacquemond M, Lecoq H. 2006. Genetic diversity of plant virus populations: towards hypothesis testing in molecular epidemiology. *Adv. Virus Res.* 67:49–87
93. Moury B, Fabre F, Senoussi R. 2007. Estimation of the number of virus particles transmitted by an insect vector. *PNAS* 104(45):17891–96
94. Mullan B, Sheehy P, Shanahan F, Fanning L. 2004. Do Taq-generated RT-PCR products from RNA viruses accurately reflect viral genetic heterogeneity? *J. Viral Hepat.* 11(2):108–14
95. Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, et al. 2016. The effect of genetic structure on molecular dating and tests for temporal signal. *Methods Ecol. Evol.* 7(1):80–89
96. Nelson MI, Viboud C, Vincent AL, Culhane MR, Detmer SE, et al. 2015. Global migration of influenza A viruses in swine. *Nat. Commun.* 6:6696
97. Ng TFF, Duffy S, Polston JE, Bixby E, Vallad GE, Breitbart M. 2011. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLOS ONE* 6(4):e19050
98. O'Fallon BD. 2013. ACG: rapid inference of population history from recombining nucleotide sequences. *BMC Bioinform.* 14:40
99. Ohshima K, Akaishi S, Kajiyama H, Koga R, Gibbs AJ. 2010. Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. *J. Gen. Virol.* 91(3):788–801
100. Ohshima K, Matsumoto K, Yasaka R, Nishiyama M, Soejima K, et al. 2016. Temporal analysis of reassortment and molecular evolution of *Cucumber mosaic virus*: extra clues from its segmented genome. *Virology* 487:188–97
101. Olarte Castillo XA, Fermin G, Tabima J, Rojas Y, Tennant PF, et al. 2011. Phylogeography and molecular epidemiology of *Papaya ringspot virus*. *Virus Res.* 159(2):132–40
102. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *PNAS* 86(8):2766–70
103. Ostfeld RS, Glass GE, Keesing F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol. Evol.* 20(6):328–36
104. Palanga E, Filloux D, Martin DP, Fernandez E, Gargani D, et al. 2016. Metagenomic-based screening and molecular characterization of cowpea-infecting viruses in Burkina Faso. *PLOS ONE* 11(10):e0165188
105. Parker IM, Gilbert GS. 2004. The evolutionary ecology of novel plant-pathogen interactions. *Annu. Rev. Ecol. Evol. Syst.* 35:675–700
106. Piry S, Chapuis M-P, Gauffre B, Papaïx J, Cruaud A, Berthier K. 2016. Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods Ecol. Evol.* 7:1463–75
107. Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *PNAS* 98(24):13757–62
108. Prasanna H, Sinha DP, Verma A, Singh M, Singh B, et al. 2010. The population genomics of begomoviruses: global scale population structure and gene flow. *Virol. J.* 7:220
109. Pybus OG, Rambaut A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10(8):540–50
110. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, et al. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *PNAS* 109(37):15066–71

112. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530(7589):228–32
113. R Dev. Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Found. Stat. Comput.
114. Rakotomalala M, Pinel-Galzi A, Mpunami A, Randrianasolo A, Ramavovololona P, et al. 2013. *Rice yellow mottle virus* in Madagascar and in the Zanzibar Archipelago; island systems and evolutionary time scale to study virus emergence. *Virus Res*. 171(1):71–79
115. Rasmussen DA, Ratmann O, Koelle K. 2011. Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Comput. Biol.* 7(8):e1002136
116. Rees EE, Pond BA, Cullingham CI, Tinline R, Ball D, et al. 2008. Assessing a landscape barrier using genetic simulation modelling: implications for raccoon rabies management. *Prev. Vet. Med.* 86(1–2):107–23
117. Rieux A, Balloux F. 2016. Inferences from tip-calibrated phylogenies: a review and a practical guide. *Mol. Ecol.* 25(9):1911–24
118. Rissler LJ. 2016. Union of phylogeography and landscape genetics. *PNAS* 113(29):8079–86
119. Rodríguez-Cerezo E, Moya A, García-Arenal F. 1989. Variability and evolution of the plant RNA virus pepper mild mottle virus. *J. Virol.* 63(5):2198–203
120. Rolland M, Glais L, Kerlan C, Jacquot E. 2008. A multiple single nucleotide polymorphisms interrogation assay for reliable Potato virus Y group and variant characterization. *J. Virol. Methods* 147(1):108–17
121. Romero-Severson E, Skar H, Bulla I, Albert J, Leitner T. 2014. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. *Mol. Biol. Evol.* 31(9):2472–82
122. Roossinck MJ. 2014. Metagenomics of plant and fungal viruses reveals an abundance of persistent lifestyles. *Virology* 5:767
123. Roossinck MJ, Martin DP, Roumagnac P. 2015. Plant virus metagenomics: advances in virus discovery. *Phytopathology* 105(6):716–27
124. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, et al. 1977. Nucleotide sequence of bacteriophage φX174 DNA. *Nature* 265:687–95
125. Sanjuán R. 2012. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLOS Pathog*. 8(5):e1002685
126. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J. Virol.* 84(19):9733–48
127. Schierup MH, Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156(2):879–91
128. Schubert J, Habekuß A, Kazmaier K, Jeske H. 2007. Surveying cereal-infecting geminiviruses in Germany—diagnostics and direct sequencing using rolling circle amplification. *Virus Res.* 127(1):61–70
129. Schulte PA, Perera FP. 1993. *Molecular Epidemiology: Principles and Practices*. San Diego: Academic
130. Seo TK, Thorne JL, Hasegawa M, Kishino H. 2002. A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics* 18(1):115–23
131. Shah VB, McRae BH. 2008. Circuitscape: a tool for landscape ecology. *Proc. Python Sci. Conf.*, 7*th*, *Pasadena*, Aug. 19–24, pp. 62–65. **https://hal.archives-ouvertes.fr/hal-00502586**
132. Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* 28(2):879–87
133. Simmons HE, Dunham JP, Stack JC, Dickins BJA, Pagán I, et al. 2012. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *J. Gen. Virol.* 93(8):1831–40
134. Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. 2014. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci. Rep*. 4:4003
135. Soubeyrand S. 2016. Construction of semi-Markov genetic-space-time SEIR models and inference. *J. Soc. Fr. Stat.* 157(1):129–52

117. This review summarizes tip dating approaches and provides a guide to performing such analyses.

137. Stainton D, Martin DP, Muhire BM, Lolohea S, Halafihi M, et al. 2015. The global distribution of *Banana bunchy top virus* reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evol.* 1(1):vev009
138. Sztuba-Solińska J, Urbanowicz A, Figlerowicz M, Bujarski JJ. 2011. RNA-RNA recombination in plant virus replication and evolution. *Annu. Rev. Phytopathol.* 49:415–43
139. Tomimura K, Špak J, Katis N, Jenner CE, Walsh JA, et al. 2004. Comparisons of the genetic structure of populations of *Turnip mosaic virus* in West and East Eurasia. *Virology* 330(2):408–23
140. Trovão NS, Baele G, Vrancken B, Bielejec F, Suchard MA, et al. 2015. Host ecology determines the dispersal patterns of a plant virus. *Virus Evol.* 1(1):vev016
141. van Etten J. 2015. *R Package Gdistance: Distances and Routes on Geographical Grids*. **https://cran.r-project.org/package = gdistance**
142. Viboud C, Nelson MI, Tan Y, Holmes EC. 2013. Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philos. Trans. R. Soc. B* 368(1614):20120199
143. Vurro M, Bonciani B, Vannacci G. 2010. Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences. *Food Secur.* 2(2):113–32
144. Wang IJ. 2010. Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol. Ecol.* 19(13):2605–8
145. Wheeler DC, Waller LA, Biek R. 2010. Spatial analysis of feline immunodeficiency virus infection in cougars. *Spat. Spatiotemporal Epidemiol.* 1(2–3):151–61
146. Worby CJ, Lipsitch M, Hanage WP. 2014. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLOS Comput. Biol.* 10(3):e1003549
147. Yasaka R, Nguyen HD, Ho SYW, Duchêne S, Korkmaz S, et al. 2014. The temporal evolution and global spread of *Cauliflower mosaic virus*, a plant pararetrovirus. *PLOS ONE* 9(1):e85641
148. Yasaka R, Ohba K, Schwinghamer MW, Fletcher J, Ochoa-Corona FM, et al. 2015. Phylodynamic evidence of the migration of turnip mosaic potyvirus from Europe to Australia and New Zealand. *J. Gen. Virol.* 96(3):701–13
149. Ypma RJF, van Ballegooijen WM, Wallinga J. 2013. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* 195(3):1055–62
150. Ypma RJF, Bataille AMA, Stegeman A, Koch G, Wallinga J, van Ballegooijen WM. 2012. Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proc. R. Soc. B* 279(1728):444–50

Résultats clés de l'Article 2

EXPLOITER L'INFORMATION GENETIQUE POUR EVALUER LA DISPERSION DES VIRUS DANS LES PAYSAGES

Cette synthèse bibliographique analyse les approches d'épidémiologie moléculaire qui ont été développées pour suivre la dispersion des virus dans les paysages. Les principaux points à retenir sont les suivants :

1. Le taux de substitution élevé des virus implique que les processus évolutifs et épidémiologiques sont observables à la même échelle de temps. Les génomes viraux peuvent donc être utilisés pour inférer la dynamique du virus dans les paysages à travers le temps et l'espace.

2. Les progrès des méthodes de caractérisation des virus ont permis d'élargir nos connaissances sur les espèces virales existantes et sur leur diversité intraspécifique.

3. Des plans d'échantillonnage appropriés sont nécessaires pour éviter les biais lorsque l'on étudie l'influence des variables biologiques et environnementales sur la structuration de la diversité des populations virales.

4. Les virus étant clonaux, les analyses exploratoires sans hypothèse sont plus appropriées que les approches classiques de génétique des populations pour décrire la structure spatiale de la diversité virale.

5. Les modèles phylogéographiques permettent d'inférer des voies d'invasion sur de grandes zones à partir de coordonnées géographiques de séquences datées.

6. De nouvelles approches combinant la génétique à l'échelle du paysage (« landscape genetics ») et la phylogéographie permettent de tester l'impact de la configuration et de la composition du paysage sur la dynamique spatio-temporelle du virus.

7. De récents modèles phylodynamiques et des modèles SEIR prenant en compte des données épidémiologiques et génétiques peuvent être utilisés pour inférer des arbres de transmission et des paramètres épidémiologiques clés (en se basant sur des séquences de virus provenant d'épidémies intensément échantillonnées).

2. Application au virus de la sharka

Afin d'estimer les paramètres épidémiologiques d'une maladie à l'échelle de l'hôte, un modèle généticospatio-temporel a été développé dans l'unité BioSP à Avignon, en collaboration avec des chercheurs de BGPI et de l'université de Glasgow (Mollentze et al. 2014, Morelli et al. 2012, Soubeyrand, 2016). Ce modèle prend en compte des données épidémiologiques et génétiques, et permet d'inférer « qui a infecté qui » dans un paysage, ainsi que des paramètres épidémiques clés.

Dans un premier temps, nous avons testé l'efficacité de ce modèle sur des données simulées. Pour cela, j'ai réalisé des simulations d'épidémies de sharka à l'aide du modèle développé par Pleydell et al. (2018) et Rimbaud et al. (2018a, 2018b) que j'ai adapté pour obtenir des données à l'échelle de l'arbre. Ces simulations ont permis de comparer par simulation la précision de la reconstruction des chaines de transmission entre les hôtes avec ou sans la prise en compte des arbres non infectés. Ce test était important pour ensuite appliquer le modèle généticospatio-temporel à des données réelles de manière efficace.

Par la suite, nous avons tenté d'estimer certains paramètres épidémiologiques de la sharka en appliquant le modèle de reconstruction des chaines de transmission sur des données réelles. Le PPV est un virus à ARN dont le génome (10 kb environ) évolue rapidement. Il nous est donc paru possible de reconstruire les chaines de transmission entre les hôtes et d'estimer les paramètres épidémiologiques sous-jacents. L'application de cette approche à la sharka n'a cependant pas été finalisée car des adaptations de la méthode sont vraisemblablement encore nécessaires pour fournir des résultats robustes. Ces travaux sont présentés dans l'article 3 pour lequel j'ai contribué à l'écriture et à la production des résultats des parties qui traitent des simulations des données de dispersion d'une épidémie et des données réelles de sharka. J'ai également participé à la mise en forme des données qui ont servi à réaliser l'inférence des chaines de transmission ainsi qu'à l'analyse des résultats.

ARTICLE 3

Accounting for uninfected hosts in transmission tree reconstruction

Coralie Picard, Sylvie Dallot, Gaël Thébaud and Samuel Soubeyrand

Accounting for uninfected hosts in transmission tree reconstruction

Picard C. (1), Dallot S. (1), Thébaud G. (1) and Soubeyrand S. (2)

(1) BGPI, INRA, Montpellier SupAgro, Univ. Montpellier, Cirad, 34398, Montpellier, France

(2) BioSP, INRA, 84914, Avignon, France.

ABSTRACT

Several approaches coupling epidemiological and evolutionary models, genetic-space-time data and appropriate inference techniques have been proposed to infer transmissions in outbreaks. These approaches are grounded on data, which generally do not contain information on hosts that are not infected during the observation period. The absence of negative data is generally caused by the large number of uninfected hosts compared to the number of infected hosts in studies where the approaches for inferring transmissions were tested. Here, we precisely study the impact of including uninfected hosts in the inference of transmissions in the context of plant epidemiology. For that purpose, we modified an existing genetic-space-time approach allowing the estimation of "who infected whom" by incorporating uninfected hosts in the underlying epidemiological model, and we assessed the advantage of incorporating such hosts in a numerical study based on simulated outbreaks of a plant pathogen (*Plum pox virus*). We showed that integration of uninfected hosts allowed reconstructing 35% of the transmissions (against 20% without it). Thus, including uninfected hosts in a joint analysis of epidemiological and genetic data provides a better understanding of the spatial epidemiology of a pathogen and provides valuable insights into transmission dynamics. Such knowledge on transmissions is crucial for designing efficient control policies.

Keywords: transmission tree, space time genetic, sharka, landscape, SEIR

1. Introduction

Epidemics caused by pathogen spread through host populations can be a high socioeconomic burden (Klinkowski, 1970; Vurro et al., 2010). In order to minimize the associated costs, governmental agencies often design management strategies relying on scientific expertise. To support public policy decision-making, scientists need to understand and predict how pathogens spread within and between host populations (Ferguson et al., 2003; Keeling et al., 2003). More specifically, the reconstruction of transmission routes during past epidemics may help to predict how the same pathogens will spread through similar populations in future outbreaks (Picard et al., 2017). Indeed, understanding the history of transmission events can highlight key drivers of transmission, provide refined estimates of epidemiological parameters and point out risk factors related to vectors, reservoirs and landscape components, which can help build epidemiological projections (Ostfeld et al., 2005).

However, identifying transmission links between hosts in a landscape remains a challenging task. Indeed, the locations of diseased individuals through time are usually consistent with many different transmission trees (i.e. "who-infected-who"). Pathogen genome sequences collected during epidemics can help discriminating between such trees, because genetic data can provide critical additional information regarding the relationships between hosts infected by measurably evolving pathogens (i.e. that fix mutations across their genome during the course of a single outbreak; Picard et al., 2017).

Various models integrating genetic and spatiotemporal data have been developed to understand transmission links between hosts (Jombart et al., 2014; Lau et al., 2015; Mollentze et al., 2014; Morelli et al., 2012; Ypma et al., 2012, 2013; Worby et al., 2014). These models enable to infer epidemiological processes, specifically the most likely transmission tree reflecting "who infected whom", and other parameters related to the infection strength, the latent period, the incubation period, the dispersal kernel (which partly determines the speed and spatial extent of disease spread), and the substitution rate (Soubeyrand, 2016). Such genetic-space-time models are generally stochastic and based on an SEIR (Susceptible, Exposed, Infectious, Removed) structure explicitly representing successive sanitary statuses of host individuals.

For now, these models have been used for animal and human diseases but, to our knowledge, they have never been applied to plant diseases. In addition, they never accounted for the localization of the uninfected hosts, since such data can be difficult to obtain, particularly for animal and human diseases. Regarding plant diseases, the locations of uninfected hosts are more easily available and

have already been used to infer disease transmissions in the absence of genetic data (Gibson, 1997; Neri et al., 2014). However, the number of uninfected hosts is generally much higher than the number of infected hosts; thus, it can become challenging to account for them in genetic-space-time models. Indeed, a key challenge in plant disease modelling is to assess the impact of incomplete host data on model predictions (Cunniffe et al., 2015).

In this context, we aimed to understand how taking into account uninfected hosts into transmission tree reconstruction can improve the estimation of "who infected whom". For that purpose, we modified an existing genetic-space-time SEIR model and its associated estimation method (Mollentze et al., 2014; Soubeyrand, 2016). Then, we assessed inference performance using simulated data obtained by coupling a micro-evolutionary model of pathogen sequences with a spatio-temporal epidemiological model built for sharka (Picard et al., in prep; Picard et al., in revision; Pleydell et al., 2018; Rimbaud et al., 2018). Sharka is one of the most damaging diseases of stone fruit trees belonging to the genus *Prunus* (e.g. peach, apricot and plum) (Cambra et al., 2006; Rimbaud et al., 2015) and it is caused by *Plum pox virus* (PPV, *Potyviru*s genus). As many RNA viruses, PPV is expected to evolve quickly and its evolutionary and epidemic dynamics are supposed to happen at similar time scales.

2. Materials and methods

2.1. Genetic-space-time SEIR model

In this article, we extended the genetic-space-time SEIR model described by Soubeyrand (2016) representing the transmissions of an infectious disease within a population of susceptible hosts and the micro-evolution of the pathogen causing the disease. The genetic-space-time SEIR model results from the coupling of a semi-Markov, individual-based, continuous-time, spatial epidemic model that governs the transitions between the sanitary statuses of individuals (S, E, I and R) and a Markovian evolutionary model that governs nucleotide substitutions in the sequence of the pathogen at the host level. This model was extended to apply it to sharka epidemics: it handles (i) the emergence of hosts across the study period, (ii) the delay between the detection of infected hosts and their removal.

Tables 1 and 2 describe the epidemiological and evolutionary events that are included in the genetic-space-time SEIR model. Mathematical details are provided in Soubeyrand (2016). Here, we simply

comment on the points related with the model extension. First, the delay between the detection of infected hosts and their removal is treated like in Morelli et al. (2012). The second extension relates to host emergence (here, host plantation), which is supposed to occur at known dates and in the healthy state. If a host is actually infected before plantation, the method for inferring transmissions should select early dates of infection for this host, which would mimic an infection at the plantation date.

Table 1. Possible events and corresponding transition rates or distributions for the semi-Markov, individual-based, continuous-time, spatial epidemic model SEIR model. Host *k* emerges at its (known) date of plantation. Then, after infection, it enters the exposed stage at the rate given in the table; this rate is defined as the sum of a basic risk α_0 and the contributions of infectious hosts at time *t* weighted by a kernel *w* computed at the distances d_{jk} between the focal host *k* and the infectious hosts. *w* was specified as the 2D exponential kernel parameterized by γ: $w(d) = \exp\left(-\frac{d}{\gamma}\right)/(2\pi\gamma^2)$. The duration of the exposed stage (latency period) and the duration between the end of the exposed stage and the detection of the infected host are drawn from a gamma distribution. Note that the gamma distribution is parameterized here by its mean and its standard deviation. Finally, host *k* is removed at the (known) uprooting date.

Description	**Event**	**Rate**	**Distribution**[a]
Host emergence	S_k: 0 → 1		Dirac(plantation date)
Infection	S_k: 1 → 0 & E_k: 0 → 1	$\alpha_0 + \alpha_1 \sum_{j \neq k} w(d_{jk}) I_j(t)$	
Beginning of infectious stage	E_k: 1 → 0 & I_k: 0 → 1		Gamma(β_1 , β_2)
Detection	I_k: 1 → 1		Gamma(δ_1 , δ_2)
End of infectious stage	I_k: 1 → 0 & R_k: 0 → 1		Dirac(uprooting date)

[a] For Infection, the equation corresponds to a rate.

Table 2. Possible events and corresponding substitution rates for the Markovian evolutionary model. Letters A, C, G and U denotes nucleotides adenine, cytosine, guanine and uracil, respectively.

Description	**Event**	**Rate**
Transition	A→G or G→A or C→U or U→C	μ_1
Transversion (type 1)	A→U or U→A or C→G or G→C	μ_2
Transversion (type 2)	A→C or C→A or G→U or U→G	μ_3

2.2. Estimation method

The estimation of model parameters and latent variables (sources of infection, infection times and durations of exposed stages) was carried out in the Bayesian framework, following the method of Soubeyrand (2016) based on the approximate genetic likelihood. This method includes the reconstruction of sequences that are transmitted at the infection events by using a parsimonious reconstruction algorithm. Only direct transmissions were reconstructed (indirect transmissions handled in Jombart et al. (2014) and Mollentze et al. (2014) were not taken into account here). For each treated dataset, the posterior distribution was evaluated with three interacting MCMC chains (chain length: 10^5; burn-in: 4000 iterations; thinning: every 100 iterations; interaction between chains: every 2000 iterations).

Two versions of the estimation method were run for each dataset: hosts that remained healthy were either included in or removed from the dataset. Incorporating healthy hosts in the estimation method amounts to compute for these hosts the probability that they have not been infected up to the end of the observation period. This probability is incorporated into the transmission likelihood (see Soubeyrand, 2016). To handle healthy hosts, we ignored the possibility that apparently healthy hosts were actually infected.

2.3. Simulated data

2.3.1 Simulation of sharka epidemics

In order to assess the performance of the estimation method (with and without the uninfected hosts), we used an existing simulation model of sharka disease (Picard et al., in prep; Picard et al., in revision; Pleydell et al., 2018; Rimbaud et al., 2018). This stochastic, spatially explicit SEIR model includes 6 epidemiological parameters characterizing the epidemic, and 21 disease management parameters enabling to simulate orchard surveillance, plantation bans and removal of infected trees. Here, we used the same variation ranges of epidemiological parameters as in by Picard et al. (in prep) to simulate 20 established epidemics. In addition, for each simulation, the epidemic spreads during 5 years without management, followed by 10 years with one survey per year performed with a detection probability of 0.66 (once an infected tree is detected, it is removed from the simulation).

These epidemics were run on a virtual landscape comprising 2508 trees grouped into 16 patches (Fig 1). To simulate this landscape, we generated a grid pattern of 10,000 squares (16 m² each) representing potential trees. Among them, we randomly selected 20 rectangular patches with

random width and length (between 2 and 21 trees). When two rectangular patches included the same trees, we grouped them into a single patch (hence the 16 final patches).

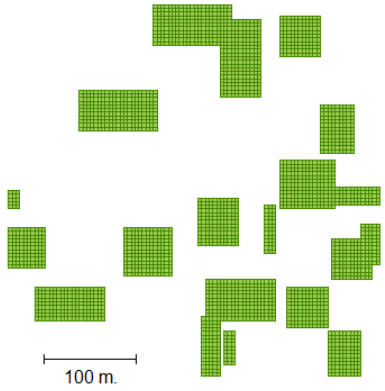


Figure 1: Simulated landscape composed of 2508 trees allocated into 16 different patches

2.3.2. Simulation of pathogen sequence evolution

The evolution of PPV in the hosts was simulated conditional on the transmission tree obtained from the simulation of the epidemics. We used sequence fragments of 10,000 nucleotides, among which a fraction was fixed. We assumed that there is no within-host diversity, i.e. at any time each host is infected by at most one genomic sequence. When a tree was infected at plantation, the sequence of the pathogen was drawn from a set of 20 slightly varying reference sequences (each of these 20 sequences were obtained by uniformly randomly modifying any nucleotide with a 5‰ chance from a reference genome). When a tree was infected by another one within the simulated landscape, the current PPV sequence infecting the source tree was used as the initial sequence in the receiving host. The substitution of nucleotides within each host was performed forward in time as a Markov chain with heterogeneous rates of substitution across the sequence. More specifically, the rate of substitution for each nucleotide was drawn from a zero-inflated gamma distribution with the probability of zero equal to 0.3 (the shape of the gamma distribution was 0.3 and its scale parameter was 10^{-4}). This substitution rate (10^{-4} subs/site/year) was estimated under a Bayesian framework (BEAST 1.8 software) using 86 heterochronous whole genome sequences of PPV isolates sampled in peach orchards of southern France from 1991 to 2008 (Dallot et al., 2016).

2.3.3. Subsampling for generating datasets used in the inference

We simulated 20 outbreaks through the 16 patches and we retrieved the geographical coordinates of the trees, as well as their plantation and removal dates. We also got the simulated dates of disease surveillance, the sanitary status of the trees at each date (symptomatic/ non-symptomatic) as well as the simulated viral sequences associated with the infected trees. In order to reduce the overall computational cost of outbreak reconstructions, we did not attempt to reconstruct the transmissions between all the trees in the landscape simultaneously, but only between trees of some patches (note that this approach reflects a frequent situation since epidemiological and genetic data are generally available for only a part of the landscape). For each outbreak, we selected the patches with the highest number of infected prunus trees (excluding patches with more than 400 trees), without exceeding a total of 1200 trees (healthy and infected) on all patches. The data corresponding to these patches were used to reconstruct transmission chains between hosts with the estimation method introduced above, both with and without the non-infected trees.

2.4. Specification of prior distributions for the inference

For the genetic-space-time SEIR model, vague exponential priors with mean 100 were used for the infection strengths α_0 and α_1, corresponding respectively to exogenous (from trees not included in the dataset) and endogenous (from trees included in the dataset) sources. In addition, informative gamma priors with mean and standard deviation equal to 331 m was specified for the mean dispersal distance 2γ (331 m was the mean dispersal distance derived from the estimation for sharka in Pleydell et al. (2018), to 1.92 yr and 0.1 for the mean incubation duration θ_1, to 0.66 yr and 0.1 for the standard deviation of the incubation duration θ_2 (these prior means were taken from Pleydell et al., 2018), to 3 yr and 0.1 for the mean duration δ_1 between the end of the exposed stage and the detection, and to 2 yr and 0.1 for the standard deviation δ_2. Vague exponential priors with mean 10^{-4} sub/site/year were used for the substitution rates μ_1, μ_2 and μ_3 (Dallot et al., 2016). From year 1, the first year of the epidemic, a normal prior with mean -100 yr and standard deviation 50 yr was used for the time of the most recent common ancestor.

3. Results

We tried to reconstruct the transmission trees with a median of 45 infected trees for each of the 20 simulations performed. Among these infected trees, on average 20 trees were infected by an external source (i. e. by a tree located outside the simulated landscape, and for which we do not have informations), and on average 26 by a tree located in the landscape, for which we have epidemiological and genetic data (Fig 2A and B).

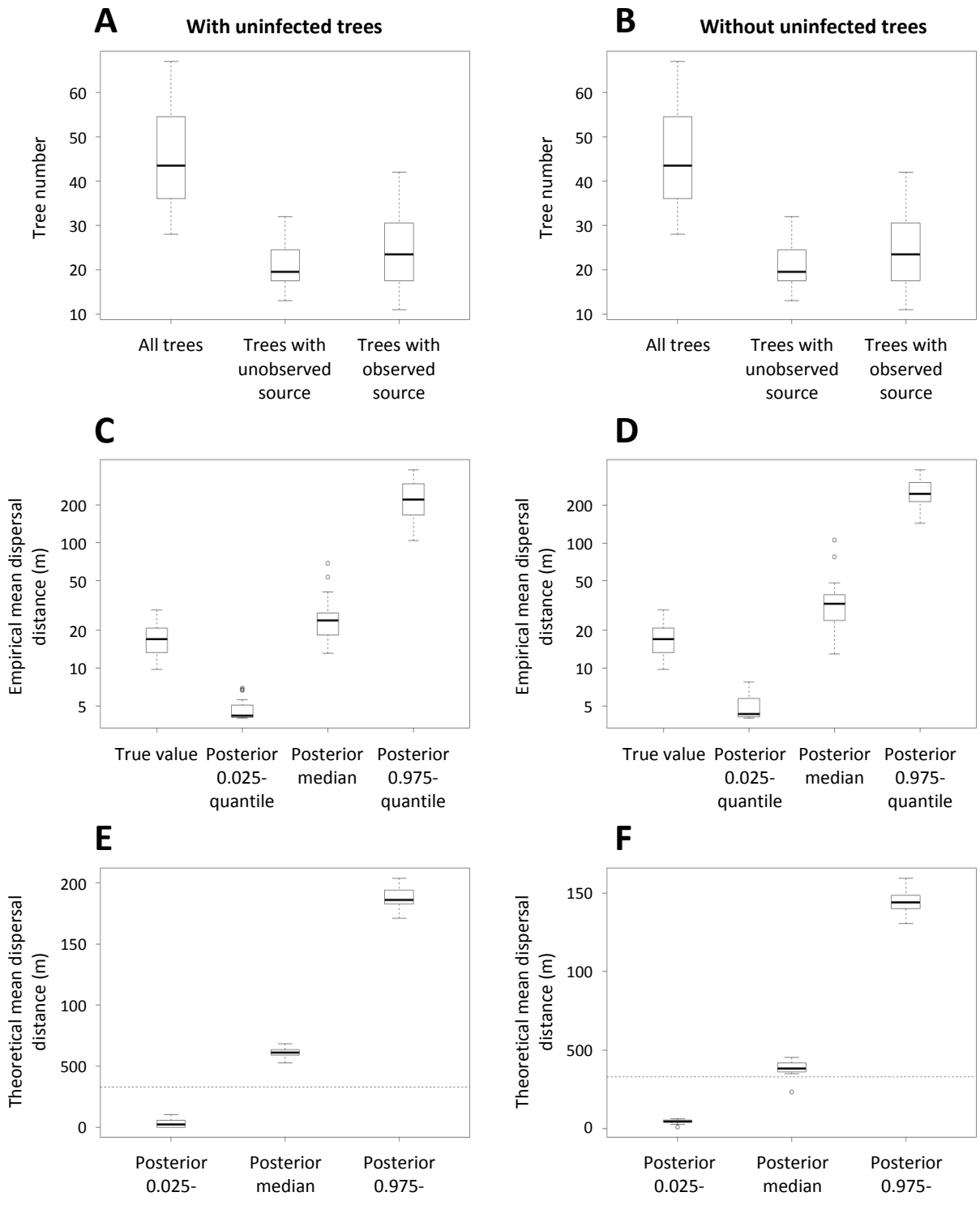


Figure 2: Results of the inference of simulated sharka epidemics. A and B: among the 20 simulated outbreaks, frequency of all infected trees, and frequency of infected trees whose source of infection is located outside (unobserved source) or inside (observed source) the selected orchards. C and D: empirical mean dispersal distance (calculated from the transmissions between trees located in the simulated landscape). E and F: Theoretical mean dispersal distance (accounting for the external transmissions). A, C and E: results with uninfected trees. B, D and F: results without uninfected trees.

The genetic-space-time model allowed us to reconstruct 35% of the transmissions when the uninfected trees were considered but only 20% without taking them into account (Fig 3). As a comparison, we would have reconstructed only 0.02% of the transmissions by randomly generating transmissions trees. In addition, we identified less transmissions when the source trees were located in the landscape (30% with the uninfected trees and 19% without), than when the trees were infected by an external source (45% with the uninfected trees and 21% without).

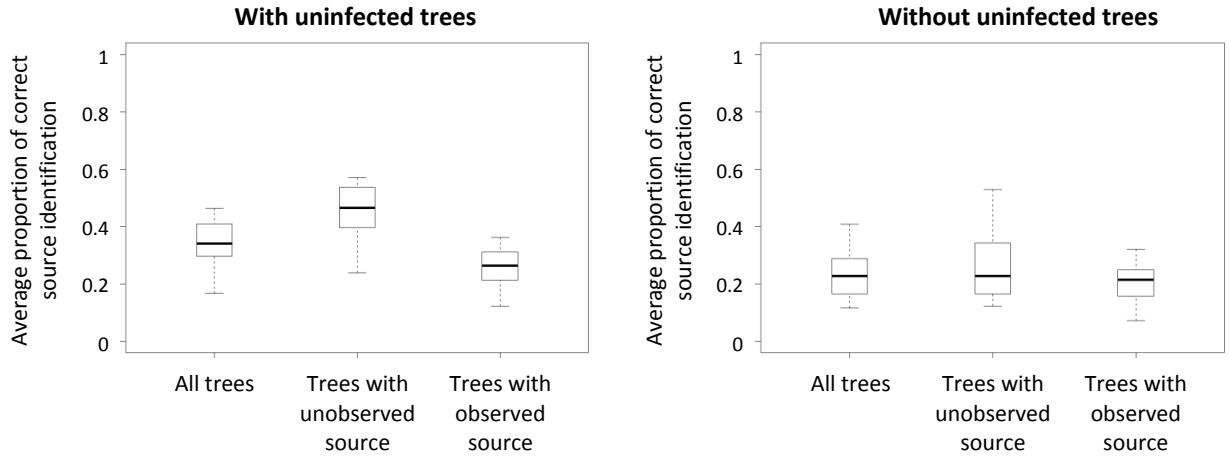


Figure 3: Accuracy of the reconstruction of transmissions assessed from simulated sharka epidemics.

The mean of the empirical dispersal (calculated from the transmissions between infected prunus trees located in the simulated landscape) was correctly estimated when taking into account the uninfected trees. Indeed, the median of the empirical dispersal was estimated at 23m (accounting for uninfected trees), and the true value (calculated from the simulations) at 18m (Fig 2C and D). Without the uninfected trees, the median of the empirical dispersal was slightly overestimated (30m). However, in both cases, the estimated quantile distributions were far from the true value of the empirical dispersal.

We also estimated the prediction accuracy of the theoretical dispersal, which differs from the empirical dispersal accounting for the external transmissions in its calculation. The estimation of the mean of theoretical dispersal was overestimated with or without the uninfected trees (although the prediction accuracy was higher without the uninfected trees). Indeed, in the sharka simulation model, the mean of theoretical dispersal was 331m, and the estimations were 600 and 400 with and without the uninfected trees respectively (Fig 2E and F). However, it was difficult to estimate this

epidemiological parameter since we performed our simulations in a small area (only slightly bigger than 331m).

4. Discussion

4.1. Conclusion and applicability of the approach

In this study, we showed how including uninfected hosts in a genetic-space-time model can improve the reconstruction of transmission trees. For that purpose, we used the example of sharka disease, for which we simulated dispersal and management, as well as the genetic sequence of the virus for each infected tree. Then, the genetic-space-time model allowed us to reconstruct the transmission links from these simulated data. We showed that accounting for the uninfected trees improved the inference of transmission links: we reconstructed 35% of the transmissions with the uninfected trees (against 20% without).

However, the epidemics were here simulated through landscapes for which the orchards are composed of few trees (the bigger simulated orchard includes only 231 trees), which represents an area with traditional arboriculture. By contrast, more recent exploitations are generally composed of bigger orchards (which can include more than 1000 trees). In such situation, the performance of the transmission tree reconstruction may be different since the landscape present less discontinuity between hosts. However, testing the genetic-space-time model in this case can increase the inference duration (which is multiplied increasing the number of trees).

Our results could allow improving numerous studies which aim to understand and to predict how pathogens spread within host populations, which could help to develop adapted management strategies to control pathogens. This approach is particularly interesting for diseases of perennial plants since they are localized at the same place during several years. However, using genetic-space-time model on diseases which spread on annual hosts can be more challenging since it is difficult, if not impossible, to follow the temporal signal included in the genetic sequences (which is essential to perform the inference). Similarly, our approach could be difficult to transpose to human and animal diseases since the hosts are generally mobile. To address this issue, it could be interesting to account for the host movements in the inference, but aside the need of lot of material, tracking them can cause ethical problems.

4.2. Transmission links inference of real sharka data

To go even further in this study, we attempted to use the genetic-space-time model on a real sharka epidemic. The material and method used is described in S1 text. However, the distribution of the transmissions between trees was unsatisfactory. Indeed, transmission links obtained with the inference were characterized by a small number of trees which infected numerous other trees located further in the landscape. We would have expected that the model infer less long distance transmissions and more local transmissions (short distance). These unsatisfactory results are probably due to a lack of information from the dataset and a maladaptation of the method for these data. Indeed, we attempted to reconstruct the transmissions trees with data only sampled over 3 consecutive years, the temporal signal was thus difficult to capture, especially because latency duration of sharka may vary from few weeks to few years.

In order to improve this preliminary work, we tried to remove from the dataset the infected trees for which we did not know the virus genetic sequence. Indeed, the inference suggested that these trees were the source of most of the transmissions, which is unlikely. In addition, we modified the prior of the model corresponding to the mutation rate in order to limit the possibility of long transmissions. Thus, we inferred transmissions links between infected trees with a transmission rate of 10^{-5} subs/site/year instead of 10^{-4} subs/site/year. Nevertheless, these two attempts to improve the inference were not satisfactory. However, we did not explore ways that could improve the estimation of transmissions. Firstly, although the inference can be much longer, we could perform it with all our available data (for now, we carried out the inference on only 6 out of 19 orchards for which we dispose of information), which may prevent some long distance transmissions. Then, for some of the trees sampled, we had both majority and minority genetic sequences (i.e. found in smaller quantities). For now, we only used the information of the majority sequences because the model only allowed accounting for a unique genetic sequence for one tree. We could modify a part of the model to take into account this information. To finish, it could be interesting to sample data over more than 3 years in order to really exploit the temporal information of the data.

ACKNOWLEDGMENTS

This work was supported by the CIRAD-UMR AGAP HPC Data Center of the South Green Bioinformatics platform (http://www.southgreen.fr). This work is supported by an INRA/ANSES scholarship to CP, and by INRA (BEcOSMASH project, funded by the SMaCH Metaprogram and the SAE2 Department) and Agropolis Fondation (E-SPACE project).

REFERENCES

Ashkenazy H., Penn O., Doron-Faigenboim A., Cohen O., Cannarozzi G., Zomer O. and Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 40:W580–W584.

Cambra M., Capote N., Myrta A. and Llácer G. (2006). *Plum pox virus* and the estimated costs associated with sharka disease. EPPO Bull. 36(2):202-204.

Cunniffe N.J., Koskella B., E. Metcalf C.J., Parnell S., Gottwald T.R. and Gilligan, C.A. (2015). Thirteen challenges in modelling plant diseases. Epidemics 10:6–10.

Dallot S., Borron, S. Bertanpetit, E., Dupuy V., Jacquot E. and Thébaud, G. (2016). Exploiting viral genetic diversity to uncover sharka dispersal at regional and local scales. In: *Building bridges between disciplines for sustainable management of plant virus diseases. IPVE 2016. Programme and abstracts* (p. 23). Presented at 13. International plant virus epidemiology symposium, Avignon, FRA (2016-06-06 - 2016-06-10). 165 p. https://prodinra.inra.fr/record/361232.

Ferguson N.M., Keeling M.J., Edmunds W.J., Gani R., Grenfell B.T., Anderson R.M. and Leach, S. (2003). Planning for smallpox outbreaks. Nature 425(6959):681–685.

Gibson G.J. (1997). Markov Chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. J. R. Stat. Soc. Ser. C Appl. Stat. 46:215–233.

Jombart T., Cori A., Didelot X., Cauchemez S., Fraser C. and Ferguson, N. (2014). Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. PLoS Comput Biol 10(1):e1003457.

Keeling,M.J., Woolhouse M.E.J., May R.M., Davies G. and Grenfell B.T. (2003). Modelling vaccination strategies against foot-and-mouth disease. Nature 421(6919):136–142.

Klinkowski M. (1970). Catastrophic plant diseases. Annu. Rev. Phytopathol. 8:37-60.

Lau M.S.Y., Marion G., Streftaris G., Gibson G. (2015). A systematic Bayesian integration of epidemiological and genetic data. PLOS Comput. Biol. 11(11):e1004633.

Mollentze N., Nel L.H., Townsend S., Roux K., Hampson K., Haydon D.T. and Soubeyrand, S. (2014). A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proc. R. Soc. Lond. B Biol. Sci. 281(1782):20133251.

Morelli M.J., Thébaud G., Chadœuf J., King D.P., Haydon D.T. and Soubeyrand, S. (2012). A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. PLoS Comput Biol 8(11):e1002768.

Neri F.M., Cook A.R., Gibson G.J., Gottwald T.R. and Gilligan, C.A. (2014). Bayesian analysis for inference of an emerging epidemic: citrus canker in urban landscapes. PLOS Comput Biol 10(4):e1003587.

Ostfeld R.S., Glass G.E. and Keesing F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol. Evol. 20(6):328–336.

Picard C., Dallot S., Brunker K., Berthier K., Roumagnac P., Soubeyrand S., Jacquot E. and Thébaud, G. (2017). Exploiting Genetic Information to Trace Plant Virus Dispersal in Landscapes. Annu. Rev. Phytopathol. Vol. 55 (in press).

Picard C., Soubeyrand S., Jacquot E. and Thébaud G. Analyzing the influence of landscape aggregation on disease spread to improve management strategies. In revision.

Picard C., Picheny V., Bonnot F., Soubeyrand S. and Thébaud G. In silico optimization of a strategy for landscape-wide plant disease management. In prep.

Pleydell D., Soubeyrand S., Dallot S., Labonne G., Chadœuf J., Jacquot E. and Thébaud, G. (2018). Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLOS Comput. Biol. 14(4):e1006085.

Rimbaud L., Dallot S., Gottwald T., Decroocq V., Jacquot E., Soubeyrand S. and Thébaud G. (2015). Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annu. Rev. Phytopathol. 53:357–378.

Rimbaud L., Dallot S., Bruchou C., Thoyer S., Jacquot E., Soubeyrand S. and Thébaud G. (2018). Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. bioRxiv 315747. (doi: https://doi.org/10.1101/315747)

Soubeyrand S. (2016). Construction of semi-Markov genetic-space-time SEIR models and inference. J. Société Fr. Stat. 157(1):129–152.

Vurro M., Bonciani B. and Vannacci, G. (2010). Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences. Food Secur. 2(2):113–132.

Worby C.J., Lipsitch M. and Hanage W.P. (2014). Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLOS Comput. Biol. 10(3):e1003549.

Ypma R.J.F., Bataille A.M.A., Stegeman A., Koch G., Wallinga J. and Van Ballegooijen, W.M. (2012). Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. Proc. R. Soc. Lond. B Biol. Sci. 279(1728):444–450.

Ypma R.J.F., Van Ballegooijen W.M. and Wallinga J. (2013). Relating phylogenetic trees to transmission trees of infectious disease outbreaks. Genetics 195(3):1055–1062.

S1 Text: Inference of transmission tree of real sharka data: materiel and method

We attempted to infer the transmissions of a real sharka epidemic. Here, we present the epidemiological and genetic data used, and we introduce the modifications made in the genetic-space-time model.

1. Epidemiological and genetic sharka data

We used data from 3 years of surveillance (2004, 2005 and 2006) of a very close set of 19 orchards located in southern France (Fig 1). Disease surveillance was based on visual inspections and PPV symptomatic trees were removed each year following their detection. The 4905 trees (among which 145 were found infected by PPV) planted in the 19 orchards were precisely geo-referenced. The plantation dates of each orchard were recorded as well as the dates of detection and removal of the infected trees. Symptomatic leaves were sampled on each detected PPV infected tree and we obtained the whole genome PPV consensus sequence for 114 infected trees. In addition, a preliminary phylogeny study evidenced that two different genetic clades of PPV were spreading in the study area (Dallot et al., 2016). The common ancestors of these two clades were reconstructed from the 145 genetic sequences sampled on infected trees thanks to FastML server (Ashkenazy et al., 2012). Here, we used the genetic-space-time model on data corresponding to only 6 patches (Fig 1).



Figure 1: Map of peach orchards with uninfected (grey) and infected trees detected in 2004 (yellow), 2005 (pink) and 2006 (blue). Only the orchards framed in black were used in the inference of transmission trees and pathogen dispersal.

2. Genetic-space-time SEIR model modifications

The genetic-space-time SEIR model was extended to handle the use of multiple ancestral sequences of the pathogen corresponding to different genetic clusters. Regarding the epidemiological and evolutionary events that are included in the model, if host *k* is infected by an exogenous source (this possibility depends on the basic risk α_0), then the ancestral sequence selected for the exogenous source is the ancestral sequence used for the genetic cluster to which the pathogen sequence collected from host *k* belongs.

In addition, we specified a new prior for parameters of the genetic-space-time SEIR model. We considered at least two introduction events from genetically differentiated PPV sources corresponding to the distinct clades in the reconstructed dated phylogeny of PPV. Based on BEAST inferences (Dallot et al., 2016), the time to the most recent common ancestor (tMRCA) of these two clades was set at 1985 (CI95% : 1981 – 1989, i.e., 19 years before the discovery of sharka disease in the area).

Résultats clés de l'Article 3

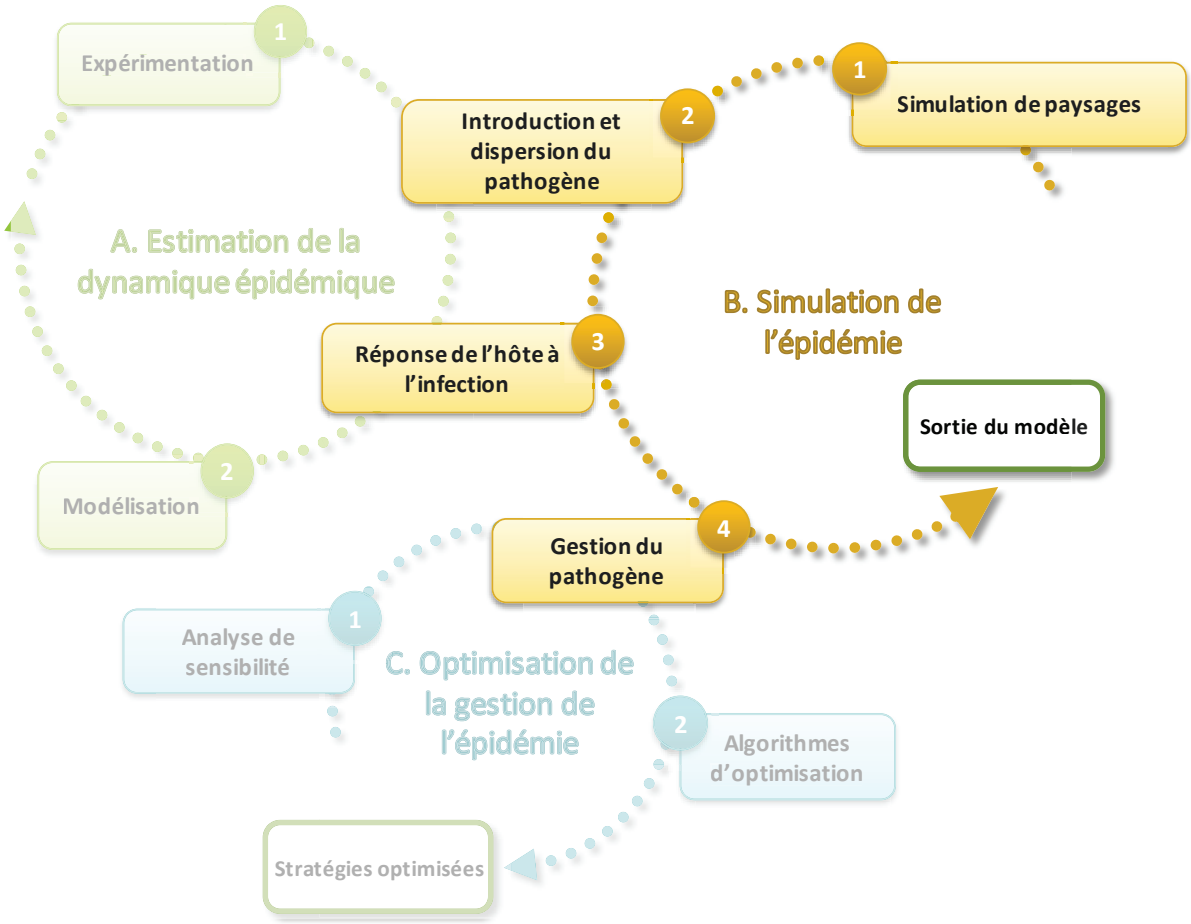
PRISE EN COMPTE DES HOTES NON INFECTES DANS LA RECONSTRUCTION DE CHAINES DE TRANSMISSION

- **Impact de la prise en compte des hôtes non infectés dans l'inférence des transmissions d'une épidémie**
  - Un modèle généticospatio-temporel déjà existant permettant d'inférer « qui a infecté qui » dans un paysage a été modifié pour prendre en compte les hôtes sensibles mais non infectés.
  - Grâce à la reconstruction d'épidémies simulées (et du processus évolutif concomitant), nous avons montré que la prise en compte des hôtes non infectés permettait d'inférer correctement 35% des transmissions (contre 20% sans).
  - Inclure les hôtes non infectés dans une analyse de données épidémiologiques et génétiques permet donc une meilleure compréhension de l'épidémiologie spatiale d'un agent pathogène et fournit des indications précieuses sur la dynamique de transmission. Une telle connaissance des transmissions est cruciale pour concevoir des politiques efficaces pour gérer les épidémies.

- **Reconstruction des chaines de transmission pour une épidémie de sharka**
  - Nous avons tenté de reconstruire les chaines de transmission pour une épidémie de sharka dans un ensemble de vergers proches. Nous disposions des coordonnées géographiques des arbres infectés et des séquences génétiques du virus correspondant.
  - Probablement à cause du manque de signal temporel, nous n'avons pas obtenu de résultats satisfaisants. Les paramètres du modèle représentant les épidémies de sharka n'ont donc pas été modifiés dans les études présentées dans la suite de cette thèse.

Chapitre 2

ETUDE DE L'INFLUENCE DU PAYSAGE SUR LES STATEGIES DE GESTION DES MALADIES



Lorsque des paramètres décrivant la dynamique d'une épidémie sont disponibles, des modèles de simulation permettent de tester différents scénarios épidémiques et/ou de gestion. Dans le cadre d'une gestion spatialisée de la maladie, ces modèles ont généralement besoin d'un paysage explicite et réaliste, d'un scénario d'introduction et de dispersion du pathogène, d'équations décrivant les changements de statut des hôtes et d'actions de gestion visant à réduire la propagation de la maladie.

Le paysage (caractérisé ici par la disposition spatiale et la forme des parcelles) n'est pris en compte que depuis peu dans les études de modélisation en épidémiologie, bien qu'il puisse avoir un fort impact. Ainsi, les études Pleydell et al. (2018) et Rimbaud et al. (2018a, 2018b) ont été réalisées sur un unique paysage. Elles ne permettent donc pas d'estimer des paramètres épidémiologiques et d'identifier des stratégies de gestion efficaces sur différents paysages. Afin d'étudier l'influence du paysage sur les stratégies de gestion de la sharka, j'ai tout d'abord modifié ce modèle pour permettre la simulation de l'épidémie dans des paysages variés.

Pour ce faire, j'ai modifié le paysage utilisé dans l'étude de Rimbaud et al. (2018a) afin d'obtenir des paysages de taille et de densité différentes. Dans cette approche, le paysage constitué de 553 parcelles a été dupliqué par 3 (avec un total de 1659 parcelles), puis par 7 (avec un total de 3871 parcelles). Des parcelles de ces paysages ont ensuite été retirées pour diminuer la densité. Des simulations de l'épidémie avec la stratégie de gestion française ont ensuite été réalisées pour évaluer l'impact des caractéristiques du paysage sur la dynamique épidémique et l'efficacité de la gestion. Plus précisément, l'influence de la taille du paysage et de la densité des parcelles cultivées a été étudiée sur deux critères : un critère agronomique et un critère économique. Le premier critère correspond au nombre moyen équivalent d'arbres pleinement productifs, et le deuxième à la valeur actuelle nette (VAN ; Rimbaud et al. 2018a).

Dans un deuxième temps, j'ai développé un algorithme simulant des paysages réalistes : il permet de définir les principales caractéristiques d'un paysage telles que le nombre de parcelles ou leur agrégation spatiale à partir d'une simulation de tessellation en T. Cet algorithme m'a permis de simuler 3 types de paysages variant par le niveau d'agrégation de leurs parcelles. Ces paysages contrastés ont été utilisés pour toutes les études présentées dans la suite de cette thèse. De même que précédemment, des simulations avec la stratégie de gestion française ont ensuite été réalisées. De plus, pour étudier l'influence du paysage sur les paramètres du modèle et identifier les paramètres clés de la propagation et de la gestion d'une épidémie, j'ai réalisé des analyses de sensibilité sur les 3 paysages définis. Ces analyses permettent de mieux comprendre les épidémies et d'identifier les paramètres de gestion les plus influents sur la VAN (ces paramètres peuvent alors être

d’une importance capitale si l’on souhaite identifier des stratégies de gestion performantes). Elles permettent également d’analyser comment l’influence de ces paramètres varie en fonction du niveau d’agrégation des parcelles.

L’article 4 détaille en partie les résultats de ce chapitre. Il expose également des résultats concernant l’optimisation des paramètres de gestion qui seront abordés dans le chapitre suivant.

ARTICLE 4

Analyzing the influence of landscape aggregation on disease spread to improve management strategies

Coralie Picard, Samuel Soubeyrand, Emmanuel Jacquot and Gaël Thébaud

Article accepté dans *Phytophatology* le 08/08/2018

Analyzing the influence of landscape aggregation on disease spread to improve management strategies

Coralie Picard, Samuel Soubeyrand, Emmanuel Jacquot and Gaël Thébaud

First, third and fourth authors: BGPI, Montpellier SupAgro, INRA, Univ. Montpellier, Cirad, TA A-54/K, 34398, Montpellier Cedex 5, France; second author: BioSP, INRA, Avignon, France.

Corresponding author: G. Thébaud
Short title: Landscape influence on disease management
E-mail address: gael.thebaud@inra.fr

ABSTRACT

Epidemiological models are increasingly used to predict epidemics and improve management strategies. However, they rarely consider landscape characteristics although they can influence the epidemic dynamics, and thus the effectiveness of disease management strategies. Here, we present a generic *in silico* approach which assesses the influence of landscape aggregation on the costs associated to an epidemic and on improved management strategies. We apply this approach to sharka, one of the most damaging diseases of *Prunus* trees, for which a management strategy is already applied in France. Epidemic simulations were carried out with a spatiotemporal stochastic model under various management strategies in landscapes differing in patch aggregation. Using sensitivity analyses, we highlight the impact of management parameters on the economic output of the model. We also show that the sensitivity analysis can be exploited to identify several strategies that are, according to the model more profitable than the current French strategy. Some of these strategies are specific to a given aggregation level, which shows that management strategies should generally be tailored to each specific landscape. However, we also identified a strategy that is efficient for all levels of landscape aggregation. This one-size-fits-all strategy has important practical implications because of its simple applicability at a large scale.

Keywords: landscape, management, optimization, SEIR, sharka, spatiotemporal model, virus

1. Introduction

Understanding epidemiological processes is crucial to anticipate outbreaks, to predict the spread of epidemics, and thus to propose optimized management strategies that aim to reduce or eliminate a disease (Ferguson et al. 2001). However, epidemics are the result of complex interactions between biological processes, human interventions and the spatial arrangement of patches in the landscape. Thus, understanding epidemics and assessing the effectiveness of disease management options is often a difficult task, especially as field trials are generally limited by regulatory, ethical and logistical constraints (particularly for large-scale experimental studies). To overcome these limitations, epidemiological models are an interesting approach because of their ability to test several epidemic and management scenarios using the best available knowledge (Cunniffe et al. 2015; Keeling and Rohani 2008; Keeling et al. 2003).

Spatially explicit models have been used to estimate epidemiological parameters such as dispersal functions (Parnell et al. 2011; Parry et al. 2014; Pleydell et al. 2018; Soubeyrand et al. 2008), infection rates (Cunniffe et al. 2014) and incubation durations (Cunniffe et al. 2014; Pleydell et al. 2018). This approach leads to disease-specific, data-calibrated models that can then be exploited to assess the efficacy of control measures, e.g., sampling frequency and intensity (Parnell et al. 2012, 2014; Soubeyrand et al. 2018), plantation density (Chan and Jeger 1994; Cunniffe et al. 2014; Cunniffe et al. 2015b; Jeger and Chan 1995), insecticide spraying frequency and location (Filipe et al. 2012), and zones and dates of removal (Cunniffe et al. 2014, 2015b; Filipe et al. 2012; Parnell et al. 2009, 2010; Sisterson and Stenger 2012).

However, these modeling studies mostly focused on only one or two management parameters, other parameters being set at their reference value. Rimbaud et al. (2018b) tried to optimize several parameters simultaneously however, like almost all previous studies, they performed simulations in a single landscape and did not consider landscape characteristics. Nevertheless, in order to study outbreaks and large-scale management strategies, considering landscape characteristics can be crucial. Indeed, they can influence epidemic dynamics, implying that the best management strategies may vary depending on the landscape (Papaïx et al. 2014). A review by Ostfeld et al. (2005) analyzed the few studies that demonstrate how spatial locations of crop patches can influence disease risk, suggesting that a true integration of the landscape within epidemiological studies would be fruitful. As a consequence, promising approaches have been developed to integrate landscape characteristics into epidemiological models. For example, it was shown that, for the purpose of eradication, the optimum radius of orchard removals increases with the level of patch aggregation and the host density in the landscape, both factors increasing epidemic spread (Parnell et al. 2009, 2010).

However, in these studies patch layout is summarized by patch centroid coordinates although plot size and shape play an important role in disease dispersal (Mikaberidze et al. 2016; Pleydell et al. 2018), and thus on the impact of disease management. Indeed, such simplification can introduce a bias in connectivity estimates when patches have different shapes and sizes, e.g., the connectivity between the centroids of two patches would erroneously be the same whatever their area. Here, we try to understand how landscape structure influences disease spread and the impact of control options thanks to simulations of disease spread and management on various landscapes.

We apply this approach to sharka, one of the most damaging diseases of trees belonging to the *Prunus* genus (e.g., peach, apricot and plum) (Cambra et al. 2006; Rimbaud et al. 2015). The causal agent of this disease, *Plum pox virus* (PPV, genus *Potyvirus*, family *Potyviridae*), is naturally transmitted by aphids in a nonpersistent manner. The presence of PPV symptoms (such as fruit deformation (Németh 1986), apparition of light green rings, mosaic, mottling, and distortions on the leaves (Rimbaud et al. 2015) reduces potential sales, occasioning a significant economic impact (Cambra et al. 2006), with yield losses up to 100% for the most sensitive cultivars. Different alternatives for sharka management strategies exist in the world (eradication, suppression, containment, or resilience) depending on the epidemic context (Rimbaud et al. 2015). In France, sharka management aims to reduce the number of PPV-infected trees to mitigate its impact (suppression); it is compulsory and defined by a national decree specifying a complex procedure based on nursery protection, frequent visual inspections of orchards and removal of symptomatic trees or, possibly, whole orchards, as well as plantation restrictions (JORF 2011; Fig. 1). In a previous study, key parameters of a sharka epidemic were identified, and an improved management strategy was highlighted for a single landscape (Rimbaud et al. 2018b). In the present article, we use the same model to analyze the influence of landscape characteristics on plant disease control. For that purpose, we first study the influence of landscape structure on *Prunus* productivity under the French management strategy (JORF 2011). Next, we use sensitivity analyses to assess the relative influence of model parameters on crop productivity depending on the level of patch aggregation in the landscape. Then, we exploit the results of these analyses to identify several efficient strategies and we study how the landscape influences their impact. This last point allows to asses if the management can be generic (i.e., if a unique management strategy is efficient for all landscapes), or should be specific to each landscape.

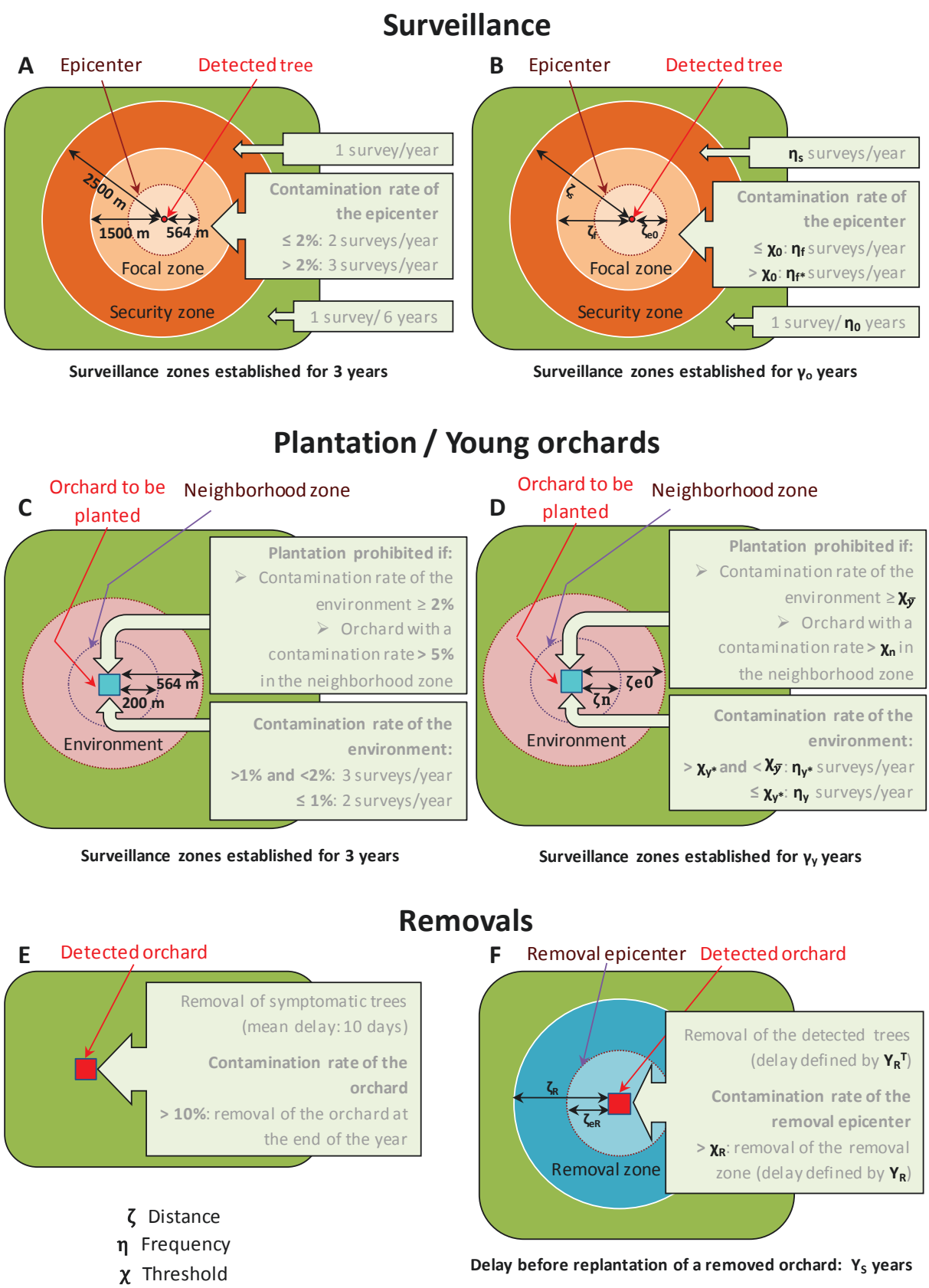


Figure 1: A, C and E, Management actions currently applied in France. B, D and F, Management actions implemented in the model. The detected orchards include at least one observed infected tree.

2. Materials and methods

2.1. Landscape generation

In this study, the landscape is considered as a set of cultivated patches (i.e., pieces of land) in a defined study area on which the pathogen may spread when trees are planted. Patches of different sizes and aggregation levels were simulated by (i) replicating real patches and (ii) simulating patches with a T-tessellation algorithm.

2.1.1. Replication of real patches

In a previous study, Pleydell et al. (2017) and Rimbaud et al. (2018a) developed a model allowing to simulate virus dispersal on a real landscape comprising 553 patches (524 ha of patches in a study area of 2730 ha). This landscape was generated from a database collected in a peach producing area in southeastern France. Here, artificial landscapes were constructed by replicating this real landscape 3 times to obtain a total of 1659 patches and 7 times to obtain 3871 patches (with the size of the study area increasing accordingly). In addition, to obtain landscapes with lower levels of patch aggregation, some of the patches were removed (subsampled) from the 2 replicated landscapes. For the landscape replicated three times, 40% and 70% of the patches were removed randomly from each of the original landscapes (with 553 patches). For the landscape replicated 7 times, 40% and 80% of the patches were removed randomly from each of the original landscapes. Three independent landscapes were generated for each subsampled landscape. An example of each landscape type is displayed in Supplementary Fig. S1. In this way we obtained three sizes of study area, with one aggregation level for the smaller one (corresponding to the real landscape) and three different aggregation levels for each of the larger study areas (and three different landscapes for each subsampled landscape).

2.1.2. Landscape simulations with T-tessellations

To avoid being dependent on a single real landscape, we also simulated new agricultural landscapes with various levels of patch aggregation and a realistic outlook (Fig. 2). Three landscapes comprising n=400 patches were simulated, thereafter called H, M and L based on the value of the aggregation parameter: d=1 (H: high aggregation), d=200 (M: medium aggregation) and d=400 (L: low aggregation) (Fig. 3; Supplementary Fig. S2). These values were chosen to represent diverse patch aggregation levels: with d=400, patches were scattered throughout the study window and with d=1, we obtained only neighboring patches in each cluster. The value of parameter p was chosen to ensure the simulation of on average 15 clusters: p=1-15/n=0.96. This parameter accounts for

landscape irregularities due to the presence of features such as soil, topology, lakes, rivers, roads or towns, which require that patches are generally grouped. We simulated 30 landscapes with these parameter values for each aggregation level.

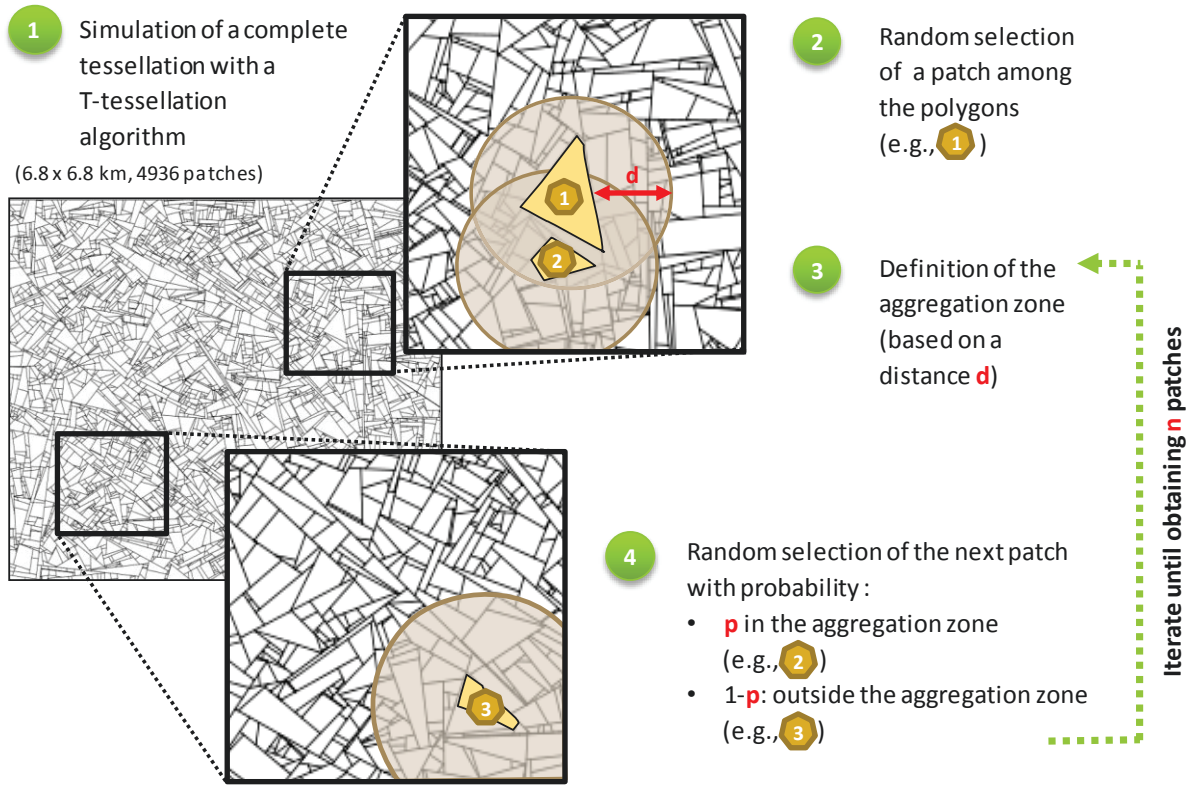


Figure 2: Algorithm for the simulation of landscapes with a specified aggregation level of patches of susceptible hosts. The aggregation level is defined by 3 parameters: n, p and d. Parameter d determines the size of an "aggregation zone" including all polygons located within d meters of a previously selected patch. In steps 2 and 4, patches are selected randomly and uniformly.

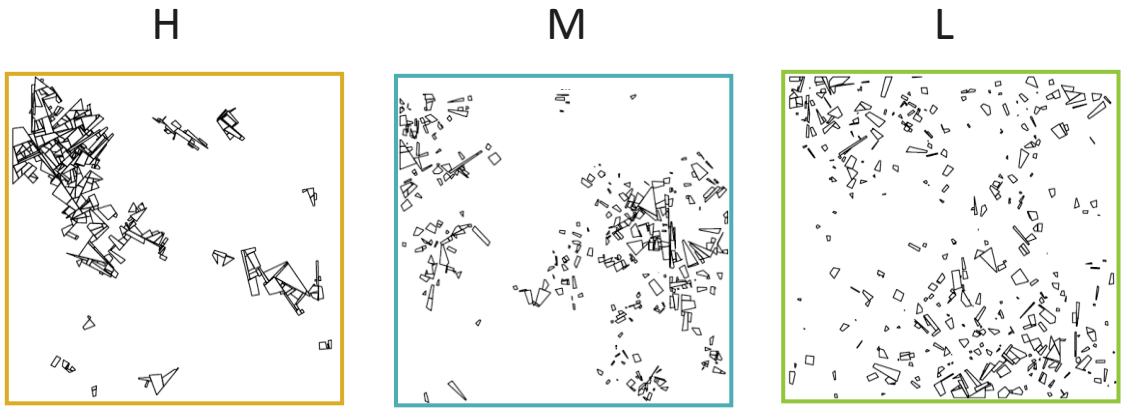


Figure 3: Examples of landscapes simulated using the process presented in Fig. 2 with parameters n=400, p=15. Three values of the aggregation parameter are used: H, d=1 (high aggregation); M, d=200 (medium aggregation); L, d=400 (low aggregation).

2.2. Epidemiological model

To simulate disease spread and management in landscapes, we used an existing stochastic, spatially explicit, SEIR (susceptible-exposed-infectious-removed) model (Pleydell et al. 2018; Rimbaud et al. 2018a, 2018b). The model is orchard-based, with a discrete time step of 1 week. At the beginning of the simulation, the trees in the patches are not infected: they are in the "susceptible" (i.e., healthy) state. The virus is introduced at the beginning of the first year of the simulation in one of the patches (defined by its connectivity quantile) and then spreads through orchards, causing changes in tree status: from "susceptible", they become "exposed" just after virus infection, "infectious hidden" (and symptomatic) after the end of the latent period, "infectious detected" after detection of the infected tree during surveys, and "removed" when the tree is removed from the patch. In addition, new introductions can also occur (with a specified probability) during the entire simulation at each patch plantation. Epidemic spread is governed by 6 epidemiological parameters (Table 1). Furthermore, a management strategy based on the French management of sharka in *Prunus* orchards is implemented as previously described (Rimbaud et al. 2018b). Briefly, a disease management strategy defined by 23 parameters (Fig. 1 and Supplementary Table S1) is applied after 5 years of epidemic to allow the spread of the virus. The model output is an economic criterion: the net present value (NPV), which corresponds to the sum of the gross margin (GM) calculated each year and updated by a discount rate (Rimbaud et al. 2018b). The GM represents the difference between the benefits generated by the cultivation of productive hosts and the costs induced by production and management actions (including surveillance, removal and replantation).

Table 1: Epidemiological parameters implemented in the model, and their variation ranges in simulations.

		Min	Max
q_k	Quantile of the connectivity of the patch of first introduction	0	1
φ	Probability of introduction at plantation	0.0046	0.0107
p_{MI}	Relative probability of massive introduction	0	0.1
W_{exp}	Expected value of the dispersal weighting variable	0.469	0.504
β	Transmission coefficient	1.25	1.39
θ_{exp}	Expected duration of the latent period (years)	1.71	2.14

2.3. Epidemic simulations and sensitivity analyses

2.3.1. Simulations with the French management

To study the influence of the landscape on productivity under the French management strategy, we performed simulations for all the landscapes described above. A realistic turnover of peach orchards was simulated on patches using a mean cultivation duration of 15 years (Rimbaud et al. 2018b). Simulations were run for 35 years (5 years without management and 30 years with management), which is a reasonable duration to assess the long-term impact of an epidemic in cultivated perennial plants. For each simulation, the 6 epidemiological parameters were drawn from uniform distributions using the bounds corresponding to sharka pathosystem (as described in Rimbaud et al. (2018b) and in Table 1) and management parameters representing the French management strategy (Fig. 1, JORF 2011). On the replicated real landscapes, 10,000 simulations were carried out on the three landscapes without subsampled patches, and 3,334 simulations were performed on each of the three replicates of the subsampled landscapes (to obtain a total of 10,000 simulations for each aggregation level). Likewise, on landscapes simulated by the T-tessellation algorithm, 334 simulations were performed for each of the 30 replicates.

2.3.2. Sensitivity analyses

The relative influence of epidemic and management parameters on disease impact was assessed for simulated sharka epidemics. For this purpose, Sobol's method for sensitivity analysis was used, which consists of: (i) the definition of target parameters and of their respective variation ranges; (ii) the generation of a numerical experimental design to explore parameter space; (iii) simulation; and (iv) the computation of Sobol's sensitivity indices which quantify the influence of each target parameter on the output variable (Faivre et al. 2013; Saltelli et al. 2008; Sobol 1993). The first-order sensitivity index of a parameter, noted SI_1, measures the main effect of this parameter whereas the total sensitivity index, noted SI_{tot}, also accounts for its interactions with other parameters. These indices are bounded by 0 and 1, a total index close to 0 meaning that the parameter has a negligible effect on the output variable.

Here, to get results specific to each level of patch aggregation, three sensitivity analyses were performed independently for the three simulated landscapes. We targeted 23 control parameters defining the implemented management strategy and 6 epidemiological parameters. Variation ranges were defined as their respective definition domain, possibly restricted using expert's opinion when this domain was infinite (Table 1, Supplementary Table S1; Rimbaud et al. 2018b). For each of the 30 landscapes (for each aggregation level), simulations were performed with 310,155 parameter

combinations generated with Sobol sequences (Sobol 1967, 1976). Then, Sobol's indices were calculated on the mean of the 30 replicates. First-order indices were estimated with the Sobol-Saltelli method (Saltelli et al. 2010; Sobol et al. 2007) whereas total indices were estimated with the Sobol-Jansen method (Jansen 1999; Saltelli et al. 2010).

2.3.3. Simulation of improved strategies

Using outputs of the sensitivity analyses we identified an improved strategy for each aggregation level. This improved strategy corresponds to the parameter combination leading to the highest NPV among the 310,155 combinations. We call these strategies "Best point H", "Best point M" and "Best point L" for each aggregation level (high, medium and low, respectively). Then, 10,000 simulations were performed with these three management strategies for the three aggregation levels as described in "Simulations with the French management" section.

We compared the mean NPV ($\overline{\mathrm{NPV}}$) and the lowest decile of the NPV (i.e., 10% of the NPV values are below the lowest decile, noted $\mathrm{NPV}_{10\%}$). This last criterion was chosen considering that farmers do not accept a management strategy which can too often lead to a low NPV. The purpose of this initial step was to assess whether a strategy that is efficient in a particular landscape remains efficient on landscapes with different characteristics. This provided an overview of the influence of landscapes features on management strategies.

2.4. Heuristic optimization of management strategies

The sensitivity analyses were carried out with 310,155 combinations of both epidemiological and management parameters. Thus, the three strategies "Best point H", "Best point M" and "Best point L" were selected because they were effective for one epidemic (characterized by the 6 epidemiological parameters). However, the other combinations of management parameter could have led to higher NPV with other epidemic parameters. Thus, we searched improved combinations of management parameters for various "epidemic cases". Each epidemic case corresponds to a set of different value ranges of the epidemiological parameters (example of one epidemic case: $q_k \in [0,0.25]$, $\beta \in [1.25,1.29]$, $\varphi \in [0.0046,0.0108]$, $p_{MI} \in [0,0.05]$, $W_{exp} \in [0.469,0.0175]$, $\theta_{exp} \in [1.71,1.925]$). The level of subdivision of the value ranges was based on the results of the sensitivity analyses, with more subdivisions for more influential epidemiological parameters. Then, each of the 310,155 parameter combinations was allocated to the corresponding epidemic case. Finally, for each epidemic case and each aggregation level, we identified the combination of management parameters leading to the highest NPV, and we performed 10,000 simulations with these strategies on the

corresponding landscape, while varying the epidemiological parameters within their respective variation ranges (Supplementary Table S1).

To finish, we selected the 10 parameter combinations corresponding to the highest $\mathrm{NPV}_{10\%}$ for each aggregation level and we performed 10,000 simulations of these strategies on the other landscapes. The strategies leading to the best $\mathrm{NPV}_{10\%}$ for each aggregation level are called respectively "Improved strategy H", "Improved strategy M" and "Improved strategy L".

3. RESULTS

3.1. Landscape organization influences the impact of management strategies

3.1.1. Landscape influence on productivity with the French management strategy

We performed simulations of epidemic spread on duplicated and simulated landscapes under the French management strategy (JORF 2011). In both cases, the NPV decreased for landscapes with increasing patch aggregation (Fig. 4).

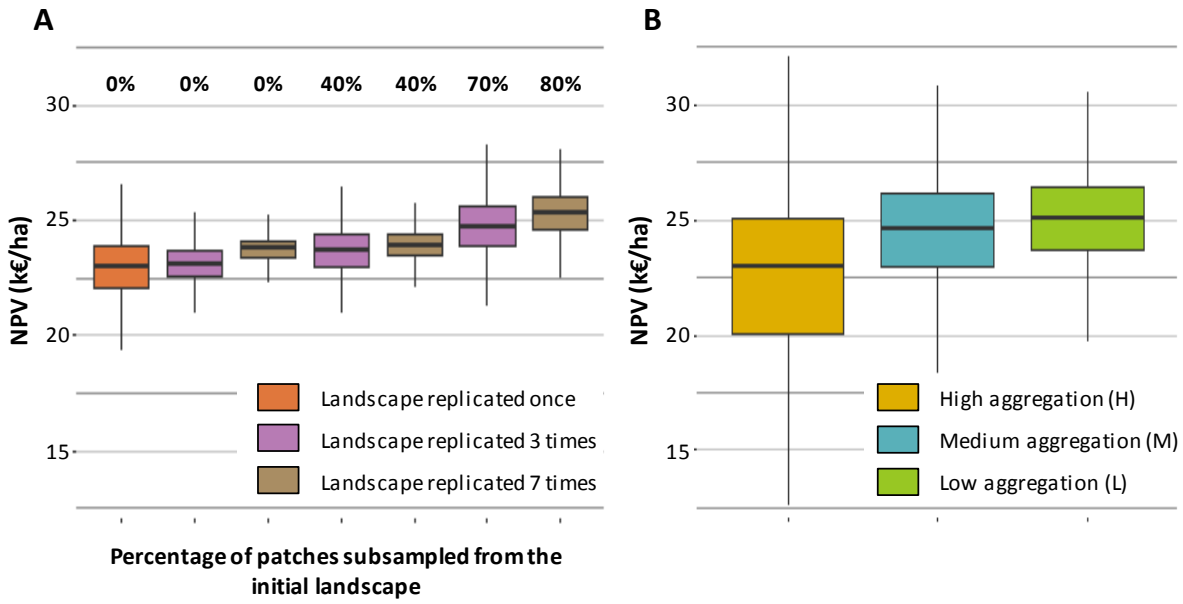


Figure 4: Distribution of the NPV for 10,000 simulations of sharka spread and management: A, on replicated landscapes; and B, on simulated landscapes.

In order to understand why the NPV is affected in landscapes with higher patch aggregation, we observed disease prevalence through time and studied the impact of patch aggregation on

components of the economic criterion for simulated epidemics (Supplementary Fig. S3). During the early years of the epidemic, the virus spreads faster in landscape H than in landscapes M and L; prevalence and incidence are therefore slightly higher. Thus, surveillance is strengthened and increases costs (inducing a lower GM). In addition, the increased number of removals leads to a decrease in the number of productive trees (the average number of productive trees per ha per year is respectively 553, 557 and 559 for H, M and L landscapes over the 30 years of the epidemic), which entails yield losses.

3.1.2. Landscape influence on sensitivity to model parameters

Three sensitivity analyses were performed for the three levels of patch aggregation on 23 management parameters and 6 epidemiological parameters in order to identify the most influential input parameters on the NPV (Fig. 5). We showed that 2 parameters related to plantation (χ_n: contamination threshold for an orchard in the neighborhood, above which the plantation of orchards is forbidden) and removals (χ_R: contamination threshold in the removal epicenter, above which orchards inside the removal zone are removed) have a strong influence on the NPV and this result does not depend on patch aggregation. The high impact of these parameters was likely due to a loss of productivity when the contamination threshold for the plantation bans and removal was too low.

However, although the two most influential parameters are the same for the three landscapes, their relative influence depends on landscape aggregation. For landscape H, the most influential contributors to the NPV were first the plantation ban threshold (χ_n; SI_{tot}=0.62) and then the removal threshold (χ_R; SI_{tot}=0.29). Conversely, for landscape L, the most influential contributors to the NPV were first the removal threshold (χ_R; SI_{tot}=0.45) and then the plantation ban threshold (χ_n; SI_{tot}=0.42). Overall, when the landscape is highly aggregated much of the variance is explained by a few parameters; conversely, when the landscape is less aggregated a larger number of parameters explain the variance observed in the simulations (Fig. 5).

To summarize, management parameters do not have the same influence on the economic criterion depending on the landscape. Optimal management parameters can therefore depend on landscape features.

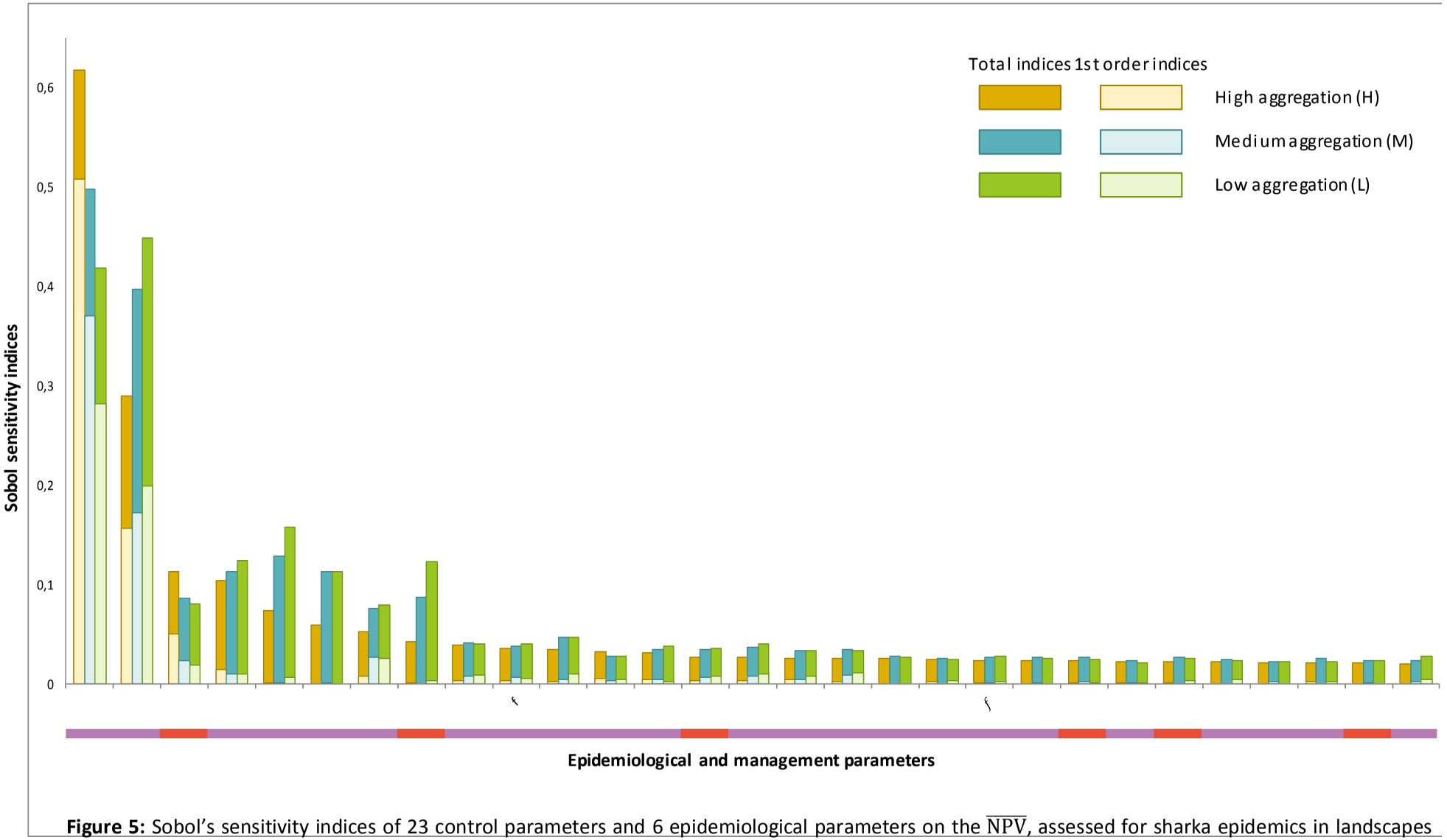


Figure 5: Sobol's sensitivity indices of 23 control parameters and 6 epidemiological parameters on the $\overline{\mathrm{NPV}}$, assessed for sharka epidemics in landscapes with 3 different aggregation levels (H, M and L). The line under the figure indicates the epidemiological parameters (red) and management parameters (purple).

3.1.3. Landscape influence on productivity for improved strategies

An improved strategy (i.e., the parameter combination resulting in the best NPV among the 310,155 tested combinations) was identified for each level of landscape aggregation, and named “Best point H”, “Best point M” and “Best point L”. For the three aggregation levels, these management strategies only very rarely involve orchard plantation bans, and only symptomatic trees are removed (and not entire orchards). In addition, surveillance zones (focal and security zones) are much smaller with these strategies than with the French management strategy (Supplementary Table S2), which reduces surveillance costs.

Then, simulations were carried out with these three strategies by varying epidemiological parameters on all the simulated landscapes (Fig. 6). Simulations performed on a landscape with the parameter combination identified for the same landscape lead to better $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{10\%}$ than with the French management strategy. Besides, these analyses show that a management strategy that is efficient for a landscape is not necessarily efficient in another, and can be less profitable than the French strategy. Indeed, the “Best point H” strategy is more profitable than the French strategy when it is applied on landscapes M and L; however, the “Best point M” and “Best point L” strategies are less profitable than the French management strategy for landscape H: $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{10\%}$ were largely lower (i.e., risk-taking is higher) than with the French strategy.

3.2. Landscape influence on improved management strategies

The sensitivity analyses show that 2 out of 6 epidemiological parameters have a high impact on the NPV (Fig. 5): $q_{\kappa,}$ the quantile of the connectivity of the patch of first introduction and β, the transmission coefficient. To define the epidemic cases (i.e., subsets of parameter values corresponding to similar epidemics), we divided the value ranges of these 2 parameters into four equal parts, and the 4 other epidemiological parameters (Φ, p_{MI}, W_{exp}, θ_{exp}) were divided into 2 equal parts. We obtained 4 x 4 x 2 x 2 x 2 x 2 = 256 epidemic cases for each level of landscape aggregation, and for each case we identified the combination of management parameters leading to the highest NPV. The majority of these strategies does not involve orchard plantation bans (in 85% of the cases for landscape H, and 89% for landscapes M and L) and does not impose removal of entire orchards (in 68% of the cases for landscape H, and 74% for landscapes M and L). In addition, surveillance zones are again much smaller for these strategies than for the French management.

For each aggregation level, (i) simulations were carried out with the corresponding 256 strategies, and (ii) the 10 parameter combinations resulting in the best $\mathrm{NPV}_{10\%}$ were retained. Simulations were then performed with these 30 combinations on all the landscapes (Supplementary Fig. S4). We observe that the impact of these management strategies is more important for landscape H than for landscapes M and L. Indeed, the $\mathrm{NPV}_{10\%}$ for landscape H varies between 15,945 €/ha and 22,987 €/ha with this 30 management strategies, between 23,110 €/ha and 24,202 €/ha for landscape M and between 23,111 €/ha and 24,616 €/ha for landscape L. The strategies leading to the best $\mathrm{NPV}_{10\%}$ ("Improved strategy H", "Improved strategy M" and "Improved strategy L") are detailed in Supplementary Table S2 and Supplementary Fig. S5. We note that the strategy leading to the best $\mathrm{NPV}_{10\%}$ is also the strategy leading to the best $\overline{\mathrm{NPV}}$ (for landscape L) or leading to a very close value to the best $\overline{\mathrm{NPV}}$ (for landscapes H and M).

Finally, we compared the NPV from simulations without management, with the French management strategy, with the three "Best point" strategies, and with the three "Improved strategies" leading to the best $\mathrm{NPV}_{10\%}$ (Fig. 6). We could find substantially improved NPV for the three levels of patch aggregation. For instance, the $\mathrm{NPV}_{10\%}$ is 17,652 €/ha with the French management strategy, 20,474 €/ha with strategy "Best point H" and 22,987 €/ha with strategy "Improved strategy H". In addition, although each landscape has a specific improved strategy, the "Improved strategy H" could be, according to the model, an acceptable compromise for all landscapes (Fig. 7). Indeed, application of the "Improved strategy H" on landscapes M and L (instead of their respective "Improved strategies") leads to a reduction of only 184 €/ha and 640 €/ha in $\mathrm{NPV}_{10\%}$ over 30 years.

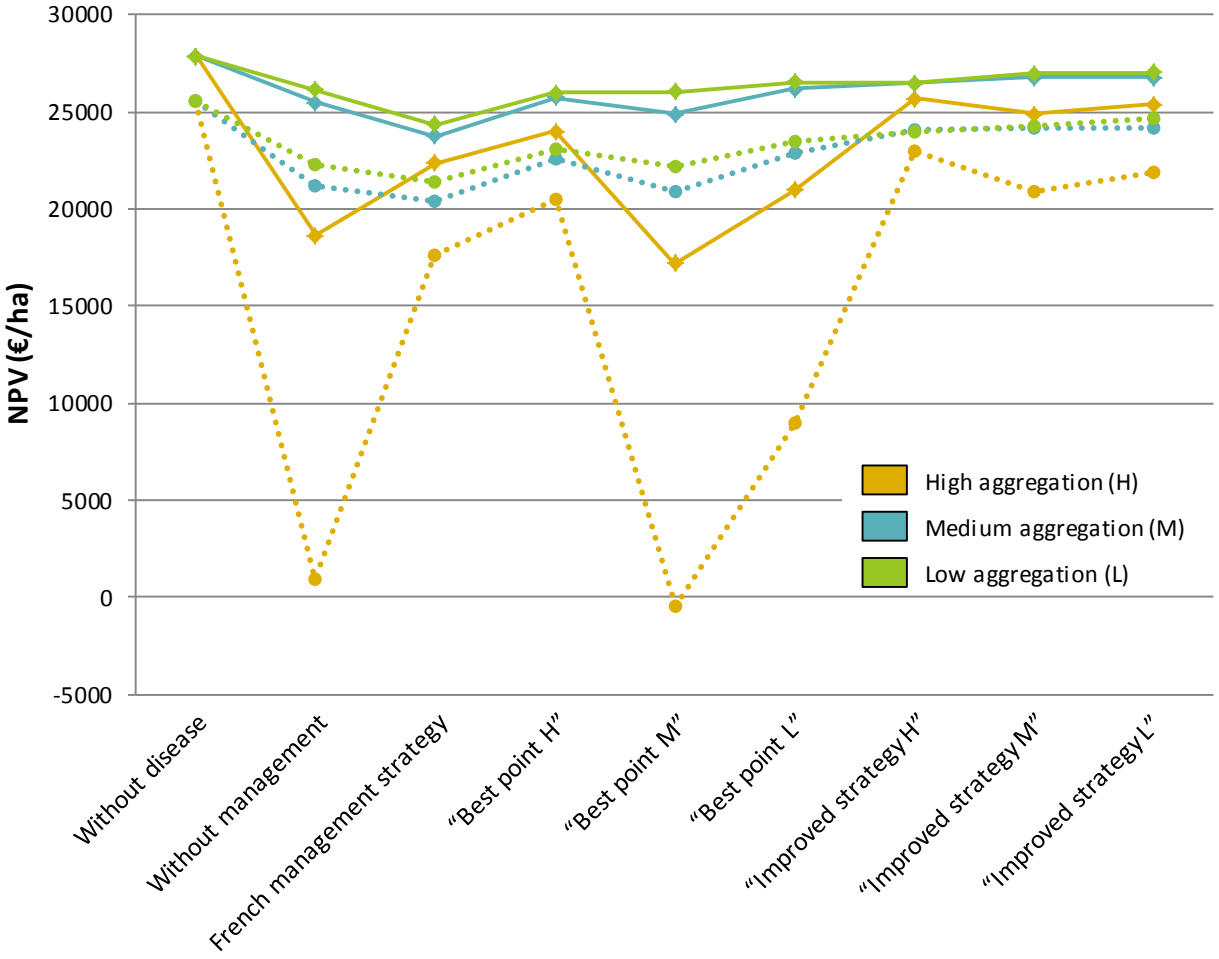
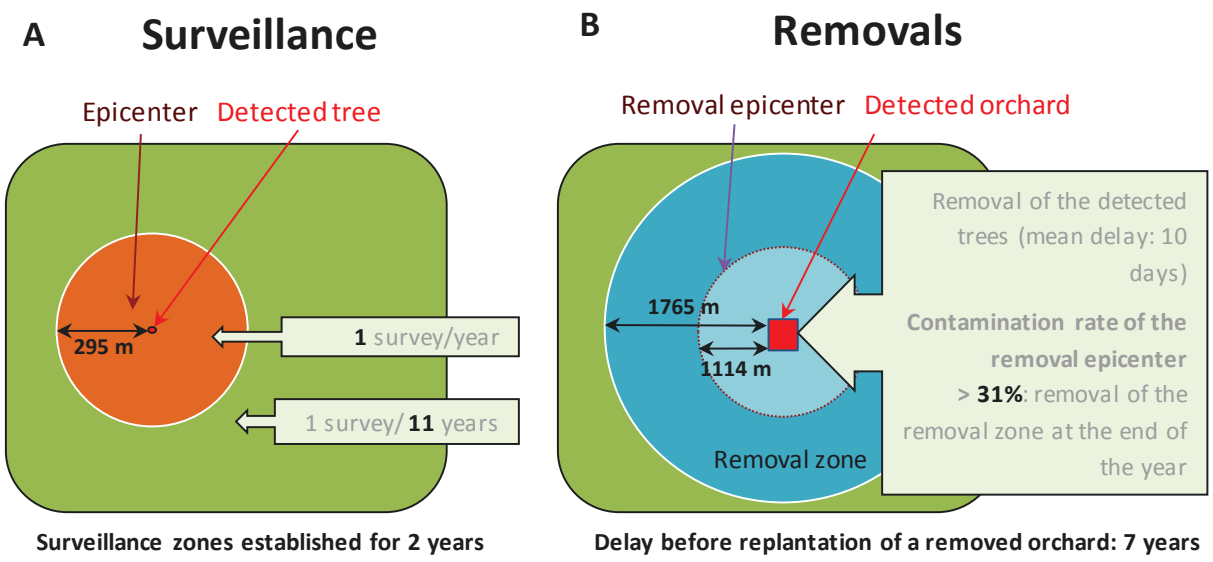


Figure 6: $\overline{\mathrm{NPV}}$ (solid lines) and $\mathrm{NPV}_{10\%}$ (dotted lines) obtained after simulations of PPV dispersal and management.



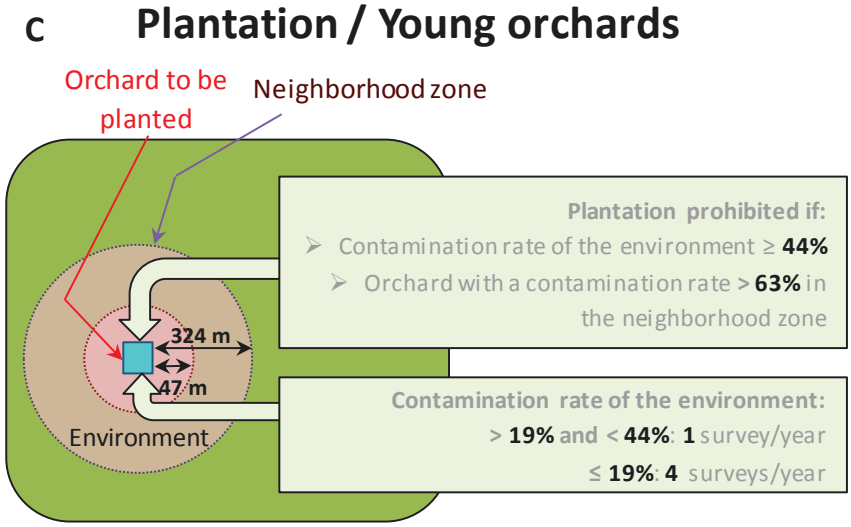


Figure 7: Management actions for the "Improved strategy H" (leading to the best lowest decile for the most aggregated landscape).

4. Discussion

This work aimed to understand how patch aggregation influences disease spread and the impact of control options. Simulations of disease spread and management within a sensitivity analysis framework showed that the landscape influences the profitability of different strategies for sharka control in peach orchards. In addition, the results of these sensitivity analyses were exploited to identify efficient strategies (more profitable than the present French management of sharka). These strategies are efficient for a specific aggregation level, but we also identified a generic strategy, namely the "Improved strategy H", that is efficient for various levels of landscape aggregation.

4.1. Influence of landscape in modeling studies

Our study shows the importance of taking landscape characteristics into account in the design and optimization of disease management strategies. First, we show that landscape aggregation influences sharka dispersal: in our simulations, profitability (NPV) increases with the distance between patches, both without sharka management or under the French management strategy (Fig. 6). In addition, we show that landscape aggregation influences the impact of management strategies, both because the relative influence of the management parameters on the NPV depends on landscape aggregation (Fig. 5), and because a management strategy which is efficient for a landscape is not necessarily efficient for the other landscapes (Fig. 6). This demonstration that the efficiency of a disease management strategy depends on landscape aggregation has important consequences for the improvement of management strategies (or, maybe less realistically, for the optimization of the landscape itself).

This result also means that such studies must be based on either real or realistic landscapes. However, as pointed out in the Introduction, generic realistic landscapes are rarely considered in epidemiological modeling studies. Because generic conclusions cannot be drawn on a single real landscape, it was important to simulate landscapes with a specified level of aggregation. Thus, we devised an algorithm based on T-tessellations to generate landscapes composed of various patches (with realistic enough shapes and sizes) that are more or less aggregated. Disease dispersal and the impact of control options might also be influenced by other landscape structures such as mountains, lakes, rivers, forests or roads (Brunker et al. 2018), species composition or proportion of suitable habitat (Ostfeld et al. 2005), including the proportion of resistant *vs*. susceptible hosts (Papaïx et al. 2014). Here we chose to focus on patch aggregation, but other landscape features might enter such models in the future if their epidemiological and economic impact is properly estimated.

4.2. *In silico* improvement of disease management

In the second part of the present study, for each level of landscape aggregation we searched improved management strategies. This was challenging since we attempted to improve a complex strategy including 23 management parameters (epidemiological modeling studies generally optimize one or two parameters at a time). To succeed, we used the results of sensitivity analyses for which numerous parameter combinations were tested. In addition, contrary to previous studies that pursued the same goal using an epidemiological criterion (Cunniffe et al. 2014, 2015b; Filipe et al. 2012; Parnell et al. 2009, 2010, 2012, 2104; Sisterson and Stenger 2012; Chan and Jeger 1994; Jeger and Chan 1995), here we balanced all costs and benefits of disease management strategies within an economic criterion (Rimbaud et al. 2018b), which is important when several parameters expressed in different units are co-optimized. Furthermore, modeling studies generally aim to improve the mean of the criterion to optimize and do not take into account the level of risk aversion of decision-makers. However, as shown by Cunniffe et al. (2015b, 2016), the optimal strategy can depend on which percentile of a criterion is optimized. Because decision-makers generally tend to minimize the risk of devastating scenarios, here we searched efficient strategies on the basis of the lowest decile of the NPV (i.e., $\mathrm{NPV}_{10\%}$). Improving this criterion allows to select management strategies that limit the proportion of epidemics causing substantial economic damage.

For different levels of patch aggregation, we identified with our simulations different improved management strategies ("Improved strategy H", "Improved strategy M" and "Improved strategy L"). Applying these strategies on the respective landscapes, we obtain better $\overline{\mathrm{NPV}}$ (as well as $\mathrm{NPV}_{10\%}$) than with previously improved strategies (Rimbaud et al. 2018b) (Supplementary Fig. S6). It may be due to the fact that these previous strategies were improved for a unique landscape and lacked robustness to changes in landscape aggregation. In addition, the number of simulations performed for each strategy may influence the results. Here, we selected 256 candidate management strategies for which we carried out 10,000 simulations where the epidemiological parameters vary, and we selected the strategy associated with an accurate estimate of the best $\mathrm{NPV}_{10\%}$. In their study, Rimbaud et al. (2018b) performed only 30 simulations for each of 310,155 random management strategies and (i) they isolated the parameter combination associated with the highest estimated $\overline{\mathrm{NPV}}$ ("Best-value strategy") and (ii) they performed a marginal optimization using the mode of the distribution of each parameter for the combinations associated with the best 1% values of NPV ("Best-percent strategy"). Our own attempt to perform such marginal optimization (not shown) failed to produce good NPV values, probably because the substantial interactions between management parameters (Fig. 5) are ignored by this approach.

The results of our heuristic optimization mean that, in theory, management could be tailored to each landscape. However, in practice, stakeholders may struggle to delineate zones that differ by their level of landscape aggregation, and to apply different strategies within the territory where they are involved. In addition, landscapes change through time, which means that strategies that are too specific to a given level of aggregation may become obsolete. Thus, such landscape-specific strategies may only be applicable when production areas with very different levels of landscape aggregation are distant enough. A practically useful alternative to such landscape-specific strategies is the identification of a robust, one-size-fits-all, strategy which could be an efficient compromise for all the landscapes. This is the case for the "Improved strategy H", which may thus be applied at a wide scale. This strategy could be interesting for stakeholders because it is both more profitable and simpler to implement than the present French management strategy. Indeed, it requires surveillance of small areas around each detected tree, very rarely involves orchard plantation bans, and almost never imposes the removal of entire orchards (we note that the last two points correspond to the most influential parameters in the sensitivity analyses).

This work is relevant to stakeholders because it shows that both landscape-specific and landscape-generic disease management strategies can be identified and improved *in silico*. Indeed, the current strategy applied in France on 11,045 ha of peach orchard (Agreste 2013) reduces economic losses in case of severe sharka epidemics, but according to our simulations on average 36 million euros could be saved by using the "Improved strategy H" over a period of 30 years for landscape H (24 million euros for landscape L), and 59 million euros for the lowest decile of the NPV (29 million for landscape L).

However, as previously mentioned (Rimbaud et al. 2018b) our results can be affected by some model assumptions (for instance, the detection probability may be overestimated). In addition, we used here a Sobol-type sensitivity analysis to improve management strategies. Although this analysis has good space-filling properties that enabled to test a huge number of parameter combinations well spread throughout the parameter space (Sobol 1976), this one is so vast that better strategies can be found between the sampled points. The main goal of the present study was to explore the impact of landscape aggregation on improved disease management strategies, but if interest lies in approaching more closely the actual optimum, one option may be to iteratively explore the parameter space as previously done (Rimbaud et al. 2018b). However, this approach involves some arbitrary choices at each iteration and inefficiencies in the use of computing resources; thus, dedicated optimization algorithms may be more efficient for future work.

ACKNOWLEDGMENTS

The authors thank Julien Papaïx for advising on landscape simulations, and Sylvie Dallot for helpful feedback throughout the work. This work was supported by the CIRAD-UMR AGAP HPC Data Center of the South Green Bioinformatics platform (http://www.southgreen.fr), and we are grateful to Sébastien Ravel for the associated help. This work is supported by an INRA/ANSES scholarship to CP, and by INRA (BEcOSMASH project, funded by the SMaCH Metaprogram and the SAE2 Department) and Agropolis Fondation (E-SPACE project).

LITERATURE CITED

Agreste. 2013. Inventaire des vergers 2013. http://agreste.agriculture.gouv.fr/enquetes/productions-vegetales-528/vergers-et-fruits/

Brunker, K., Lemey, P., Marston, D. A., Fooks, A. R., Lugelo, A., Ngeleja, C., Hampson, K., and Biek, R. 2018. Landscape attributes governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs. Mol. Ecol. 27(3):773-788.

Cambra, M., Capote, N., Myrta, A., and Llácer, G. 2006. Plum pox virus and the estimated costs associated with sharka disease. EPPO Bull. 36:202–204.

Chan, M. S., and Jeger, M. J. 1994. An analytical model of plant virus disease dynamics with roguing and replanting. J. Appl. Ecol. 31:413–427.

Cunniffe, N. J., Laranjeira, F. F., Neri, F. M., DeSimone, R. E., and Gilligan, C. A. 2014. Cost-effective control of plant disease when epidemiological knowledge is incomplete: modelling Bahia Bark Scaling of Citrus. PLoS Comput. Biol. 10:e1003753.

Cunniffe, N. J., Koskella, B., E. Metcalf, C. J., Parnell, S., Gottwald, T. R., and Gilligan, C. A. 2015a. Thirteen challenges in modelling plant diseases. Epidemics 10:6–10.

Cunniffe, N. J., Stutt, R. O. J. H., DeSimone, R. E., Gottwald, T. R., and Gilligan, C. A. 2015b. Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. PLoS Comput. Biol. 11:e1004211.

Cunniffe, N. J., Cobb, R. C., Meentemeyer, R. K., Rizzo, D. M., and Gilligan, C. A. 2016. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc. Natl. Acad. Sci. USA 113:5640–5645.

Faivre, R., Ioos, B., Mahévas, S., Makowski, D., and Monod, H. 2013. Analyse de sensibilité et exploration de modèles. Editions Quae, Versailles, France.

Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature. 413:542–548.

Filipe, J. A., Cobb, R. C., Meentemeyer, R. K., Lee, C. A., Valachovic, Y. S., Cook, A. R., Rizzo D. M., and Gilligan C.A. 2012. Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden Oak Death in Northern Californian forests. PLoS Comput Biol. 8:e1002328.

Jansen M. J. W. 1999. Analysis of variance designs for model output. Comput. Phys. Commun. 117(1–2):35-43.

Jeger, M. J., and Chan, M. S. 1995. Theoretical aspects of epidemics: uses of analytical models to make strategic management decisions. Can. J. Plant Pathol. 17:109–114.

JORF 2011. Arrêté du 17 mars 2011 relatif à la lutte contre le Plum Pox Virus, agent causal de la maladie de la Sharka, sur les végétaux sensibles du genre Prunus. République Française. JORF n°0067 du 20 mars 2011. Paris, France. http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023728913&dateTexte=&categorieLien=id.

Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G., and Grenfell, B. T. 2003. Modelling vaccination strategies against foot-and-mouth disease. Nature. 421:136–142.

Keeling, M. J., and Rohani, P. 2008. Modeling Infectious Diseases in Humans and Animals. Princeton University Press, New Jersey, USA.

Mikaberidze, A., Mundt, C. C., and Bonhoeffer, S. 2016. Invasiveness of plant pathogens depends on the spatial scale of host distribution. Ecol. Appl. 26:1238–1248.

Németh, M. 1986. Virus, mycoplasma and Rickettsia diseases of fruit trees. Akadémiai, Kaidó, Budapest, Hungary.

Ostfeld, R. S., Glass, G. E., and Keesing, F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol. Evol. 20:328–336.

Papaïx, J., Adamczyk-Chauvat, K., Bouvier, A., Kiêu, K., Touzeau, S., Lannou, C., and Monod H. 2014. Pathogen population dynamics in agricultural landscapes: The Ddal modelling framework. Infect. Genet. Evol. 27:509–520.

Parnell, S., Gottwald, T. R., Van den Bosch, F., and Gilligan, C. A. 2009. Optimal strategies for the eradication of Asiatic Citrus Canker in heterogeneous host landscapes. Phytopathology. 99:1370–1376.

Parnell, S., Gottwald, T. R., Gilligan, C. A., Cunniffe, N. J., and Van den Bosch, F. 2010. The effect of landscape pattern on the optimal eradication zone of an invading epidemic. Phytopathology. 100:638–644.

Parnell, S., Gottwald, T. R., Irey, M. S., Luo, W., and Van den Bosch, F. 2011. A stochastic optimization method to estimate the spatial distribution of a pathogen from a sample. Phytopathology. 101:1184–1190.

Parnell, S., Gottwald, T. R., Gilks, W. R., and Van den Bosch, F. 2012. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. J. Theor. Biol. 305:30–36.

Parnell, S., Gottwald, T. R., Riley, T., and Van den Bosch, F. 2014. A generic risk-based surveying method for invading plant pathogens. Ecol. Appl. 24:779–790.

Parry, M., Gibson, G. J., Parnell, S., Gottwald, T. R., Irey, M. S., Gast, T. C., and Gilligan, C. A. 2014. Bayesian inference for an emerging arboreal epidemic in the presence of control. Proc. Natl. Acad. Sci. USA 111:6258–6262.

Pleydell, D. R. J., Soubeyrand, S., Dallot, S., Labonne, G., Chadœuf, J., Jacquot, E., and Thébaud G. 2018. Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLoS Comput. Biol. 14(4): e1006085.

Rimbaud, L., Dallot, S., Gottwald, T., Decroocq, V., Jacquot, E., Soubeyrand, S., and Thébaud G. 2015. Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annu. Rev. Phytopathol. 53:357–378.

Rimbaud, L., Bruchou, C., Dallot, S., Pleydell, D. R. J., Jacquot, E., Soubeyrand, S., and Thébaud, G. 2018a. Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. R. Soc. Open Sci. 5:171435.

Rimbaud, L., Dallot, S., Bruchou, C., Thoyer, S., Jacquot, E., Soubeyrand, S., and Thébaud G. 2018b. Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. bioRxiv 315747. (doi: https://doi.org/10.1101/315747)

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana M. and Tarantola S. 2008. Global sensitivity analysis: the Primer. Wiley.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181:259–270.

Sisterson, M. S., and Stenger, D. C. 2012. Roguing with replacement in perennial crops: conditions for successful disease management. Phytopathology. 103:117–128.

Sobol, M. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. 7(4):86-112.

Sobol, M. 1976. Uniformly distributed sequences with an additional uniform property. USSR Comput. Math. Math. Phys. 16(5):236-242.

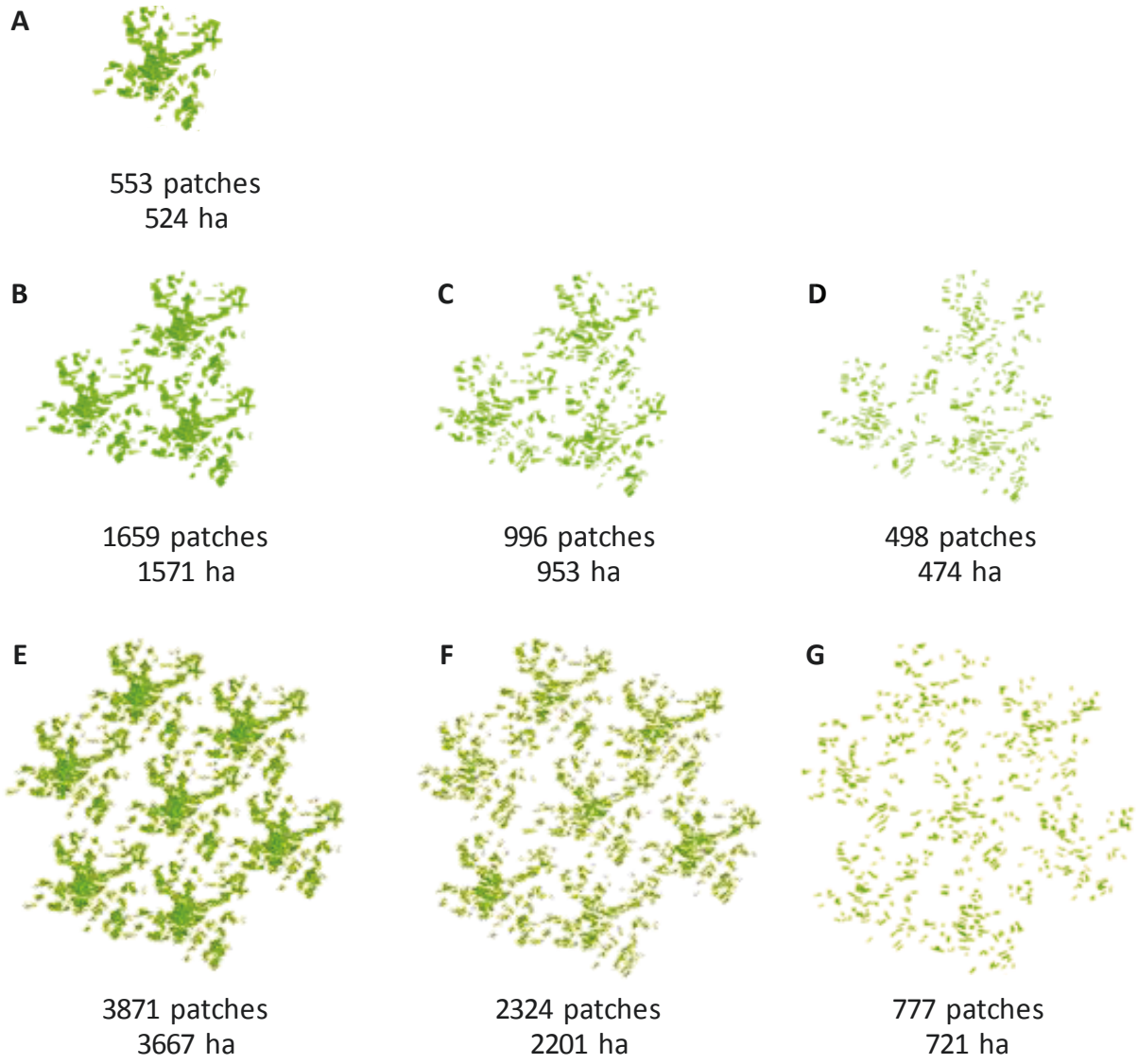
Sobol, M. 1993. Sensitivity analysis for non-linear mathematical models. Math. Modell. Comput. Exp. 1:407-414.

Sobol M., Tarantola, S., Gatelli, D., Kucherenko, S. S., and Mauntz, W. 2007. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliab. Eng. Syst. Safe 92:957–960.

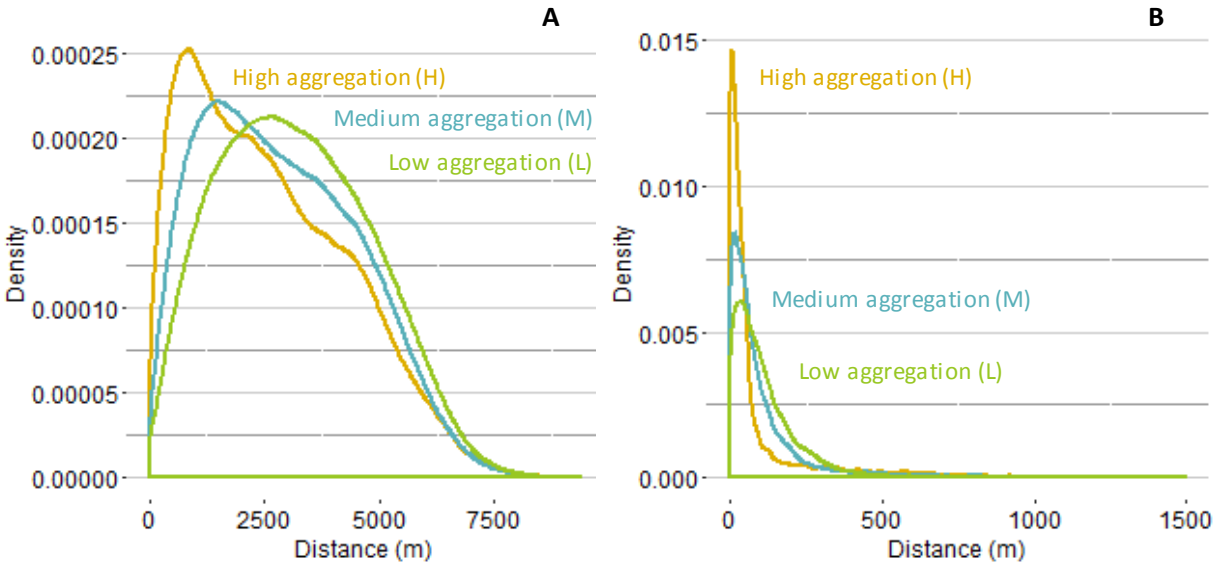
Soubeyrand, S., Held, L., Höhle, M., and Sache, I. 2008. Modelling the spread in space and time of an airborne plant disease. J. R. Stat. Soc. Ser. C Appl. Stat. 57:253–272.

Soubeyrand, S., Jerphanion, P., Martin, O., Saussac, M., Manceau, C., Hendrikx, P., and Lannou, C. 2018. Inferring pathogen dynamics from temporal count data: the emergence of Xylella fastidiosa in France is probably not recent. New Phytol. (doi:10.1111/nph.15177)

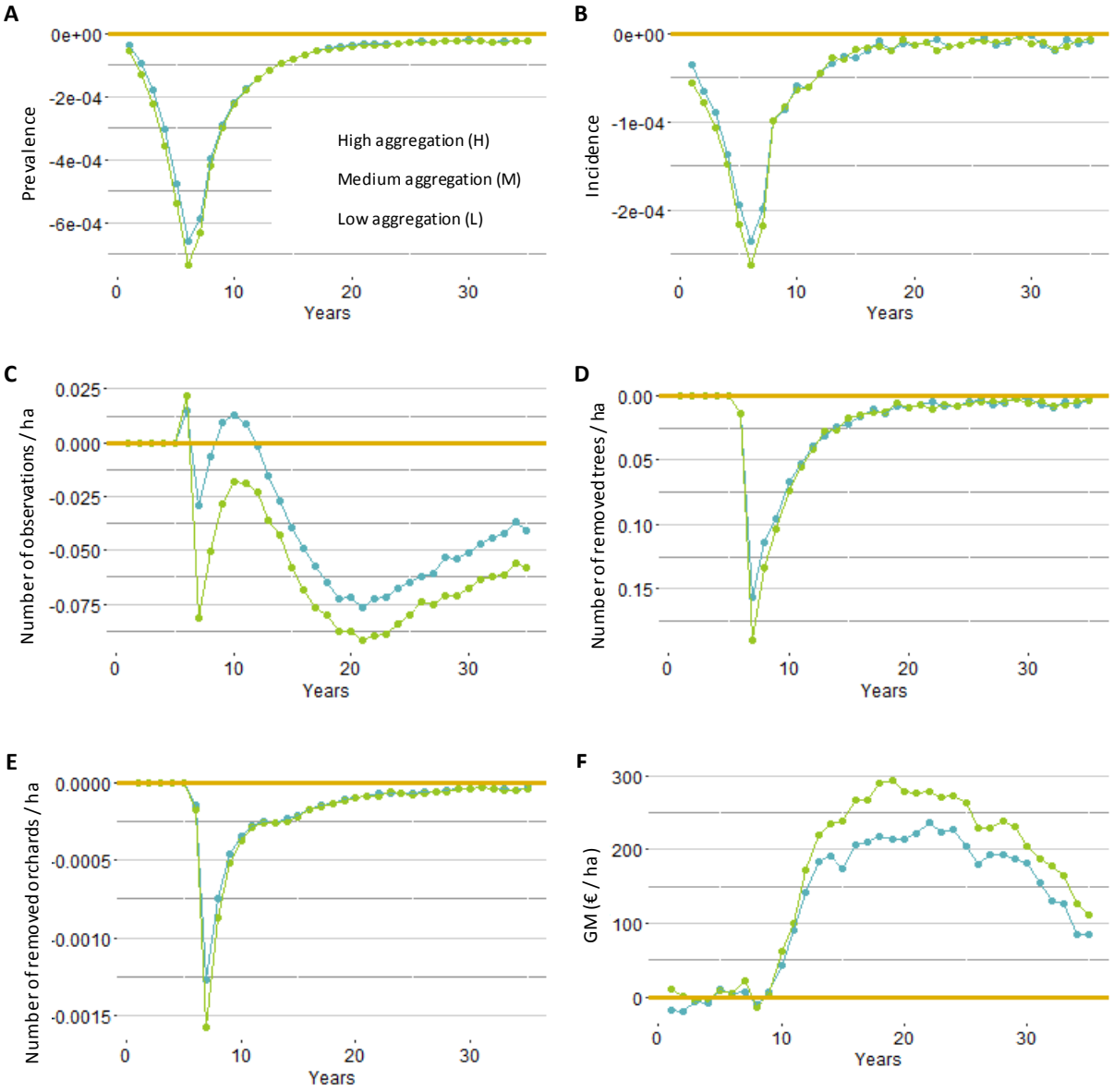
SUPPORTING INFORMATION



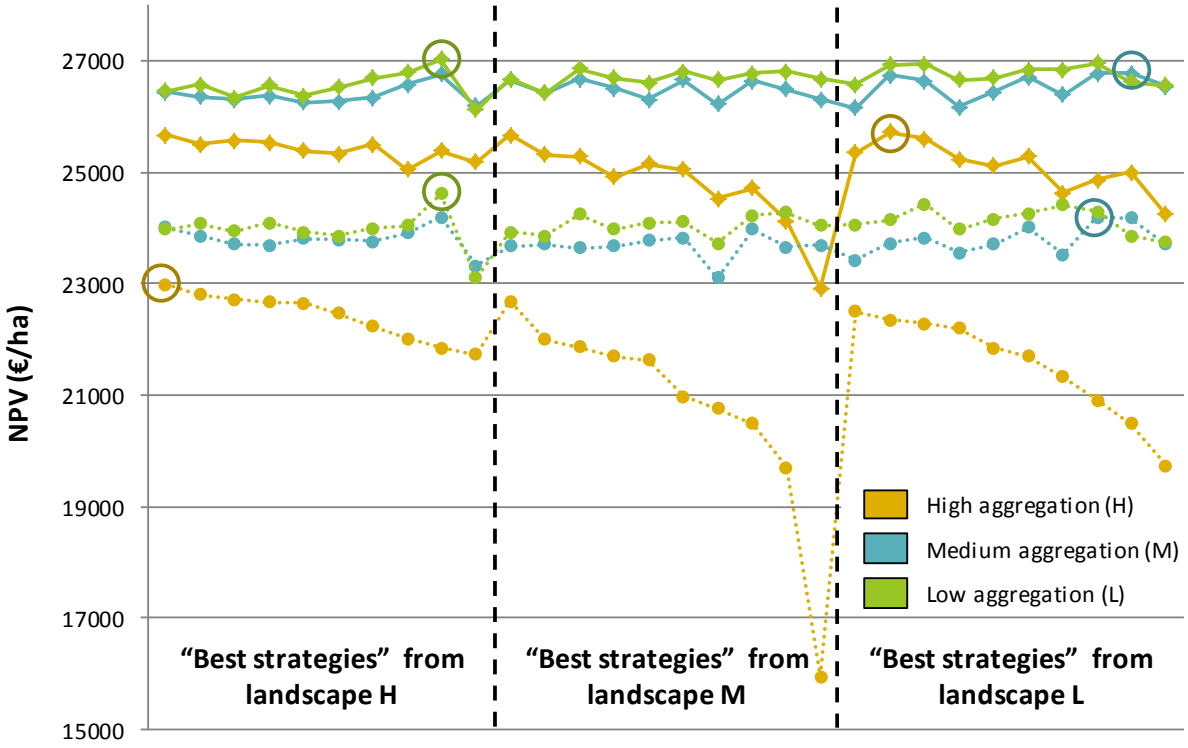
Supplementary Fig. S1: Duplicated and subsampled landscapes. A, real landscape of 553 peach patches (green polygons). B, real landscape duplicated 3 times. C and D, examples of landscapes obtained after the random removal of 40% and 70% of the patches from landscape B, respectively. E, real landscape duplicated 7 times. F and G, examples of landscapes obtained after the random removal of 40% and 80% of the patches from landscape E, respectively.



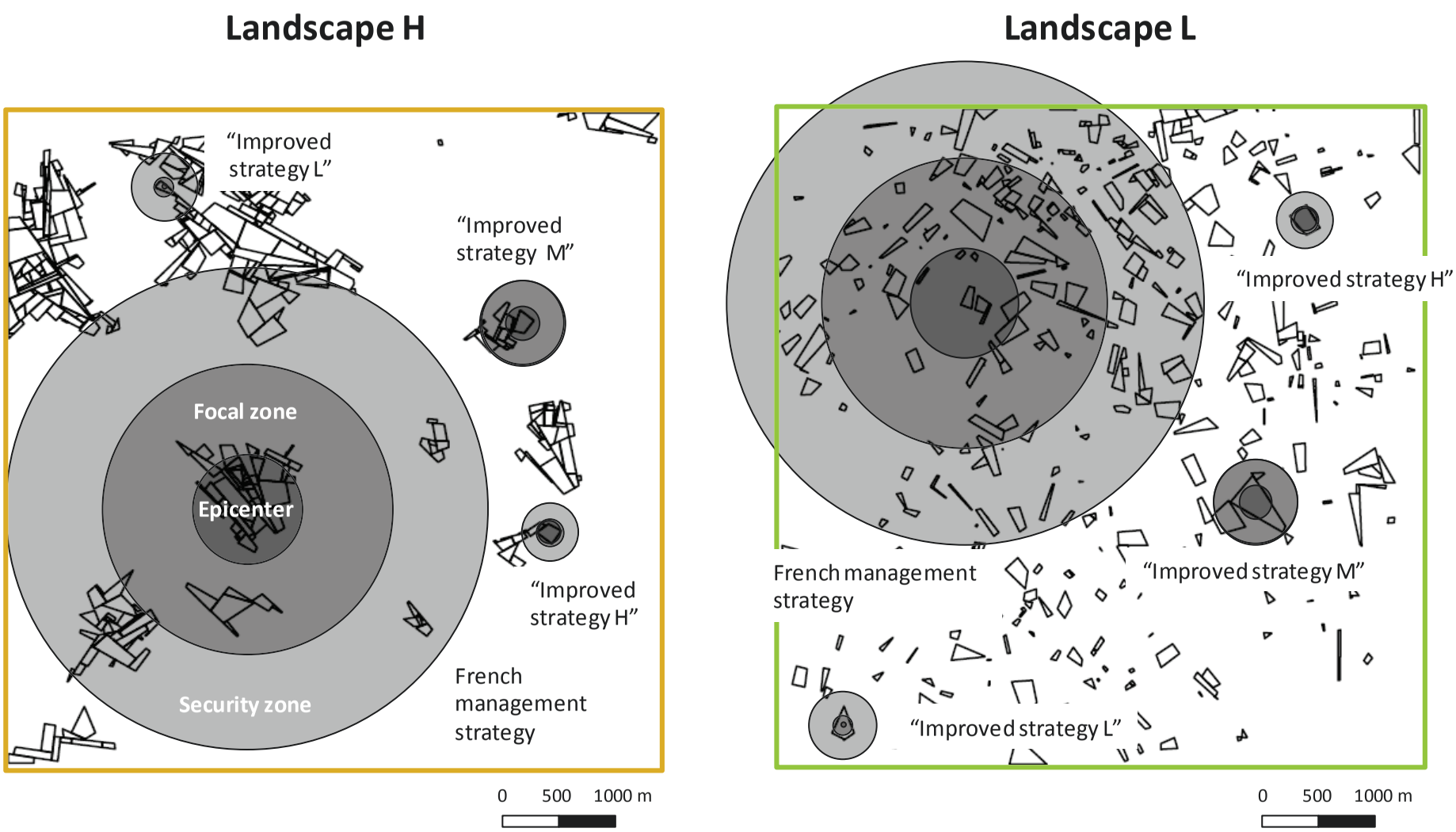
Supplementary Fig. S2: Probability densities of the distances for landscapes H, M and L between the centroids of: A, all the patches and B, all nearest neighbor patches.



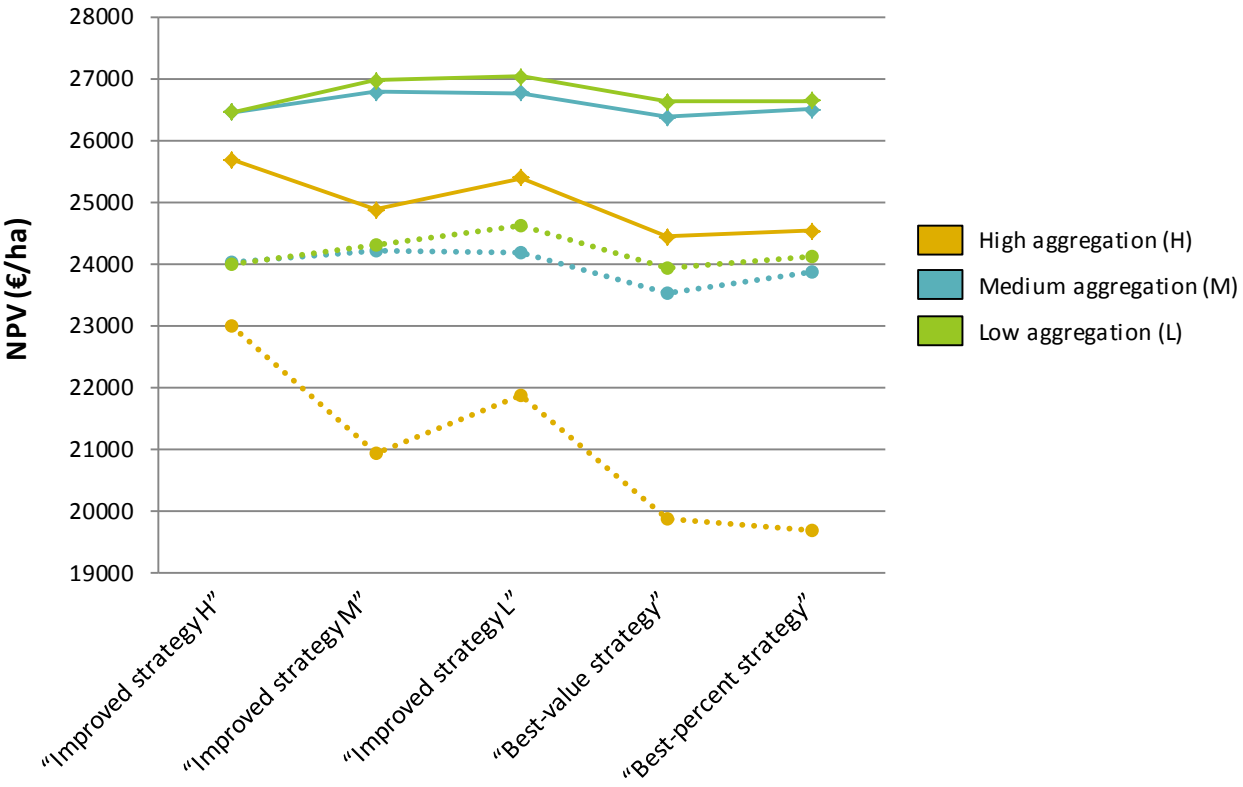
Supplementary Fig. S3: Evolution of NPV components on simulated landscapes with the French management strategy. These components are: A, prevalence; B, incidence; C, number of observations per ha; D, number of removed trees per ha; E, number of removed orchards per ha; F, gross margin (€/ha). For each component, the blue (resp. green) line represents the difference between landscape M (resp. L) and landscape H. Note, these lines are above the yellow line when the NPV component is higher than for landscape H (and vice versa).



Supplementary Fig. S4: $\overline{\mathrm{NPV}}$ (solid lines) and $\mathrm{NPV}_{10\%}$ (dotted lines) obtained after simulation of PPV dispersal and its management. Simulations were performed with the 30 improved combinations of management parameters (10 strategies leading to the best $\mathrm{NPV}_{10\%}$ among 256 strategies identified for the 3 aggregation levels), on the 3 levels of landscape aggregation. The circled points represent the best values of $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{10\%}$ for each aggregation level. The strategies corresponding to the best $\mathrm{NPV}_{10\%}$ are called "Improved strategy H" (high aggregation level), "Improved strategy M" (medium aggregation level) and "Improved strategy L" (low aggregation level).



Supplementary Fig. S5: Schematic representation of the surveillance zones on landscapes H (high aggregation level) and L (low aggregation level). Four management strategies are represented : the French strategy and the 3 "Improved strategies".



Supplementary Fig. S6: $\overline{\mathrm{NPV}}$ (solid lines) and $\mathrm{NPV}_{10\%}$ (dotted lines) obtained after simulations of PPV dispersal and management. Simulations are carried out with the 3 "Improved strategies" and with the strategies improved by Rimbaud et al. (2018b). Simulations are performed for 3 levels of landscape aggregation.

Supplementary Table S1: Epidemiological and management parameters implemented in the previously developed model with minimum and maximum values corresponding to the variation range of each parameter in the sensitivity analysis.

		Min	Max
	Epidemiological parameters		
q_K	Quantile of the connectivity of the patch of first introduction	0	1
φ	Probability of introduction at plantation	0.0046	0.017
p_{MI}	Relative probability of massive introduction	0	0.1
W_{exp}	Expected value of the dispersal weighting variable	0.469	0.504
β	Transmission coefficient	1.25	1.39
θ_{exp}	Expected duration of the latent period (years)	1.71	2.14
	Management parameters		
ρ	Probability of detection of a symptomatic tree	0	0.66
δ	Mean delay before removal of a detected tree (days)	-	-
Y_R^T	(Boolean) Individual trees are removed: 0: after a mean delay of 10 days 1: at the end of the year	0	1
Y_R	(Boolean) Whole orchards are removed: 0: after a mean delay of 10 days 1: at the end of the year	0	1
γ_S	Delay before replantation of a removed orchard (years)	0	10
γ_o	Duration of observation zones (years)	0	10
γ_y	Duration of young orchards (years)	0	10
ζ_s	Radius of security zones (m)	0	5800
$r\zeta_f$	Ratio of the focal area over the security area	0	1
$r\zeta_{eO}$	Ratio of the observation epicenter area over the focal area	0	1
ζ_n	Radius of the close neighborhood (m)	0	5475
ζ_R	Radius of the removal zone (m)	0	5800
$r\zeta_{eR}$	Ratio of the removal epicenter area over the removal area	0	1
$1/\eta_0$	Maximal period between 2 observations (years)	1	15
η_s	Observation frequency in security zones ($year^{-1}$)	0	8
η_f	Observation frequency in focal zones ($year^{-1}$)	0	8
η_{f*}	Modified observation frequency in focal zones ($year^{-1}$)	0	8
η_y	Observation frequency in young orchards ($year^{-1}$)	0	8
η_{y*}	Modified observation frequency in young orchards ($year^{-1}$)	0	8
χ_o	Contamination threshold in the observation epicenter, above which the observation frequency in focal zone is modified	0	1
χ_y	Contamination threshold in the environment around young orchards, above which the plantation of orchards is forbidden	0	1
$r\chi_{y*}$	Ratio (over $\chi\bar{y}$) of the contamination threshold in the environment, above which the observation frequency in young orchards is modified	0	1
χ_n	Contamination threshold on an orchard in the neighborhood, above which the plantation of orchards is forbidden	0	1
χ_R	Contamination threshold in the removal epicenter, above which orchards inside the removal zone are removed	0	0.34

Supplementary Table S2: Parameter combinations for the main management strategies.

Management parameters	French management strategy	"Best point strategies"			"Improved strategies"		
		High aggregation (H)	Medium aggregation (M)	Low aggregation (L)	High aggregation (H)	Medium aggregation (M)	Low aggregation (L)
ρ	0.66	0.47	0.06	0.53	0.65	0.30	0.48
Y_R^T	0	0	0	0	0	0	0
Y_R	1	1	0	1	1	1	0
γ_S	0	8	4	3	7	6	5
γ_o	3	10	0	0	2	9	1
γ_y	3	5	1	8	1	6	0
ζ_s	2500	252	461	3941	295	442	349
$r\zeta_f$	0.38	0.25	0.06	0.54	0.23	0.97	0.08
$r\zeta_{eO}$	0.60	0.51	0.83	0.49	0.68	0.06	0.16
ζ_n	200	4316	1437	4645	324	1592	2974
ζ_R	0	1309	2322	3372	1765	51	974
$r\zeta_{eR}$	0	0.91	0.03	0.87	0.63	0.88	0.89
$1/\eta_0$	6	8	15	10	11	9	7
η_s	1	2	4	3	1	0	0
η_f	2	6	4	2	0	1	3
η_{f*}	3	8	7	1	1	0	7
η_y	2	0	2	0	4	0	5
η_{y*}	3	5	5	6	1	4	8
χ_o	0.02	0.81	0.14	0.35	0.60	0.95	0.59
χ_y	0.02	0.75	0.15	0.56	0.44	0.29	0.25
$r\chi_{y*}$	0.50	0.15	0.18	0.80	0.44	0.56	0.17
χ_n	0.05	0.99	0.26	0.76	0.63	0.60	0.79
χ_R	0.10	0.14	0.27	0.26	0.31	0.30	0.32

Résultats clés de l'Article 4 (parties 2.1, 2.2, 2.3 et 3.1)

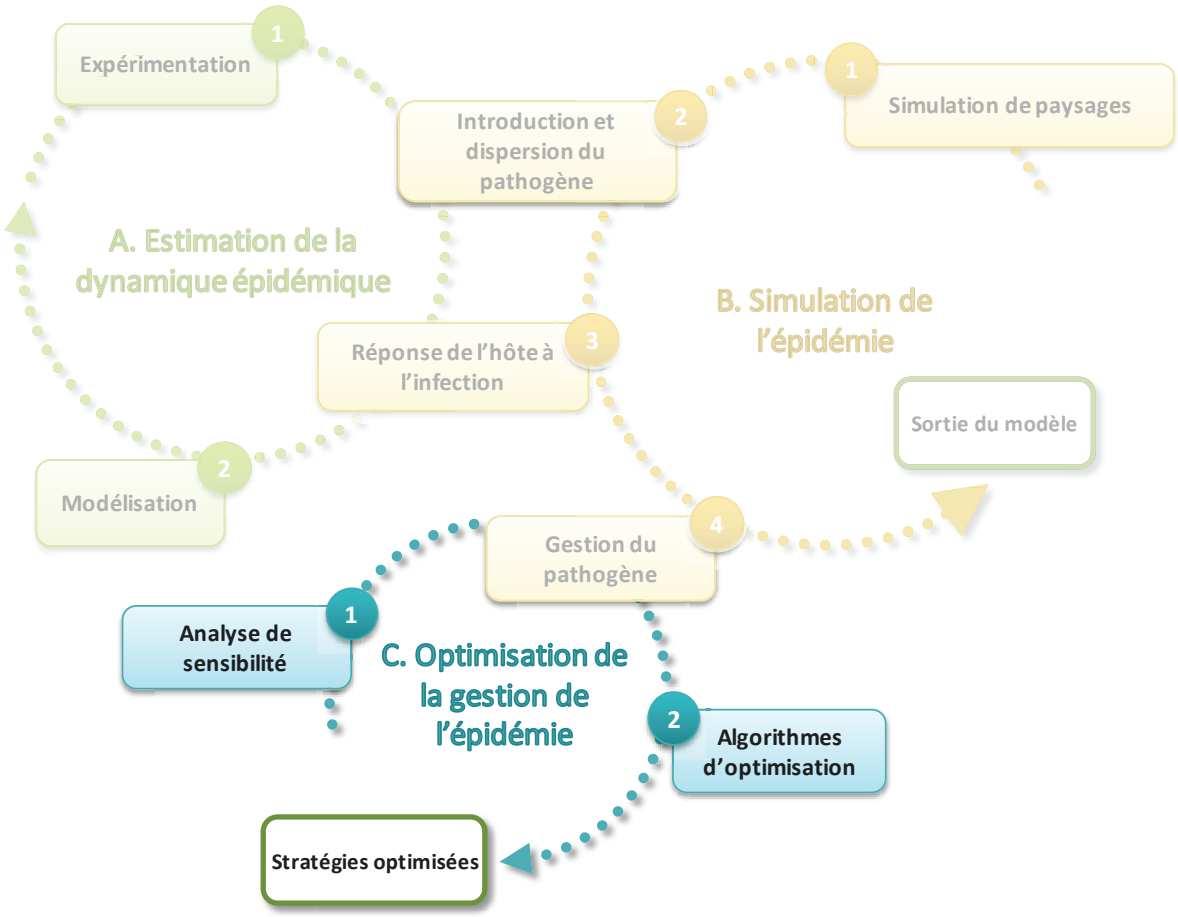
ANALYSE DE L'INFLUENCE DE L'AGREGATION DU PAYSAGE SUR LA PROPAGATION DES MALADIES POUR AMELIORER LES STRATEGIES DE GESTION

- **Une méthode pour étudier l'influence du paysage sur les stratégies de gestion**
  - Des paysages de taille et de densité différentes ont été simulés à partir de données géographiques associées à un parcellaire réel, ainsi que des paysages avec différents niveaux d'agrégation à partir d'un algorithme de tessellation en T.
  - Sur ces différents paysages, des épidémies de sharka et la gestion française de cette maladie ont été simulées.
  - Une analyse de sensibilité a été réalisée sur des paysages correspondant à 3 niveaux d'agrégation différents.

- **Le paysage influence les stratégies de gestion de la sharka**
  - Les simulations montrent que quelle que soit la taille du paysage, plus les parcelles sont agrégées, plus l'épidémie se répand vite, et plus les pertes économiques sont conséquentes (que ce soit avec ou sans gestion). L'organisation des parcelles cultivées dans un paysage a donc de l'influence sur les bénéfices de la production de pêches.
  - Deux paramètres de gestion avec une forte influence sur la VAN ont été mis en évidence grâce à l'analyse de sensibilité : ils concernent les interdictions de planter des vergers ainsi que les arrachages de vergers appartenant à une même zone géographique. Une attention particulière devra donc leur être portée dans l'optique d'optimiser les paramètres de gestion des épidémies de sharka.

Chapitre 3

OPTIMISATION DE LA STRATEGIE DE GESTION D'UNE EPIDEMIE



Dans le chapitre précédent il a été montré que lors d'une épidémie, les caractéristiques du paysage peuvent avoir de l'influence sur un critère économique (la VAN), en présence ou en absence de stratégie de gestion. L'objectif de ce chapitre 3 est de proposer une méthode pour optimiser les stratégies de gestion d'une épidémie à l'échelle du paysage, c'est-à-dire identifier les combinaisons de paramètres de gestion permettant d'obtenir la VAN la plus élevée possible. Plus spécifiquement, nous montrons comment identifier des stratégies de gestion efficaces contre des épidémies de sharka sur des paysages caractérisés par différents niveaux d'agrégation des parcelles.

Pour cela, une première approche d'optimisation basée sur des résultats d'analyses de sensibilité permet d'explorer une partie de l'espace des paramètres de gestion et d'identifier des stratégies efficaces. Une deuxième approche permettant d'explorer tout l'espace des stratégies possibles a ensuite été mise en œuvre pour résoudre notre problème d'optimisation. Pour aller plus loin, nous avons également travaillé sur l'optimisation des stratégies de gestion des épidémies lorsque des variétés résistantes de pêchers commenceront à être disponibles sur le marché, ainsi que sur la répartition optimale de ces variétés résistantes dans le paysage.

1. Optimisation des paramètres de gestion d'une épidémie grâce à l'analyse de sensibilité

Afin d'identifier des stratégies de gestion efficaces, les résultats des analyses de sensibilité effectuées précédemment ont été étudiés.

Dans un premier temps, pour chaque niveau d'agrégation du paysage, les stratégies conduisant à la meilleure VAN dans les analyses de sensibilité ont été sélectionnées. Pour évaluer leur efficacité et avoir un aperçu de l'influence du paysage, des simulations ont été réalisées avec ces stratégies sur les 3 paysages en faisant varier les paramètres épidémiologiques. Cependant, la sélection de ces stratégies est dépendante des paramètres épidémiologiques utilisés dans le plan d'échantillonnage (le plan d'échantillonnage a été réalisé avec à la fois 23 paramètres de gestion et 6 paramètres épidémiologiques, ce qui signifie que la VAN obtenue pour chaque stratégie de gestion dépend des 6 paramètres épidémiologiques utilisés).

Dans un deuxième temps, pour contourner cette dépendance aux paramètres épidémiologiques de l'analyse de sensibilité, nous avons défini des cas épidémiques (un cas épidémique correspondant à une combinaison de gammes de variation des 6 paramètres épidémiologiques). Dans chaque type de paysage et pour chaque cas épidémique, la stratégie conduisant à la meilleure VAN dans les analyses de sensibilité a été retenue et des simulations ont été réalisées en faisant

varier les paramètres épidémiologiques. Pour chaque type de paysage, les 10 meilleures stratégies ont ensuite été sélectionnées puis testées sur les autres paysages. Enfin, nous avons retenu les 3 meilleures stratégies correspondant aux 3 niveaux d'agrégation des paysages.

Les résultats de ce chapitre sont détaillés dans la dernière partie de l'article 4 présenté dans le chapitre précédant.

Résultats clés de l'Article 4 (parties 2.4, 3.1 et 3.2)

ANALYSE DE L'INFLUENCE DE L'AGREGATION DU PAYSAGE SUR LA PROPAGATION DES MALADIES POUR AMELIORER LES STRATEGIES DE GESTION

- **L'analyse de sensibilité permet d'identifier des stratégies de gestion optimisées**
  - Les résultats de 3 analyses de sensibilité ont été exploités pour identifier des stratégies de gestion efficaces pour 3 paysages différant par leur niveau d'agrégation.
  - Certaines de ces stratégies sont plus efficaces *in silico* que la stratégie de gestion française.

- **Les stratégies de gestion optimisées dépendent des caractéristiques du paysage**
  - Des stratégies de gestion optimisées spécifiques à un niveau d'agrégation du paysage ont été identifiées.
  - Une stratégie générique (efficace pour tous les paysages) a également été identifiée, ce qui est important en pratique car il peut être difficile pour les gestionnaires du risque de délimiter des zones qui diffèrent par leur niveau d'agrégation du paysage. Cette stratégie n'inclut que de très rares interdictions de planter de nouveaux vergers (lorsque le taux de contamination dans la zone environnante est trop élevé) et très peu d'arrachages de vergers entiers ; par ailleurs, elle requiert moins de surveillance des vergers que la stratégie de gestion française.

2. Optimisation des paramètres de gestion d'une épidémie grâce à un algorithme d'optimisation

L'étude présentée précédemment a montré qu'il est possible d'améliorer les stratégies de gestion d'une épidémie grâce aux résultats d'une analyse de sensibilité. Néanmoins, cette méthode est limitée par le nombre de combinaisons de paramètres de gestion explorées (310.155 dans notre cas). Des stratégies plus efficaces n'ont peut-être pas été testées ; c'est pourquoi une approche permettant d'explorer tout l'espace des stratégies possibles a été utilisée. Cette approche d'optimisation est basée sur un métamodèle de krigeage. Elle permet d'explorer l'espace des stratégies possibles de manière parcimonieuse, et de s'orienter progressivement vers les combinaisons de paramètres les plus efficaces économiquement.

Cependant, pour maximiser l'efficacité de cette approche dans le cadre de notre problème d'optimisation, un des défis a été de redéfinir par distorsion l'espace des paramètres (*warping*), en supprimant les combinaisons de paramètres qui caractérisent des gestions identiques. Cette méthode est présentée dans l'article 5, qui compare les résultats d'optimisations réalisées avec ou sans cette étape de distorsion. Dans le cadre de ma thèse, j'ai contribué à la production des résultats et à l'écriture de la partie qui traite de la description du modèle sharka et qui expose la problématique et de celle qui analyse la performance de l'étape de distorsion lors de l'optimisation de la gestion de la sharka.

Nous avons ensuite appliqué cette approche au problème de l'optimisation des stratégies de gestion de la sharka. Nous avons optimisé la stratégie de gestion de cette maladie sur la base de deux critères : la moyenne de la VAN et la moyenne des 10% des VAN les plus faibles. Nous avons réalisé des optimisations pour les 3 types de paysages (avec des niveaux d'agrégation des parcelles différents, à la fois dans le cas d'épidémies émergentes (faible prévalence avant la mise en place de la gestion) et dans le cas d'épidémies installées (forte prévalence avant la gestion). L'approche d'optimisation utilisée et les résultats des optimisations sont détaillés dans l'article 6.

ARTICLE 5

Impact of input warping on the Bayesian optimisation of the management of a plant disease using a complex epidemiological model

Victor Picheny, Coralie Picard and Gaël Thébaud

Impact of input warping on the Bayesian optimisation of the management of a plant disease using a complex epidemiological model

Victor Picheny · Coralie Picard · Gaël Thébaud

Abstract Optimizing black-box numerical models remains a challenge in many research fields. In this work, we focus on a Bayesian optimization approach, accounting for local invariances of the model with respect to its input variables. More precisely, we incorporate the prior knowleddge that the model is insensitive to variations of some of its input variables when other input variables take a particular value. To this end, we propose a new warping technique applied to the parameter space that encode the invariances. This approach is tested on a simulation model of sharka disease spread and management that exhibits several invariances. We analyze the contribution of the warping on the Bayesian optimization of sharka control options. We show that the warping step significantly improves the rate of convergence of the BO algorithm.

Keywords Bayesian optimisation, warping, spatio temporal model, sharka

1 Introduction

Mathematical models are increasingly used in many research fields to understand and optimize a process. For instance, they are useful in epidemiology to predict epidemics and to propose efficient control options [4, 5, 18, 33, 1, 14, 34, 9]. However, these epidemiological studies are moslty focused on improving one control option which generally depends on only one or two parameters in their model, although various control actions are usually applied simultaneously to manage an epidemic. All these actions could be jointly optimized but taking into account numerous management

V. Picheny
MIAT, Université de Toulouse, INRA, Castanet-Tolosan, France
Tel.: +33561285551
E-mail: victor.picheny@inra.fr

C. Picard
BGPI, Montpellier SupAgro, INRA, Univ. Montpellier, Cirad, TA A-54/K, 34398, Montpellier

G. Thébaud
to do

parameters in an optimization problem can be difficult, especially when the management efficiency depends on the interaction between these parameters.

In this study, we analyse a simulation model of sharka disease spread and management. This disease, caused by a virus transmitted by aphids, is one of the most damaging diseases of stone fruit trees belonging to the genus *Prunus* (e.g. peach, apricot and plum) [3, 25]. Our model includes epidemiological parameters which vary between simulations, and various landscapes on which the virus can spread, which means that this model is stochastic. In addition, management parameters allow to simulate orchard surveillance. Here, we aim to optimize these management parameters using a efficient optimization algorithm.

Within the wide range of potential approaches to solve such optimization problems, black-box optimization methods have proven to be popular in this context [28], in particular because they are in essence non-intrusive: they only require pointwise evaluations of the model at hand (output value for a given set of inputs), as opposed to knowing the underlying mechanisms of the model, structural information, derivatives, etc. This greatly facilitates implementation and avoids developping taylored algorithms. In this work, we focus more particularly on the so-called *Bayesian optimization* (BO) approaches [17, 30], which are well-suited to tackle stochastic and expensive models.

In some cases, the user possesses relevant information regarding his model that could facilitate the optimization task. Accounting for this information within a black-box optimization framework (or rather: *grey box*) may be a challenging task as it is, in essence, unnatural. In this work, we focus on a particular type of information, which we refer to as *local invariance*: for some values of a subset of parameters, it is known that the model is insensitive to another subset of parameters. As an illustration, take a function y that depends on two discs, parameterized by $r_1 \in [0, r_{\max}]$ (radius of the first disc) and $r_2 \in [0, r_{\max}]$ (radius of the second disc) with $r_1 > r_2$. An action A_1 is conveyed on the first disc and another action A_2 on the second. Setting $r_1 = 0$, we have $r_2 = 0$, thus for any value of A_2, y is not impacted. Taking into account such invariances would avoid wasting computational resources exploring those regions. Moreover, it would avoid the problem of having local *plateaus* of the optimization landscape, that are likely to slow down the optimization process or even prevent convergence to an optimum.

Intuitively, one may want to rework the definition of the parameters to optimize over in order to remove the invariances. However (as we show in 2), such a reformulation is not always possible. Here, we propose to keep the optimisation problem unchanged, and convey the invariance information to the BO algorithm directly, by applying a *warping* [31, 32] to the parameter space. In essence, it amounts to applying a specific deformation of the parameter space that reflects the invariance.

The remainder of this paper is structured as follow. Section 2 describes the sharka model and its invariances. Section 3 presents the basics of Bayesian optimization and Section 4 our warping strategy. Finally, section 5 analyses the efficiency of the warping on the sharka model.

2 Model description and problem set-up

The simulation model that we analyze in this work is a stochastic, spatially explicit, SEIR (susceptible-exposed-infectious-removed) model that simulates sharka spread and management actions [including surveillance, removals and replantations 22, 26, 27].

This model is orchard-based, with a discrete time step of one week. It allows to perform simulations on landscapes composed of uncultivated areas and patches on which peach trees are grown. The patches can be more or less aggregated in the landscape however, we only use in this work the 30 landscapes with a high level of patch aggregation as described by Picard et al. [19]. During the simulation, the trees in the patches are characterized by different states. When the simulation begins, they are not infected: they are in the "susceptible" state. Then, the virus is introduced the first year of the simulation in one of the patches and spreads through orchards (new introductions can also occur during the entire simulation on all patches). The virus causes changes in tree status: from "susceptible", they become "exposed" (infected but not yet infectious or symptomatic), "infectious hidden" (after the end of the latent period), "infectious detected" (when specific symptoms are detected on the tree during a survey), and "removed" (when the tree is removed from the patch). The model output is an economic criterion, the net present value (NPV), which accounts for the benefit generated by the cultivation of productive trees and the costs induced by fruit production and disease management [27].

In order to simulate wide range of epidemic and management scenarios, the model includes 6 epidemiological and 23 management parameters [27, 19]. In this work, we will use the 6 epidemiological parameters and only 10 management parameters to performed some optimizations quickly. Among the 23 management parameters, we removed parameters corresponding to plantation restrictions, removals, and surveillance of young orchards. The parameters we kept include distances of 3 zones for which the surveys are more or less frequent as well as their duration, the probability of the infected tree detection, and a contamination threshold which can request to increase the surveillance frequency in a focal zone. Details of management parameters used in this study are presented in Fig.1 and Table 1 (this table also includes the variation ranges of the parameters in the model).

Here, we aim to optimize the management strategy of the disease (i.e. to find the combination of management parameters allowing to obtain the best NPV), taking into account the epidemic stochasticity. However, we note that some combinations of management parameters can represent the same management, which may cause problems in the optimization process. Indeed, we observe that some management parameters are not useful when other parameters have a value of 0, which means that they can take any values without modifying the simulation. For example, when a zone radius is 0, the associated surveillance frequency have no impact on the NPV (regardless its value). The methodological developments that are proposed in this work address this issue by removing the parameter combinations which lead to the same management. The parameter invariances removed from the model are listed in Table 2.

Table 1 Management parameters implemented in the previously developed model with minimum and maximum values corresponding to the variation range of each parameter.

		Min	Max
ρ	Probability of detection of a symptomatic tree	0	0.66
γ_O	Duration of observation zones (year)	0	10
ζ_s	Radius of security zones (m)	0	5800
ζ_f	Radius of focal zones (m)	0	5800
ζ_{eO}	Radius of observation epicenter (m)	0	5800
$1/\eta_0$	Maximal period between 2 observations (year)	1	15
η_s	Observation frequency in security zones (year $^{-1}$)	0	8
η_f	Observation frequency in focal zones (year $^{-1}$)	0	8
η_{f*}	Modified observation frequency in focal zones (year $^{-1}$)	0	8
χ_o	Contamination threshold in the observation epicenter, above which the observation	0	1

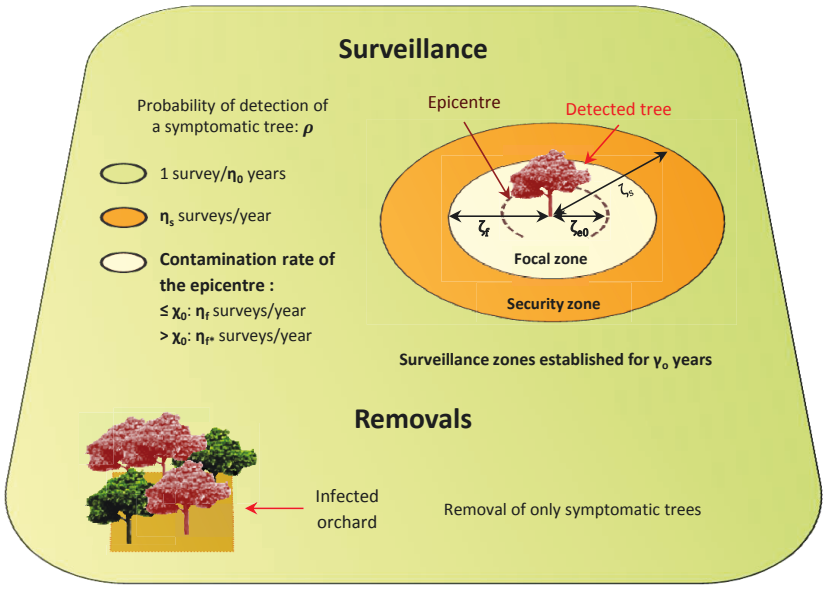


Fig. 1 Management actions implemented in the model.

3 Basics of Bayesian optimization

3.1 Gaussian process modeling

Bayesian optimization can be seen as a modernization of the statistical response surface methodology for sequential design [2], where the basic idea is to replace an

Table 2 Invariances of management parameters. For instance, when $\gamma_O = 0$ or when $\rho = 0$, χ_o does not influence the model output.

Management parameters	OR	OR	OR
χ_o	$\gamma_O = 0$	$\rho = 0$	
ζ_{eO}	$\gamma_O = 0$	$\zeta_s = 0$	$\rho = 0$
ζ_f	$\gamma_O = 0$	$\zeta_s = 0$	
η_{f*}	$\gamma_O = 0$	$\rho = 0$	
ζ_s	$\gamma_O = 0$	$\eta_s = 0$	
η_s	$\gamma_O = 0$		
η_f	$\gamma_O = 0$		

expensive-to-evaluate function by a cheap-to-evaluate surrogate one. In BO, Gaussian process (GP) regression, or kriging, is used to provide flexible response surface fits. GPs are attractive in particular for their tractability, since they are simply characterized by their mean $m(.)$ and covariance (or kernel) $k(.,.)$ functions, see e.g., Rasmussen and Williams [24]. In the following, we consider zero-mean processes ($m = 0$) for the sake of conciseness.

Conditionally on n noisy observations $\mathbf{f} = (f_1, \ldots, f_n)$, with independent, centered, Gaussian noise, that is, $f_i = y(\mathbf{x}_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \tau_i^2)$, the predictive distribution of y is another GP, with mean and covariance functions given by:

$$\mu(\mathbf{x}) = \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{f}, \tag{1}$$

$$\sigma^2(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}'), \tag{2}$$

where T denotes the tranposition operator, $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_n))^\top$ and $\mathbf{K} := (k(\mathbf{x}_i, \mathbf{x}_j) + \tau_i^2 \delta_{i=j})_{1 \leq i,j \leq n}$, δ standing for the Kronecker function.

Commonly, $k(.,.)$ belongs to a parametric family of covariance functions such as the Gaussian and Matérn kernels, based on hypotheses about the smoothness of y. Corresponding hyperparameters are often obtained as maximum likelihood estimates, see e.g., Rasmussen and Williams [24] or Roustant et al [29] for the corresponding details.

Note that in general, *stationary* covariances are used, i.e. k only depends on the distance $\|\mathbf{x} - \mathbf{x}'\|$ and not on the locations $\mathbf{x}$ and $\mathbf{x}'$. This implies that the unconditional joint probability distribution of the process does not change when shifted in the $\mathbb{X}$ space, which is in contradiction with the notion of local invariance.

3.2 Optimization

BO typically tackles optimization problems of the form:

$$\begin{aligned} &\min \quad y(\mathbf{x}) \\ &s.t.\ \mathbf{x} \in \mathbb{X}, \end{aligned}$$

with $\mathbb{X} \in \mathbb{R}^d$ is usually a bounded hyperrectangle and $y : \mathbb{R}^d \to \mathbb{R}$ is a scalar-valued objective function.

Optimization amounts here to choosing a sequence of points $\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+N}$ at which the function y is evaluated. Sequential design decisions, so-called *acquisitions*, are based on the GP model and judiciously balance exploration and exploitation in search for global optima. The GP model is updated after each new value is calculated.

In the noiseless setting ($\tau = 0$), the canonical acquisition function is *expected improvement* (EI) [13]. Define $f_{\min} = \min_{i=1,\ldots,n} y_i$, the smallest y-value seen so far, and let $I(\mathbf{x}) = \max\{0, f_{\min} - Y(x)\}$ be the *improvement* at x. $I(x)$ is largest when $Y(\mathbf{x})$ has substantial distribution below $f_{\min}$. The expectation of $I(x)$ over $Y(x)$ has a convenient closed form, revealing balance between exploitation ($\mu(x)$ under $f_{\min}$) and exploration (large $\sigma^n(x)$):

$$\mathbb{E}\{I(x)\} = (f_{\min} - \mu(x))\Phi\left(\frac{f_{\min} - \mu(x)}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{f_{\min} - \mu(x)}{\sigma(x)}\right), \qquad (3)$$

where Φ (ϕ) is the standard normal cdf (and pdf respectively).

When y is only available through noisy evaluations, the EI acquisition cannot be used directly. Several authors have tackled this issue; we refer to [21] for a review on the topic. We chose here to focus on the *reinterpolation method* proposed in [11], which is based on the use of an instrumental noiseless kriging model, built from the original one. First, the (noisy) kriging predictions at the DOE points $\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_n)$ are computed. Then, a reinterpolating model is built, by using the same covariance kernel and parameters and the same experimental design, but the observation vector is replaced by $\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_n)$ and the noise variance is set to zero. Since this latter model is noise-free, the classical EI can be used as the infill criterion. Once the new design is chosen and the evaluation is performed, both kriging models are updated.

4 Bayesian optimization with invariances

4.1 Definition of local invariances

We first introduce the following notation (this is purely notation, no actual permutation is performed):

$$y(\mathbf{x}) = y(x_i, \mathbf{x}_J, \mathbf{x}_{-iJ}) \qquad (4)$$

$$\mathbb{X} = \{\mathbb{X}_i, \mathbb{X}_J, \mathbb{X}_{-iJ}\} \qquad (5)$$

Definition 1 (Simple) We call *simple invariance* the following case: y is invariant with respect to $\mathbf{x}_J$ (J a subset of $\{1, \ldots, d\} \setminus i$) if $x_i = c_i$ ($i \in \{1, \ldots, d\}$):

$$y(c_i, \mathbf{x}_J, \mathbf{x}_{-iJ}) = y(c_i, \mathbf{x}'_J, \mathbf{x}_{-iJ}), \qquad \forall \mathbf{x}_J, \mathbf{x}'_J \in \mathbb{X}_J, \mathbf{x}_{-iJ} \in \mathbb{X}_{-iJ}.$$

This corresponds for instance to the last line of Table 2: the observation frequency η_f does not have an effect on the model if the duration of observation γ_O is set to zero.

Definition 2 (Or) We call *"or" invariance* the following case: y is invariant with respect to $\mathbf{x}_J$ (J a subset of $\{1,\ldots,d\} \setminus I$) if there exists at least one $i \in I$ such that $x_i = c_i$ (I a subset of $\{1,\ldots,d\} \setminus J$):

$$y(c_i, \mathbf{x}_{I\setminus i}, \mathbf{x}_J, \mathbf{x}_{-IJ}) = y(c_i, \mathbf{x}_{I\setminus i}, \mathbf{x}'_J, \mathbf{x}_{-IJ}), \qquad \forall \mathbf{x}_J, \mathbf{x}'_J \in \mathbb{X}_J, \mathbf{x}_{I\setminus i}, \in \mathbb{X}_{I\setminus i}.$$

This corresponds for instance to the first line of Table 2: the contamination threshold in the observation zone χ_o does not have an effect on the model if the duration of observation γ_O is set to zero or if the probability of detection ρ is set to zero.

Definition 3 (Linear) We call *linear invariance* the following case: y is invariant with respect to $\mathbf{x}_J$ (J a subset of $\{1,\ldots,d\} \setminus I$) if $\mathbf{A}\mathbf{x}_I = \mathbf{b}$, with I a subset of $\{1,\ldots,d\} \setminus J$), $\mathbf{A}$ a matrix of size $p \times \mathrm{Card}(I)$ and $\mathbf{b}$ a vector of size p:

$$y(\mathbf{x}_I, \mathbf{x}_J) = y(\mathbf{x}_I, \mathbf{x}'_J), \qquad \forall \mathbf{x}_J, \mathbf{x}'_J \in \mathbb{X}_J, \text{ if } \mathbf{A}\mathbf{x}_I = \mathbf{b}.$$

There are two particular cases worth noting:

- setting $p = \mathrm{Card}(I)$, $\mathbf{A} = \mathbb{I}_p$ and $\mathbf{b} = \mathbf{c}_I$ results in an "AND" condition: y is invariant with respect to $\mathbf{x}_j$ if, $\forall i \in I, x_i = c_i$;
- setting $p = 1$, $\mathbf{A} = [1, -1]$ results in an invariance under the condition $x_{i1} = x_{i2}$.

This invariance case is not illustrated in this work with the sharka problem optimization presented here (with 10 management parameters). However, we may have this situation if we use all the parameters implemented in the model. For instance, a parameter γ_y (not used here) is implemented in the model. It corresponds to the duration of an observation zone for young orchards. In this case, the radius of observation epicenter ζ_{eO} does not have an effect on the model if the duration of observation zones γ_O is set to 0 AND if the duration of an observation zone for young orchards γ_y is also set to 0.

4.2 Principle of input warping

There are several ways of incorporating structural information into Gaussian processes. One is to work on the kernel function k [8, 6]. Another, which is the one we use here, is to transform the original input space $\mathbb{X}$ into a *warped* one $\tilde{\mathbb{X}}$ and index the GP on $\tilde{\mathbb{X}}$, so that the new topology directly reflects the structural information [32, 15].

Consider for simplicity a single invariance over x_J when $x_i = c_i$. A simple way to handle this problem is to distort locally the space so that the subspace $\{(x_i, \mathbf{x}_J) | x_i = c_i\}$ collapses to a single point, for instance with $\mathbf{x}_J$ at its average value: $(c_i, \overline{\mathbf{x}_J})$.

Hence, we are seeking warping functions of the form:

$$\begin{aligned} &: \mathbb{X} \to \widetilde{\mathbb{X}} \\ &\mathbf{x} \mapsto \widetilde{\mathbf{x}} = \psi(\mathbf{x}) \end{aligned}$$

such that:

1. $\psi(x_i, \mathbf{x}_J, \mathbf{x}_{-iJ}) = (c_i, \overline{\mathbf{x}_J}, \mathbf{x}_{-iJ})$ if and only if $x_i = c_i$;
2. restricted to $\mathbb{X}\setminus(c_i, ., .)$ and $\widetilde{\mathbb{X}}\setminus(c_i, \overline{\mathbf{x}_J}, .)$ is a diffeomorphism.

In addition, we will search for deformations that decrease monotonically when $|x_i - c_i|$ increases, that is:

$$\left(\left(x_i, \mathbf{x}_J, \mathbf{x}_{-iJ}\right), \psi\left[\left(x_i, \mathbf{x}_J, \mathbf{x}_{-iJ}\right)\right]\right) \leq d\left(\left(x_i', \mathbf{x}_J, \mathbf{x}_{-iJ}\right), \psi\left[\left(x_i', \mathbf{x}_J, \mathbf{x}_{-iJ}\right)\right]\right) \quad \text{if } |x_i - c_i| \leq |x_i' - c_i|,$$

for some distance $d(.,.)$.

Since the $\mathbf{x}_J$ dimension collapses to $\overline{\mathbf{x}_J}$ at $x_i = c_i$, we write:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j})\, \alpha(x_i, c_i), \tag{6}$$

with $\alpha(x_i, c_i)$ an attenuation function such that:

1. $\alpha(c_i, c_i) = 0$;
2. α increases monotonically with $|x_i - c_i|$;
3. $0 < \alpha \leq 1, \forall x_i \neq c_i$.

Condition 1 ensures that $\widetilde{\mathbf{x}_j} = \overline{\mathbf{x}_j}$ when $x_i = c_i$ (all the dimensions in J collapse).

4.3 Warping for a simple invariance

In the simple invariance case, we propose linear and correlation-based attenuation functions:

$$\alpha_{\text{lin}}(x_i, c_i) = \frac{|x_i - c_i|}{\delta_i}, \tag{7}$$

$$\alpha_{\text{cor}}(x_i, c_i) = 1 - r(x_i, c_i), \tag{8}$$

where r is a $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ correlation function. Typically, δ_i may be set to the range of variation of x_i, so that the condition $\alpha \leq 1$ is ensured. Choosing r as the generalized exponential correlation, we have:

$$\alpha_{\text{exp}}(x_i, c_i) = 1 - \exp\left[-\left(\frac{|x_i - c_i|}{\theta_i}\right)^d\right], \tag{9}$$

with θ_i and d positive parameters to be tuned.

Figure 2 shows a 2D rectangular space distorted by three warpings, when the invariance is on a boundary of x_1.

4.4 Warping for linear invariances

For simplicity, we consider first the particular linear case where $\mathbf{A} = \mathbb{I}_p$ and $\mathbf{b} = \mathbf{c}_I$, that is where invariances occur when a set of variables takes simultaneously a set of critical values: $\mathbf{A}\mathbf{x}_I = \mathbf{b}$, or equivalently $\mathbf{x}_I = \mathbf{c}_I$. In that case, a possible warping is:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j})\, \alpha_I(\mathbf{x}_I, \mathbf{c}_I). \tag{10}$$

with α_I now a multivariate attenuation function ($\mathbb{R}^{\text{Card}(I)} \times \mathbb{R}^{\text{Card}(I)} \to \mathbb{R}$), so that, similarly to the simple case:

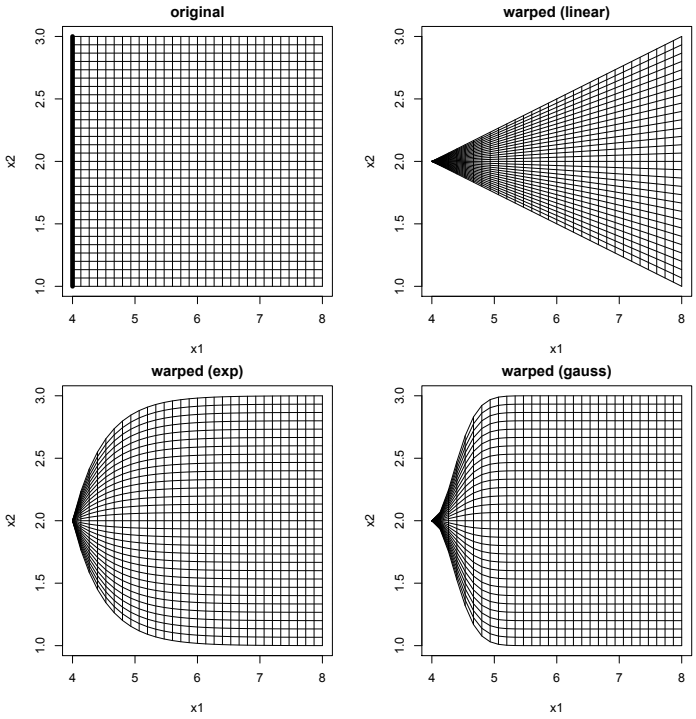


Fig. 2 Three deformations of a 2D space. The local invariance is at $x_1 = 0$, highlighted with larger lines.

1. $\alpha_I(\mathbf{c}_I, \mathbf{c}_I) = 0$;
2. α_I increases monotonically with $d(\mathbf{x}_I, \mathbf{c}_I)$ (for some distance $d(.,.)$;
3. $0 < \alpha_I \leq 1, \forall \mathbf{x}_I \neq \mathbf{c}_I$.

As in the simple case, linear and correlation-based warpings can be defined as:

$$\alpha_{\text{lin}}(\mathbf{x}_I, \mathbf{c}_I) = \frac{1}{\text{Card}(I)} \sum_{i \in I} \frac{|x_i - c_i|}{\delta_i}, \tag{11}$$

$$\alpha_{\text{cor}}(\mathbf{x}_I, \mathbf{c}_I) = 1 - r_I(\mathbf{x}_I, \mathbf{c}_I), \tag{12}$$

with r_I a $\mathbb{R}^{\text{Card}(I)} \times \mathbb{R}^{\text{Card}(I)} \to \mathbb{R}$ correlation function as in 9.

Generalizing to the affine case $\mathbf{A}\mathbf{x}_I = \mathbf{b}$, the warping function is the same as in Equation 10, with now:

$$\alpha(\mathbf{x}_I, \mathbf{c}_I) = 1 - r_A(\mathbf{A}\mathbf{x}_I, \mathbf{b}). \tag{13}$$

4.5 Combining warpings

Independent conditions Now, we consider that we have a series of invariance conditions, defined with respect to sets $I_1, \ldots, I_n$ and corresponding $J_1, \ldots, J_n$. If $J_k \cap J_l = \emptyset$, $1 \leq j \neq k \leq n$ and $I_i \cap J_k = \emptyset$, $1 \leq j, k \leq n$, the set of warped

variables are distinct from the set on which the conditions are written, the invariance conditions are written only once for each variable. In that case, the warpings can be applied independently.

Combinations of simple conditions: "OR" invariance Now, we consider the case when y is invariant w.r.t. a set $\mathbf{x}_J$ for different conditions on sets $I_1, \ldots, I_n$ (that, for $\mathbf{x}_{I_1} = \mathbf{c}_{I_1}$ OR $\mathbf{x}_{I_2} = \mathbf{c}_{I_2}$ OR ...). If $J \cap I_i = \emptyset$, $1 \leq i \leq n$, the warping function we propose is:

$$\forall j \in J, \quad \widetilde{x_j} = \overline{x_j} + (x_j - \overline{x_j}) \prod_{I \in \{I_1, \ldots, I_n\}} \alpha_I(\mathbf{x}_I, \mathbf{c}_I). \tag{14}$$

We see directly that the product of α's ensure that $\widetilde{x_j} = \overline{x_j}$ if any $x_i = c_i$, and the distortion reduces only when *all* the x_i's are far from the c_i's.

"Circular" conditions Difficulty only arises when some variables appear in both I_l's and J_m's sets. Take for instance a "reciprocal" condition, e.g., y is invariant w.r.t. $\mathbf{x}_J$ when $\mathbf{x}_I = \mathbf{c}_I$, and invariant w.r.t. $\mathbf{x}_I$ when $\mathbf{x}_J = \mathbf{c}_J$. In that case, applying independently warping functions would lead to:

$$\begin{aligned} \psi(\mathbf{c}_I, \mathbf{x}_J, \mathbf{x}_{-IJ}) &= (\mathbf{c}_I, \overline{\mathbf{x}_J}, \mathbf{x}_{-IJ}), \\ \psi(\mathbf{x}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}) &= (\overline{\mathbf{x}_I}, \mathbf{c}_J, \mathbf{x}_{-IJ}), \\ \text{but: } \psi(\mathbf{c}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}) &= (\mathbf{c}_I, \mathbf{c}_J, \mathbf{x}_{-IJ}), \end{aligned}$$

which induces a discontinuity.

In that case, a simple solution is to fix the non influent variable to its critical value instead of its average, hence applying:

$$\forall k \in K = (\cup_{1 \leq l \leq n} I_l) \cap (\cup_{1 \leq m \leq n} J_m), \widetilde{x_k} = c_k + (x_k - c_k) \prod_{i \in I_k} \alpha(x_i, c_i) \tag{15}$$

Remark This formula does not apply in the affine case (Equation 13).

We show the deformations on a 2D space on Figure 3, where the two critical values are on the boundaries of x_1 and x_2. Here, the warping of Equation 15 is applied on each variable ($K = \{1, 2\}$). Again, except for the linear warping, the local topology is preserved far from the critical edges.

4.6 Illustration

Finally, Figure 4 shows four deformations of the unit cubic space, for each of the following invariances:

- AND: y is invariant w.r.t. x_3 when x_1 AND x_2 are equal to zero (equation 10 with $I = 1$ and $J = \{1, 2\}$);
- OR: y is invariant w.r.t. x_3 when x_1 OR x_2 are equal to zero (equation 14 with $I = 1$ and $J = \{1, 2\}$);

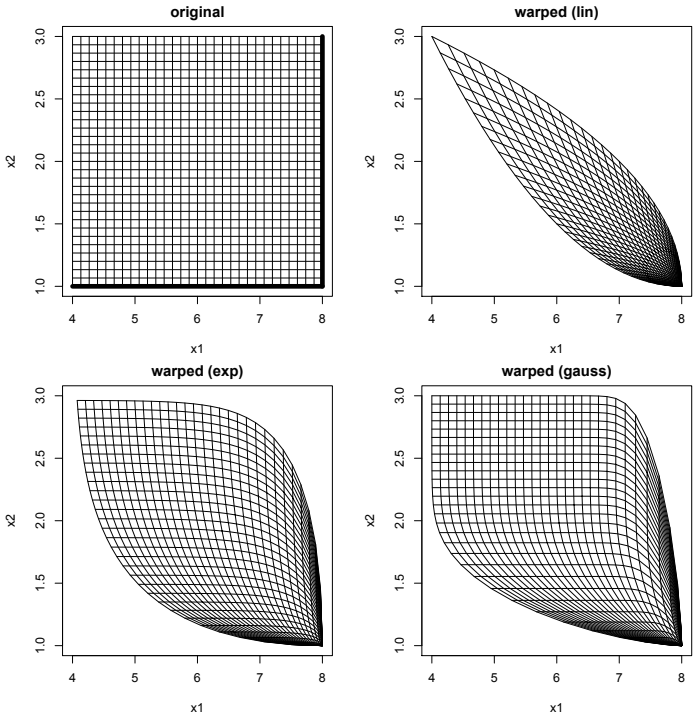


Fig. 3 Three deformations of a 2D space, with invariance at $x_1 = 8$ OR $x_2 = 1$, highlighted with larger lines.

- LINEAR: y is invariant w.r.t. x_3 when $x_1 = x_2 = 0$ (equation 13 with $I = 1$, $J = \{1, 2\}$, $\mathbf{A} = [1, -1]$ and $\mathbf{b} = 0$);
- CIRCULAR: y is invariant w.r.t. a- x_2 if $x_1 = 0$, b- x_3 if $x_2 = 0$, c- w.r.t. x_1 if $x_3 = 0$ (equation 15 with $K = \{1, 2, 3\}$, $C = [0, 0, 0]$ and $I_1 = 3$, $I_2 = 1$, and $I_3 = 2$).

On all cases, a Gaussian warping (exponential with $d = 2$) is applied, with range parameter $\theta = 0.3$.

4.7 Warping parameters tuning

The linear warping has the advantage of being parameter-free, which comes at a price of a profund modification of the problem topology. The correlation-based warpings have the capability of creating more localized distortions, but depend on range parameters (the θ_i's in Equation 9). Those may be estimated by likelihood maximization along with the GP covariance parameters [32, 15].

However, we found in our numerical experiments that choosing the same correlation function for the GP and the warping, and fixing the warping ranges to be 1/10th of the GP ones provided very satisfactory results, while avoiding the extra computational burden.

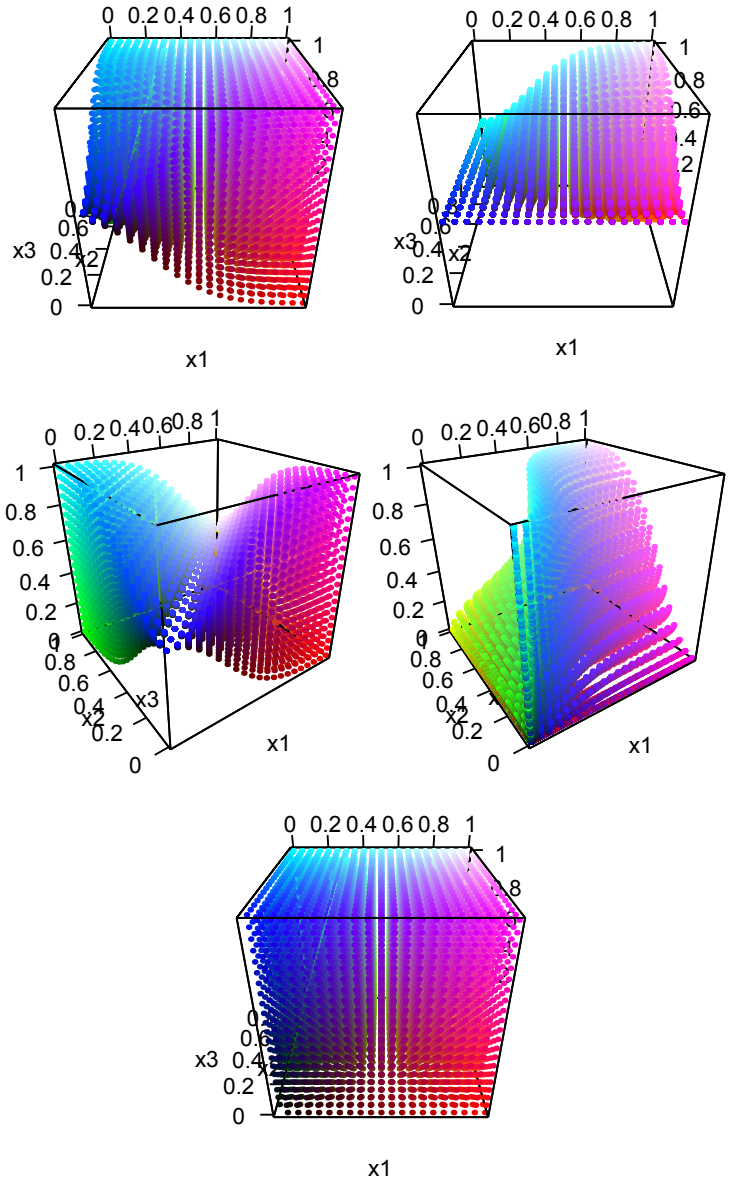


Fig. 4 Four deformations of the unit cube under different invariances: AND (top left), OR (top right), LINEAR (middle left), CIRCULAR (middle right). The bottom figure shows the orginal space.

Note that in the case of linear invariances, choosing the range of the correlation r_A is non-trivial, as it is not directly linked to design variables. A possible solution is $\theta_A = \mathbf{A}^T \boldsymbol{\theta}_I$.

4.8 Bayesian optimization on warped spaces

A decisive advantage of warping over alternative approaches is that it does not require any change of the BO apparatus. The GP modeling step is performed in the warped space $\tilde{\mathbb{X}}$ instead of the original one $\mathbb{X}$, that is, a standard GP model (i.e. stationary) is fitted to the transformed design of experiments $\{\tilde{\mathbf{x}}_1, y_1\}, \ldots, \{\mathbf{x}_n, y_n\}$.

The acquisition maximization can be done directly in the original space:

$$\mathbf{x}_{n+1} \in \arg\max_{\mathbf{x}\in\mathbb{X}} EI\left[\psi(\mathbf{x})\right]. \tag{16}$$

Note that EI would exhibit the same invariances as the objective function.

Figure 5 shows unconditional realizations of GPs originally defined in the warped space but shown in the original space (using the inverse of the transformation ψ), for each of the warpings of Figure 2. We see that the invariance at $x_1 = 1$ is ensured. The linear warping induces a strong anisotropy, while with the two other warpings, the process seems stationary far from the critical value.

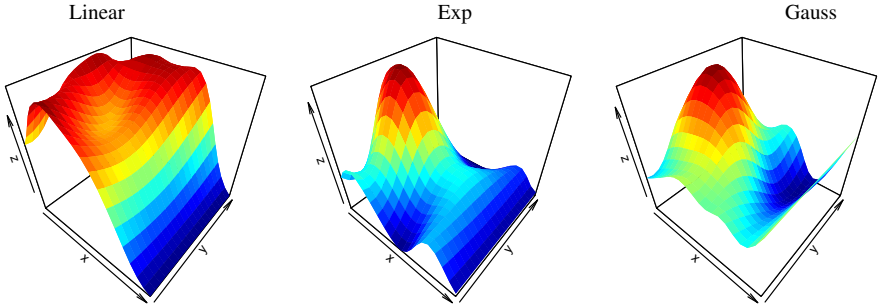


Fig. 5 Three GP realizations using warping functions as shown previously.

5 A warping-based Bayesian optimization of the Sharka model

5.1 Numerical setup

5.1.1 Experiments description

To assess the benefits of including the warping step in the optimization process (i.e. reducing the parameter space removing the combinations which lead to the same management), we conducted 50 independent optimizations of sharka management parameters with and without the warping step. Warping is applied to seven variables, following Table 2, to account for two simple invariances η_s, η_f, two combined ones χ_o, η_{f*}, and three implying "circular" conditions: ζ_{eO}, ζ_f and ζ_s. On all cases, we used a Matérn 5/2 correlation-based warping.

The economic criterion to optimize was the mean of the NPV ($\overline{NPV}$). For this to happen, we randomly selected 50 times 200 management strategies using a maximin

Latin hypercube sampling design [7]. Then, for each sampling design of 200 strategies, we performed 2 optimizations in parallel: with and without the warping step. For each optimization, we performed sequentially 200 iterations allowing to choose 200 new strategies, resulting in a total of 400 evaluated strategies. For each evaluated strategy, the objective function is computed by averaging over 1,000 simulations (carried out with different random seeds) to take into account the variability due to the epidemic and landscape characteristics.

5.1.2 Bayesian optimization setup

For all experiments, we used the same GP modeling setup, that is, an unknown constant trend (ordinary kriging, [16]) and Matérn 5/2 covariance function [24, Chapter 4]. The acquisition function maximized at each step is the expected improvement on the *reinterpolating* model. The maximization is performed by a large-scale random search followed by a local optimization starting for the optimum found by the random search (i.e. the evaluated points in the optimization process are chosen around the best current $\overline{NPV}$) All experiments were conducted in `R` [23], using code adapted from the `DiceOptim` package [20].

5.2 Results

We firstly compared the optimization results by substracting the $\overline{NPV}$ achieved using the optimization with the warping step and the optimization without the warping step (obtained from the same sampling design). In 24 out of the 50 optimization cases, we obtained better $\overline{NPV}$ with the warping step than without. This point is illustrated by the probability density function which is centered on 0 (Fig.6). This result means that with 200 iterations in the optimization, the final optimization result is not impacted by the use of the warping.

However, we showed that the warping can impact the optimization speed (Fig.7). Indeed, at the 3^{rd} iteration, the gap between the yellow (with warping) and the blue (without warping) lines is already 3957euro/ha. In addition, to reach $\overline{NPV}$=16,400euro/ha, we needed in average only 96 iterations in the optimization process with warping against 144 iterations without warping.

To go further, we performed a nonlinear regression of $\overline{NPV}$ obtained for all the selected strategies during the optimization process with and without the warping step, and we compared the growth parameter c of the following regression: $\overline{NPV} = A + be^{-cx_i}$. This parameter was higher with (0.26) than without (0.18) warping.

In addition, we can visually observe that the warping step allow to improve the optimization speed on the Fig.8 and Supplementary Fig.1. These figures were represented with a specific algorithm based on empirical distribution functions [10]. Briefly, we uniformly defined 100 α values within a specified range. Then, for each iteration performed in the optimization process (i.e. for each of the 200 evaluated strategies), we add: the number of optimizations (among 50) which exceed α_1, the number of optimizations which exceed α_2, ..., the number of optimizations which exceed α_{100}. We used $\alpha \in$ [0;18,012.12] Supplementary (Fig.1) and $\alpha \in$ [10,000;18,012.12]

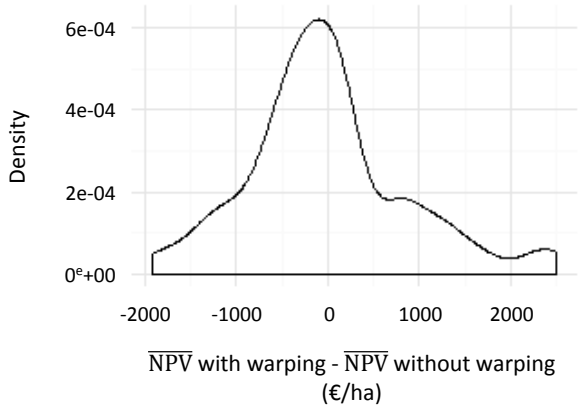


Fig. 6 Comparison of $\overline{NPV}$ obtained at the end of the optimization with and without warping.

(Fig.8). The value 18,012.12 corresponds to the maximal value of $\overline{NPV}$ identified in

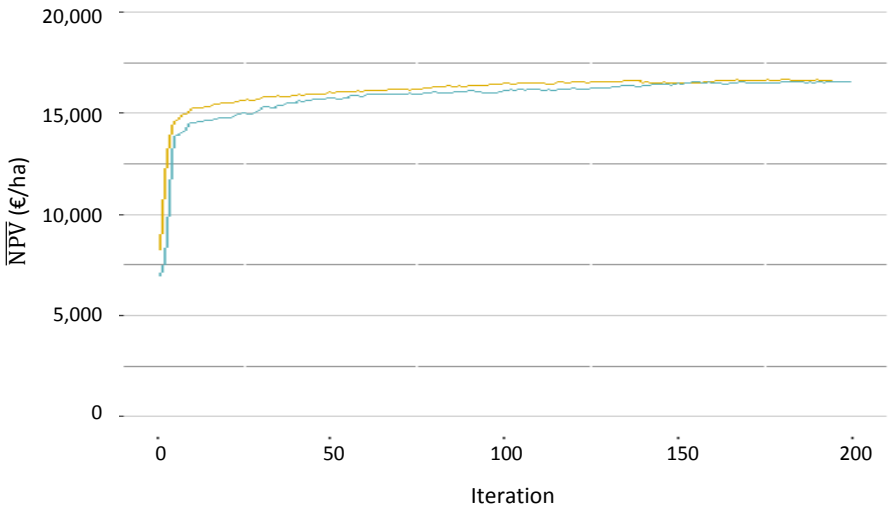


Fig. 7 Comparison of $\overline{NPV}$ obtained during optimizations with and without warping. Yellow and blue lines represent the mean of the $\overline{NPV}$ selected at each iteration for the 50 optimizations respectively perfomed with and without the warping step.

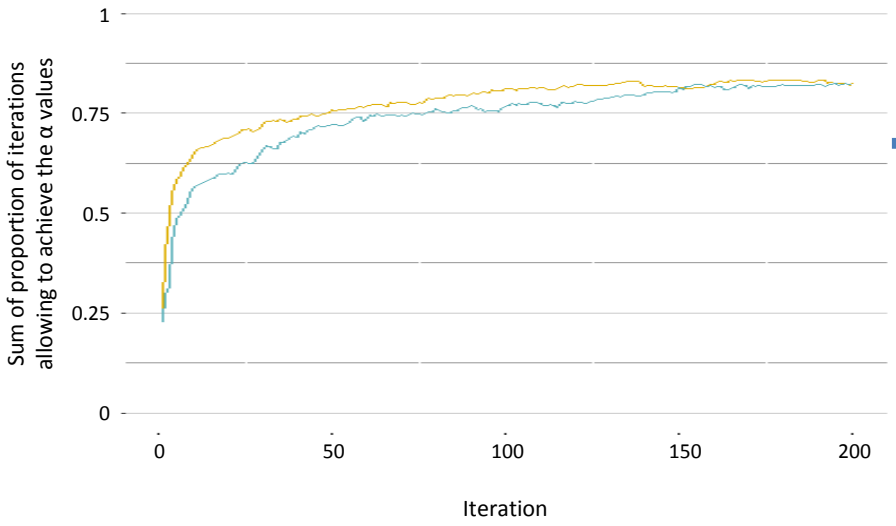


Fig. 8 Results of the algorithm using empirical distribution functions [10] with (yellow) and without (blue) warping ($\alpha \in$ [10,000;18,012.12]).

6 Conclusion

In this study, we showed how a Bayesian optimization process can be improved by accounting for some prior structural information: the insensitivity of the model with respect to a subset of its input variables when another subset of inputs takes a particular value. Such *local invariances* were exhibited by our spatiotemporal model simulating sharka management, characterized by 10 parameters related to the surveillance of the orchards. In this example, the invariances arise because parameters (radius of different zones, surveillance frequency in each zone, detection probability of infected trees, and duration of observation zones) are strongly related. Indeed, we easily note, for instance, that when the detection probability takes a value of 0, numerous other parameters do not influence the model results.

To tackle this problem, we proposed to use a warping of the input space, that here amounted to remove locally dimensions of the input space. The warping we used is based on correlation functions, making it very simple to implement while allowing sufficient flexibility. A particular advantage of input warping over other approaches is that it can be straightforwardly embedded in a BO algorithm.

We applied this Bayesian optimization process to the spatio-temporal sharka model. We performed various optimizations of its management parameters firstly with the use of warping (which allows accounting for the invariances) and then without. We showed that both approaches led to the same maximal $\overline{NPV}$, but the the optimization process with warping was substantially faster, showing that the warping efficiently reduced the search space without altering the exploration / exploitation trade-off.

As future steps for this research, we could first embed learning the warping parameters together with the parameters of the GP covariance in a single likelihood maximization step. Another room for improvement is to adapt the EI maximization

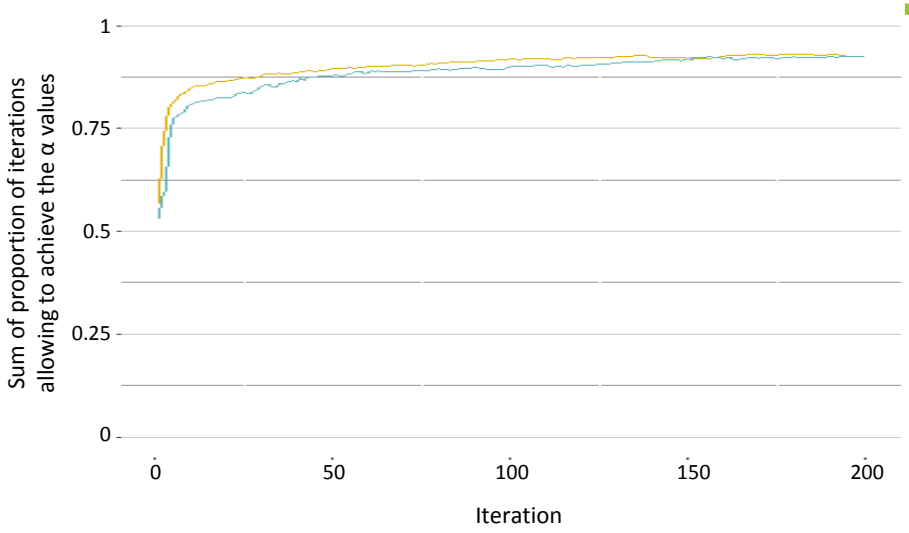
step to the new topology induced by the warping (here, on all experiments the EI was maximized over the original space). Finally, the optimization strategy pursued here used a large fixed number of replicates (1,000) for each evaluated design. Combining warping with an efficient adaptative scheme to handle replicates [12] would drastically reduce the cost of the optimization.

References

1. Bajardi P, Barrat A, Savini L, Colizza V (2012) Optimizing surveillance for livestock disease spreading through animal movements. Journal of the Royal Society Interface 9(76):2814–2825
2. Box GE, Draper NR (1987) Empirical model-building and response surfaces. John Wiley & Sons
3. Cambra M, Capote N, Myrta A, Llácer G (2006) Plum pox virus and the estimated costs associated with sharka disease. EPPO Bulletin 36(2):202–204
4. Cunniffe NJ, Koskella B, Metcalf CJE, Parnell S, Gottwald TR, Gilligan CA (2015) Thirteen challenges in modelling plant diseases. Epidemics 10:6–10
5. Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA (2016) Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in california. Proceedings of the National Academy of Sciences of the United States of America 113(20):5640–5645
6. Duvenaud D (2014) Automatic model construction with gaussian processes. PhD thesis, University of Cambridge
7. Fang KT, Li R, Sudjianto A (2005) Design and modeling for computer experiments. Chapman and Hall/CRC
8. Ginsbourger D, Durrande N, Roustant O (2013) Kernels and designs for modelling invariant functions: From group invariance to additivity. In: mODa 10–Advances in Model-Oriented Design and Analysis, Springer, pp 107–115
9. Grechi I, Ould-Sidi MM, Hilgert N, Senoussi R, Sauphanor B, Lescourret F (2012) Designing integrated management scenarios using simulation-based and multi-objective optimization: Application to the peach tree–myzus persicae aphid system. Ecological Modelling 246:47–59
10. Hansen N, Auger A, Ros R, Finck S, P P (2010) Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009. Proceedings of the 12th annual conference companion on Genetic and evolutionary computation pp 1689–1696
11. J Forrester AI, Keane AJ, Bressloff NW (2006) Design and analysis of" noisy" computer experiments. AIAA journal 44(10):2331–2339
12. Jalali H, Van Nieuwenhuyse I, Picheny V (2017) Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise. European Journal of Operational Research 261(1):279–301
13. Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13(4):455–492
14. Kompas T, Ha PV, Nguyen HTM, East I, Roche S, Garner G (2017) Optimal surveillance against foot-and-mouth disease: the case of bulk milk testing in aus-

tralia. Australian Journal of Agricultural and Resource Economics 61(4):515–538
15. Marmin S, Ginsbourger D, Baccou J, Liandrat J (2018) Warped gaussian processes and derivative-based sequential designs for functions with heterogeneous variations. SIAM/ASA Journal on Uncertainty Quantification 6(3):991–1018
16. Matheron G (1963) Principles of geostatistics. Economic Geology 58(8):1246–1266
17. Mockus J (2012) Bayesian approach to global optimization: theory and applications, vol 37. Springer Science & Business Media
18. Mushayabasa S, Tapedzesa G (2015) Modeling the effects of multiple intervention strategies on controlling foot-and-mouth disease. BioMed research international
19. Picard C, Soubeyrand S, Jacquot E, Thébaud G (2018) Analyzing the influence of landscape aggregation on disease spread to improve management strategies. (under review)
20. Picheny V, Ginsbourger D (2014) Noisy kriging-based optimization methods: a unified implementation within the DiceOptim package. Computational Statistics & Data Analysis 71:1035–1053
21. Picheny V, Wagner T, Ginsbourger D (2013) A benchmark of kriging-based infill criteria for noisy optimization. Structural and Multidisciplinary Optimization 48(3):607–626
22. Pleydell DRJ, Soubeyrand S, Dallot S, Labonne G, Chadœuf J, Jacquot E, Thébaud G (2018) Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLoS Computational Biology 14(4):e1006085
23. R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`
24. Rasmussen CE, Williams C (2006) Gaussian Processes for Machine Learning. MIT Press, URL `http://www.gaussianprocess.org/gpml/`
25. Rimbaud L, Dallot S, Gottwald T, Decroocq V, Jacquot E, Soubeyrand S, Thébaud G (2015) Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annual Review of Phytopathology 53:357–378
26. Rimbaud L, Bruchou C, Dallot S, Pleydell DRJ, Jacquot E, Soubeyrand S, Thébaud G (2018) Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. Royal Society Open Science 5(1):171435
27. Rimbaud L, Dallot S, Bruchou C, Thoyer S, Jacquot E, Soubeyrand S, Thébaud G (2018) Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. BioRxiv 315747
28. Rios LM, Sahinidis NV (2013) Derivative-free optimization: a review of algorithms and comparison of software implementations. Journal of Global Optimization 56(3):1247–1293
29. Roustant O, Ginsbourger D, Deville Y (2012) DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. Journal of Statistical Software 51(1):1–55, URL `http://www.jstatsoft.org/v51/i01/`

30. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N (2016) Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE 104(1):148–175
31. Snelson E, Ghahramani Z, Rasmussen CE (2004) Warped gaussian processes. In: Advances in neural information processing systems, pp 337–344
32. Snoek J, Swersky K, Zemel R, Adams R (2014) Input warping for bayesian optimization of non-stationary functions. In: International Conference on Machine Learning, pp 1674–1682
33. Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse ME, Grenfell BT, Keeling MJ (2006) Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 440(7080):83
34. VanderWaal K, Enns EA, Picasso C, Alvarez J, Perez A, Fernandez F, Gil A,



Supplementary Figure 1 Results of the algorithm using empirical distribution functions [10] with (yellow) and without (blue) warping ($\alpha \in [0;18,012.12]$).

Résultats clés de l'Article 5

AMELIORATION DE L'OPTIMISATION BAYESIENNE D'UN MODELE EPIDEMIOLOGIQUE COMPLEXE DU VIRUS DE LA SHARKA PAR L'UTILISATION DU WARPING

- **Développement d'une approche d'optimisation prenant en compte les invariances locales**
  - Une approche d'optimisation bayésienne a été modifiée pour prendre en compte les invariances locales des paramètres d'un modèle. Ces invariances correspondent aux configurations où, par construction du modèle, la variable de sortie prend la même valeur pour plusieurs combinaisons de valeurs des variables d'entrée.
  - Cette nouvelle approche est basée sur la distorsion de l'espace des paramètres d'entrée.

- **La distorsion permet d'accélérer la convergence de l'algorithme d'optimisation**
  - L'approche d'optimisation bayésienne développée a été testée sur le modèle simulant la propagation et la gestion de la sharka. La contribution de l'étape de distorsion sur l'optimisation des stratégies de gestion a été analysée.
  - L'étape de distorsion ne permet pas d'améliorer le résultat de l'optimisation si suffisamment d'itérations sont réalisées. Cependant, elle permet de converger plus rapidement vers l'optimum.

ARTICLE 6

In silico optimization of a strategy for landscape-wide plant disease management

Coralie Picard, Victor Picheny, François Bonnot, Samuel Soubeyrand and Gaël Thébaud

In silico optimization of a strategy for landscape-wide plant disease management

Coralie Picard[1], Victor Picheny[2], François Bonnot[1], Samuel Soubeyrand[3], Gaël Thébaud[1*]

[1] BGPI, INRA, Montpellier SupAgro, Univ. Montpellier, Cirad, TA A-54/K, Campus de Baillarguet, 34398, Montpellier Cedex 5, France
[2] MIAT, INRA, 31326, Castanet-Tolosan, France
[3] BioSP, INRA, 84914, Avignon Cedex 9, France.
* Corresponding author
Email: gael.thebaud@inra.fr (GT)

Abstract

Plant and animal diseases are generally managed at the scale of individual farms by making 'one problem, one solution' tactical decisions, often based on the use of treatments. To reduce between-farm transmission, and thus disease prevalence, landscape-scale disease management can be used. Such management is motivated by an objective of reducing treatments or by the need for collective action to ensure the control of non-treatable and/or quarantine pathogens. However, identifying an efficient landscape-scale management is not easy because the management can depend on numerous parameters, and experiments are often impossible. Therefore, models have been used to optimize these parameter values. Until now, this approach has been applied mostly to deterministic models with few parameters because it does not easily scale up to more complex management strategies embedded in spatially-explicit stochastic epidemic models. Here, we show how a generic *in silico* approach built on a global optimization algorithm can be used to optimize plant disease management. We apply this approach to sharka, the most damaging disease of *Prunus* trees, whose management involves surveillance, removals and plantation bans. These actions provide many degrees of freedom in the definition of landscape-wide surveillance intensity and removal of infected individuals. Here, we propose to optimize this management strategy by using a spatiotemporal stochastic model simulating epidemic dynamics and management on three landscape types differing by their level of patch aggregation. More specifically, we identified optimized combinations of parameter values leading to the highest net present value (NPV), an economic criterion balancing management costs and the profit generated by productive trees. For both emerging and established epidemics, we identified strategies that are more profitable than the current French strategy. It turns out that some strategies are effective for all landscapes, which has interesting implications in practice. Such optimization process can be applied to other complex disease management issues.

Author summary

Plant diseases are complex because they depend on pathogen characteristics, human interventions, as well as the organization of the patches in the landscape. Thus, identifying efficient control strategies to limit the epidemic damage constitutes a major challenge. The design of management strategies often rely on expert opinions, although they are not based on field trials at large spatiotemporal scales. Therefore, these strategies are not necessarily optimal for various landscapes and several years. Here, we present an approach to optimize the landscape-wide management of a plant disease. This approach is based on an algorithm able of identifying the most efficient management strategies of a plant disease on a model simulating pathogen dispersal and management. This approach could be very useful for risk managers who provide advices or propose law texts to manage a plant pathogen.

Introduction

The control of infectious diseases is often based on the use of pharmaceutical products for animals and pesticides for plants (and disease vectors). Global targets to reduce pesticide use, along with the increasing number of disease (re-)emergence events, have fostered the development of more complex management strategies combining surveillance and control actions at the landscape scale. The rationale for coordination beyond the scale of the individual farm is to prevent disease spread by matching the intrinsic spatial and temporal scales of the epidemic [1]. However, such strategies depend on the combination of various management options whose parameters are difficult to optimize because wide-scale experiments are often impossible (for ethical, logistical and economic reasons).

To overcome these experimental limitations, epidemiological models have been developed and helped deciding how to control invasive pathogens [2]. Thanks to their ability to test *in silico* a wide range of epidemic and management scenarios at large spatiotemporal scales [3], models can rapidly identify promising management strategies and assess their long-term effects [4]. For instance, models helped to identify optimal vaccination and culling strategies for animal diseases [5–7] and to optimize livestock surveillance [8–10]. In plant health, some studies optimized removal dates and areas [11–16], sampling frequency and intensity [17] or space between host plants [11,18]. However, these studies mostly focused on a single management parameter and not on complex management strategies (with several parameters).

Studies from others scientific disciplines optimized several parameters at once (i.e. found the set of input management parameters of a model that maximizes or minimizes the output of interest: cost, production, etc.) [19–23]. However, few of them deal about the improvement of disease control [24]. Such optimization can be challenging because the parameter space to explore is frequently very large, especially in the presence of interactions between parameters. Usually, the algorithms used to overcome this difficulty follow the same basic steps [25]: (i) generation of candidate management strategies, (ii) simulation, (iii) evaluation and selection, and (iv) possibly further loops where new candidates are generated based on the first results (if the parameters are not numerous, it is possible to present a complete list of all possibilities, but this cannot be applied to systems which combine many parameters with many options). Among the numerous optimization methods [26], nonintrusive approaches are increasingly used to solve such optimization problems. These approaches only require pointwise evaluations of the model at hand (an output value for a given set of inputs), and do not require knowledge of the underlying mechanisms of the model [19,24,27]. In

addition, epidemiological studies aiming to optimize outbreak management are faced with a major problem: integrating the epidemic variability, which partly comes from host and pathogen characteristics. To account for this variability, models are often stochastic (a management strategy does not have the same impact depending on the epidemic), which requires specific optimization algorithms. In the last few years, algorithms for optimization with heterogeneous noise have been developed [28]. They have already been used in several disciplines [29,30] and could be applied to optimization problems in disease management.

Here, we aim to optimize a complex plant disease management at the landscape scale using a stochastic epidemiological model. Our approach is applied to sharka which causes much damage on prunus trees [31,32]. In France, a national decree defines a management strategy to control this disease [33], which requires orchard surveillance, removal of symptomatic trees (or sometimes whole orchards), as well as plantation restrictions. This is a complex strategy defined in a high-dimensional parameter space (Fig 1, S1 Table, column 'French management'). Previous studies already identified management strategies that are more efficient than the present French strategy for either a single specific landscape [34] or for various landscapes differing by their aggregation level [35]. However, these strategies were derived from the results of sensitivity analyses, which evaluate around 300,000 predefined parameter combinations spread throughout the parameter space. Even such a vast number of parameter combinations sparsely sample the parameter space; thus, such approach risks missing better combinations of disease management parameters. Thus, in this article, we optimized sharka management strategy using a numerical algorithm (adapted to stochastic optimization problems) which can explore parameter space more thoroughly. This algorithm is presented in the materials and methods section.

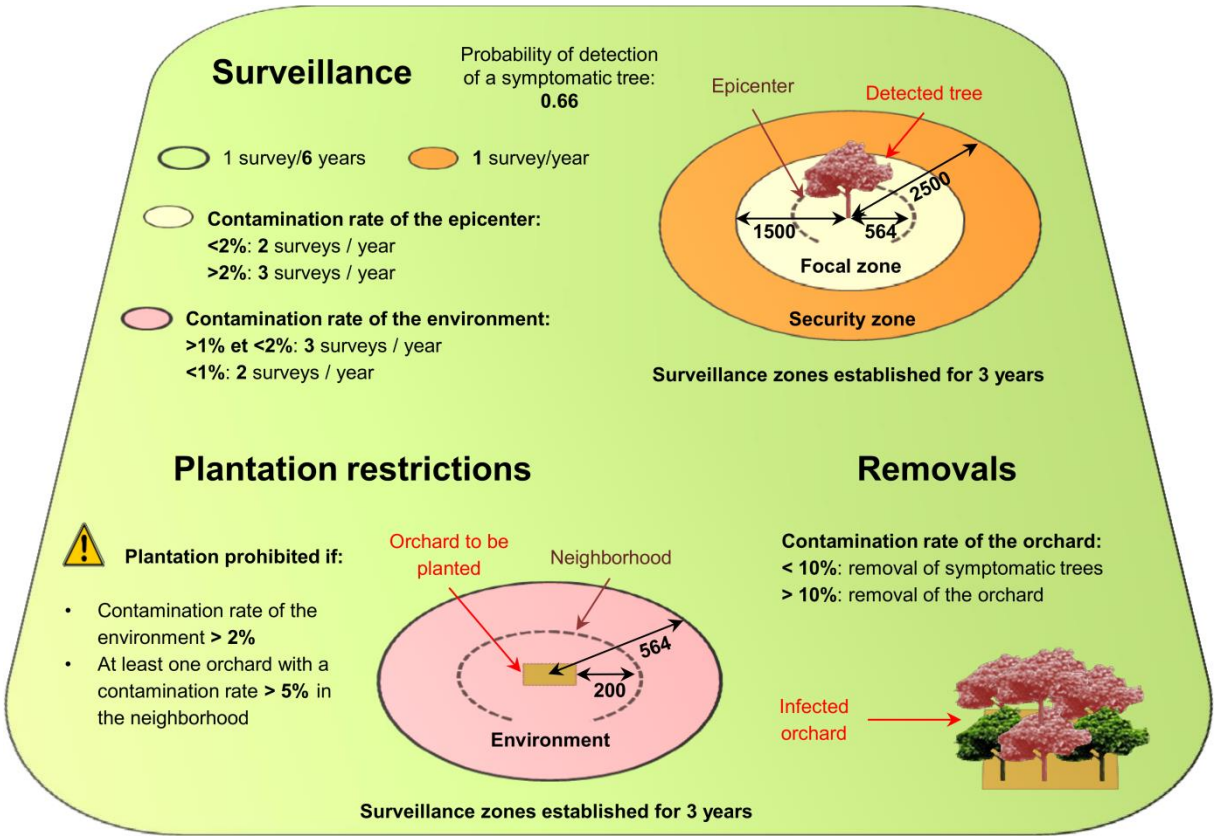


Fig 1. Management actions currently applied in France against sharka.

Materials and methods

Simulation of sharka spread and management

To simulate sharka outbreaks, we used a stochastic, spatially explicit, SEIR (susceptible-exposed-infectious-removed) model that was previously developed [34–37]. This model is orchard-based and works with a discrete time step of one week. It simulates disease spread and management in landscapes varying by the aggregation level of the patches on which peach trees are grown.

More specifically, at the beginning of the simulation (year 1), each orchard is set up with a specified age and a removal date. Here, we draw these dates from their exact asymptotic distribution as presented in S1 Text (rather than simulating them as described by Rimbaud et al. and Picard et al. [34,35]). Then, during simulations, trees are characterized by different states: “susceptible” (healthy), “exposed” (infected but not yet infectious or symptomatic), “infectious hidden” (after the end of the latent period), “infectious detected” (when specific symptoms are detected on the tree during a survey) and "removed" (when the tree is removed from the patch). The epidemic process and the

transitions between the different states can be found in [34,36,37]. In this study, we also modified the orchard plantation density compared to [34,35] : we used 719 trees/ha.

Different model parameters enable us to simulate a wide range of epidemic management scenarios. The epidemic itself is characterized by 6 epidemiological parameters [36] (Table 1). Inspired by sharka management in France and the US, flexible management options are implemented using 21 parameters (this management is illustrated in the figure 2 in [38], and detailed in S1 Table). Management starts a predefined number of years after virus introduction. In addition, simulations can be performed on 3 landscape types varying by their level of patch aggregation (landscape H: high aggregation level, landscape M: medium aggregation level, landscape L: low aggregation level, [35]). The model includes 30 landscape replicates of the 3 aggregation levels (these 90 landscapes are composed of the same number of patches). In simulations performed for one aggregation level, stochasticity stems from the random sampling of (i) the landscape (among the 30 landscapes of the corresponding aggregation level), (ii) the epidemiological parameters among their value ranges (Table 1), (iii) parameter values in their distribution (throughout the simulations) [36].

Table 2. Variation ranges of epidemiological parameters for emerging and established epidemics. Values in bold highlight the differences between emerging and established epidemics.

		Before management				**During management**	
		Emerging epidemics		**Established epidemics**		**Emerging and established epidemics**	
		Min	**Max**	**Min**	**Max**	**Min**	**Max**
q_k	Quantile of the connectivity of the patch of first introduction	0	1	0	1	0	1
φ	Probability of introduction at plantation	0,0046	0,0107	**0,02**	**0,02**	0,0046	0,0107
p_{MI}	Relative probability of massive introduction	0	0,1	**0,4**	**0,4**	0	0,1
W_{exp}	Expected value of the dispersal weighting variable	0,469	0,504	0,469	0,504	0,469	0,504
β	Transmission coefficient	1,25	1,39	1,25	1,39	1,25	1,39
θ_{exp}	Expected duration of the latent period (years)	1,71	2,14	1,71	2,14	1,71	2,14

The model output is an economic criterion, the net present value (NPV), which accounts for the benefit generated by the cultivation of productive trees and the costs induced by fruit production

and disease management actions (including surveillance, removals and replantations). It is calculated as described by [34], with a slight modification of the cost of access to a patch for surveys (S2 Text).

Optimization scenarios

Because the model is stochastic, several simulations were necessary to assess the result of a combination of management parameters, and thus to optimize them. We optimized the sharka management strategy on the basis of 2 criteria: the mean NPV (noted $\overline{\mathrm{NPV}}$) and the mean of the 10% "worst" NPVs obtained with simulations including various epidemiological parameters and several landscapes (noted $\mathrm{NPV}_{\overline{10\%}}$). This second criterion was chosen to reduce the likelihood of significant losses. It corresponds to a measure of risk aversion while being numerically more stable than the quantile because it accounts for the entire distribution tail. In addition, this criterion is not as volatile as the worst NPV case. Note that $\mathrm{NPV}_{\overline{10\%}}$ is sometimes referred to as conditional value-at-risk (CVaR, [39]) or as Bregman's superquantile [40,41].

We optimized strategies for 3 landscapes differing by their level of patch aggregation (low, medium and high), and for 2 types of epidemics (emerging and established epidemics). These epidemic types differ in the duration of virus spread without management at the beginning of the simulation (5 years and 15 years for emerging and established epidemics, respectively), and by the values of two epidemiological parameters during this period: at plantation, the probability of sharka introduction and the probability of a massive introduction are higher for established epidemics (Table 1). Disease management is then applied during 30 years for both epidemic types, and the NPV is calculated over this period.

Optimization algorithm

Optimizing sharka management was challenging since the evaluation of the $\overline{\mathrm{NPV}}$ or $\mathrm{NPV}_{\overline{10\%}}$ for a given strategy required repeated calls to the simulator. However, the number of replicates and the total number of evaluated strategies were severely limited by the overall computational cost. As a first consequence, the $\overline{\mathrm{NPV}}$ or $\mathrm{NPV}_{\overline{10\%}}$ were accessible only through noisy estimates (i.e. the estimation error could not be neglected), hence requiring the use of an algorithm adapted to stochastic optimization problems. Secondly, the small number of evaluated strategies implied the use of a parsimonious algorithm (ruling out e.g. most metaheuristics of pattern search algorithms). Thus, we followed a Bayesian optimization (BO) strategy and used an algorithm adapted from the R package DiceOptim [42], initially proposed in [43]. In short, BO works as follow: a first set of strategies chosen evenly distributed in the design space [44], is evaluated by running the simulator. A

kriging model [45] is fitted to these data. Then, additional strategies for which the simulator is run are chosen sequentially according to a so-called infill criterion calculated using the kriging model, the model being updated after each new value is calculated.

Here, we conducted an independent optimization for each epidemic level (emerging and established epidemics) and patch aggregation level, and for both $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$. For the initial BO step, we randomly selected 400 management strategies using a maximin Latin hypercube sampling design [44]. Then, 1000 new strategies were chosen sequentially following the algorithm in [43], resulting in a total of 1400 evaluated strategies. For each evaluated strategy, 1000 repeated simulations were carried out. The standard deviation of the resulting $\overline{\mathrm{NPV}}$ or $\mathrm{NPV}_{\overline{10\%}}$ estimate was calculated, using respectively the sample standard deviation and bootstrap.

The optimization variables contained both continuous and discrete elements, which made it challenging to maximize the infill criterion and prevented us from using directly the DiceOptim package. To address this issue, the infill criterion was maximized over 100,000 or 110,000 randomly generated candidate points (Fig. 2). Finally, a warping step was applied before each step of candidate point generation to reduce parameter space and remove the combinations leading to the same management (Picheny et al. in prep). For instance, when the radius of a zone was 0 for a parameter combination, the associated surveillance frequencies had no impact on the NPV. Thus, the algorithm did not need to explore all combinations but only those that propose a different management.

To summarize, we performed optimizations for 2 criteria ($\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$) x 3 landscape types x 2 epidemic levels (emerging and established epidemics), making a total of 12 optimizations (see Fig 2 for an example of the optimization process of the $\overline{\mathrm{NPV}}$ for one landscape type). Then, with each optimized strategy, we performed 10,000 simulations of sharka epidemics on the 3 landscapes to obtain (almost) noiseless estimates of $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$. As a final step we used the knowledge gained from the results of the independent optimizations to simplify the optimized strategies. For example, if a strategy indicates that the trees located in a focal zone have to be surveyed more frequently when a specified threshold is reached in the epicenter, and that this threshold is very high (rarely reached), the corresponding action (strengthened surveillance frequency) was removed from the model. Then, we tested some strategies by mixing elements of several optimized strategies. Two analyses were carried out to assess the robustness of the identified management strategies in different epidemic contexts: (i) the best optimized strategy found for emerging epidemics was tested in the case of established epidemics and vice versa, and (ii) the optimized strategies were evaluated for doubled and tripled values for the range of the transmission coefficient β.

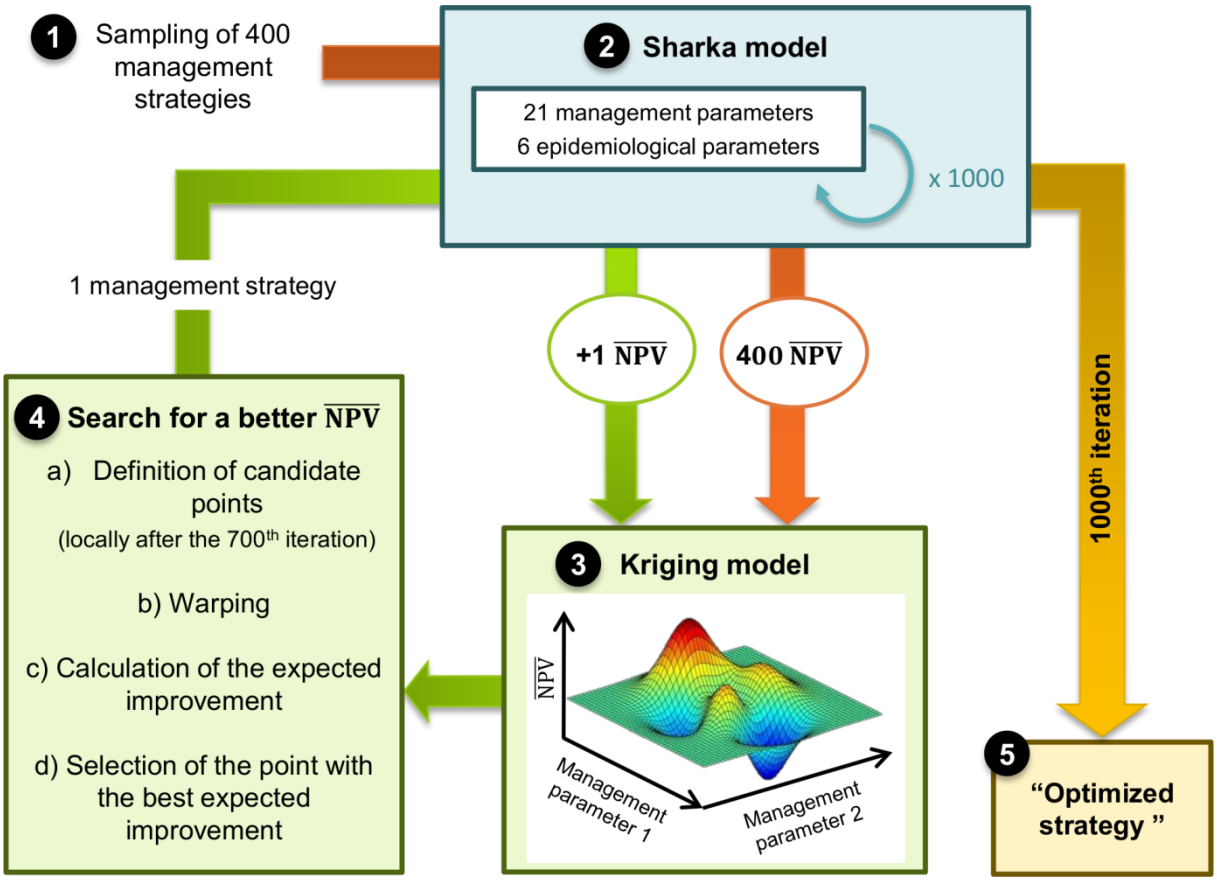


Fig 2. Optimization process of the NPV mean ($\overline{NPV}$) for one level of patch aggregation. 1. Sampling of 400 management strategies (one strategy corresponds to 21 management parameters). 2. Simulations of each management strategy for 1,000 different epidemics (one epidemic corresponds to 6 epidemiological parameters), and calculation of the $\overline{NPV}$. 3. Definition of a kriging model: example of a model with 2 management parameters. 4. Search for a better $\overline{NPV}$: a) before the 700th iteration, 100,000 candidate points of the parameter space are chosen randomly and 10,000 more are chosen locally around the best current $\overline{NPV}$; after the 700th iteration, 110,000 candidate points of the parameter space are chosen locally around the best current $\overline{NPV}$, b) warping step is applied to reduce the parameter space and remove the parameter combinations which lead to the same management, c) the expected improvement (EI) is calculated for all the candidate points (the highest this value, the more the model considers that this combination must be explored to optimize the result), d) the combination with the highest EI is selected and 1,000 simulations of epidemic and management (corresponding to this combination) are carried out. The kriging model is then updated with the new $\overline{NPV}$. Steps 2, 3 and 4 are repeated 1000 times. 5. At the end of iteration 1000, the best strategy is selected.

Results

Optimization results

The optimization algorithm used for emerging and established epidemics allowed to improve progressively the $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ for all our criteria (Fig 3). Both criteria were firstly improved after 700 iterations, then after the local optimization (1000 iterations), and to finish, with the simplification step (S2 Table). In addition, by mixing elements of several optimized strategies, we identified for each epidemic type a strategy that is efficient for both criteria ($\overline{\mathrm{NPV}}$, $\mathrm{NPV}_{\overline{10\%}}$) and for all landscape aggregation levels (Fig 3). For instance, taking the French management strategy as a reference, for landscape H the $\overline{\mathrm{NPV}}$ was improved from 22,073 €/ha to 27,045 €/ha and the $\mathrm{NPV}_{\overline{10\%}}$ from 11,698 €/ha to 22,897 €/ha for emerging epidemics, and for established epidemics the $\overline{\mathrm{NPV}}$ was improved from 973 €/ha to 17,455 €/ha and the $\mathrm{NPV}_{\overline{10\%}}$ from -24,587 €/ha to 1,907 €/ha.

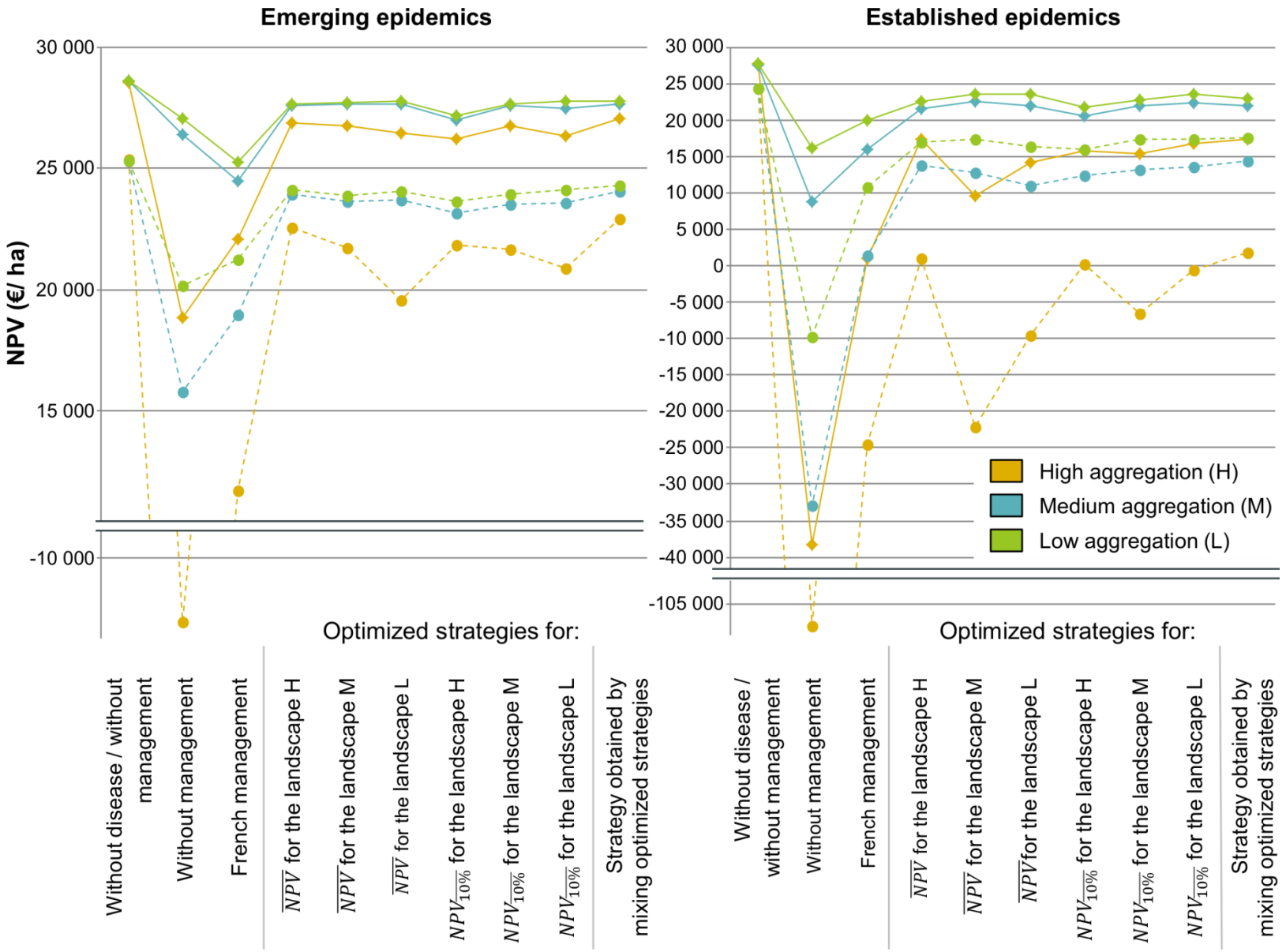


Fig 3. $\overline{\mathbf{NPV}}$ and $\mathbf{NPV_{\overline{10\%}}}$ obtained after 10,000 simulations of PPV dispersal and management. The $\overline{\mathbf{NPV}}$ is represented by solid lines and the $\mathbf{NPV_{\overline{10\%}}}$ by dotted lines. For each epidemic type, the presented optimized strategies correspond to the 6 strategies obtained after the simplification step, and the strategy obtained by mixing the optimized strategies.

The optimized strategies for both emerging and established epidemics (obtained by mixing some optimized strategies) are simpler to implement in practice than the present French management (Fig 4, S1 Table). Indeed, only the symptomatic trees need to be removed, no plantation ban is imposed, and a single surveillance zone is required (with no particular surveillance for young orchards).

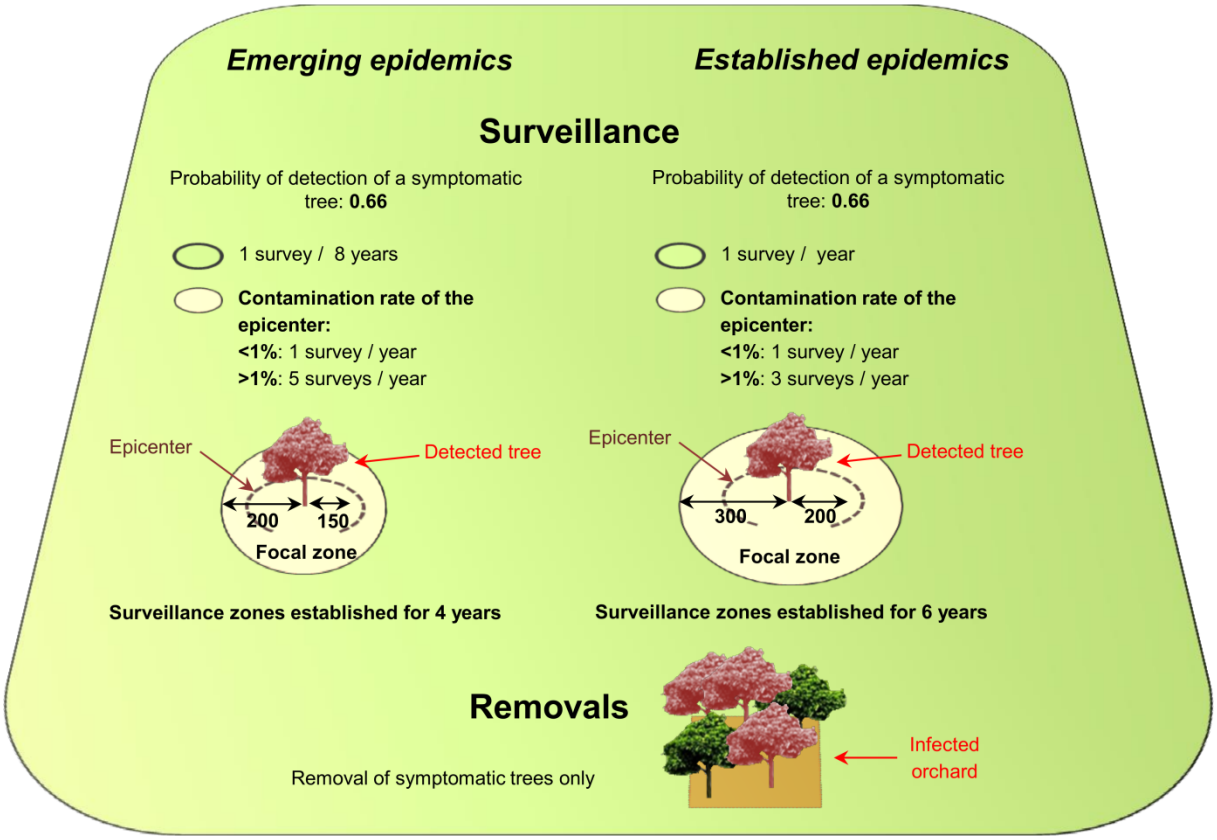


Fig 4. Management actions for the optimized strategies for emerging and established epidemics.

Details of strategy impact on NPV components

The assessment of optimized strategies impact on the NPV components (Fig 5) showed that financial products (due to fruit sales) is the component with the highest impact, followed by surveillance costs and then by plantation and removal costs (which have a minor impact on the NPV). Plantation bans and whole orchard removals (which reduce fruit sales quantity) thus have a strong impact on the NPV. This explains why the French management strategy (which imposes plantation bans and orchard removals) leads to less fruit sales than the optimized strategies for both emerging and established epidemics and for all landscape types.

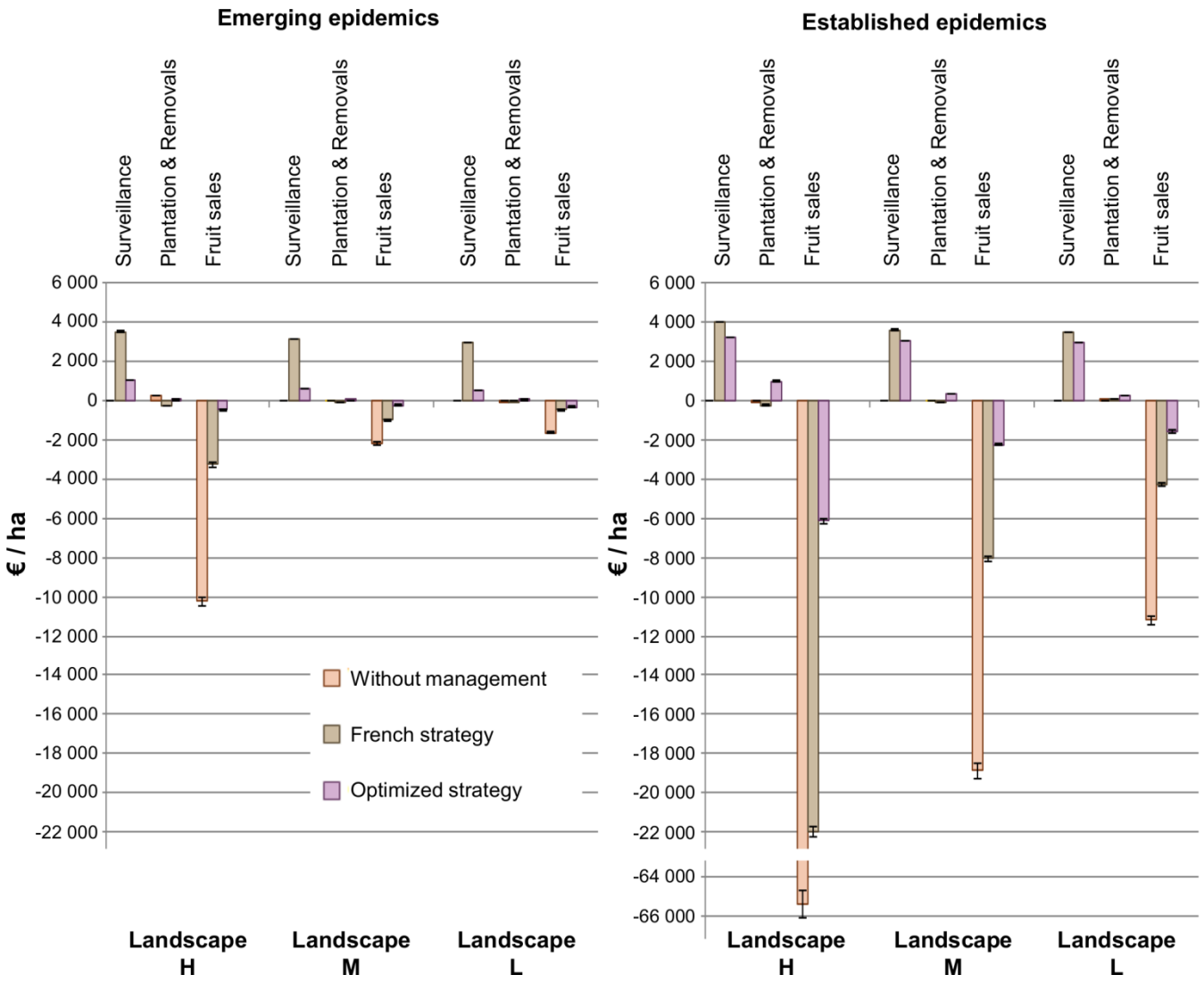


Fig 5. Comparison of the details of NPV components. Barplots represent the difference between simulations without disease (and without management) and (i) with disease and without management (red), (ii) with the French management strategy (brown), and (iii) with the optimized strategy (purple). When the value is higher than 0, costs or products of the simulation are higher than those of simulations without disease (and vice versa). Simulations were performed 10,000 times on landscapes with three levels of patch aggregation: high (H), medium (M) and low (L).

The strong impact on the NPV of the parameters corresponding to removals and plantation bans can be visualized on Fig. 6 and S1 Fig [46]. Indeed, by setting all the management parameters to their optimal value, and by modifying individually the values of the parameters corresponding to removals and plantation bans, we can observe a high fluctuation of the NPV values. For instance, parameter χ_R (contamination threshold in the removal epicenter, above which orchards inside the removal zone are removed) has a strong influence on the NPV (the lower its value, the highest the number of whole orchards removed and the lower the NPV). Conversely, the variation of surveillance parameters does not have influence on the NPV regarding the metamodels. However, it is important

to keep in mind that if all parameters were not set to their optimal values, the influence of the observed parameters on the NPV could have been different.

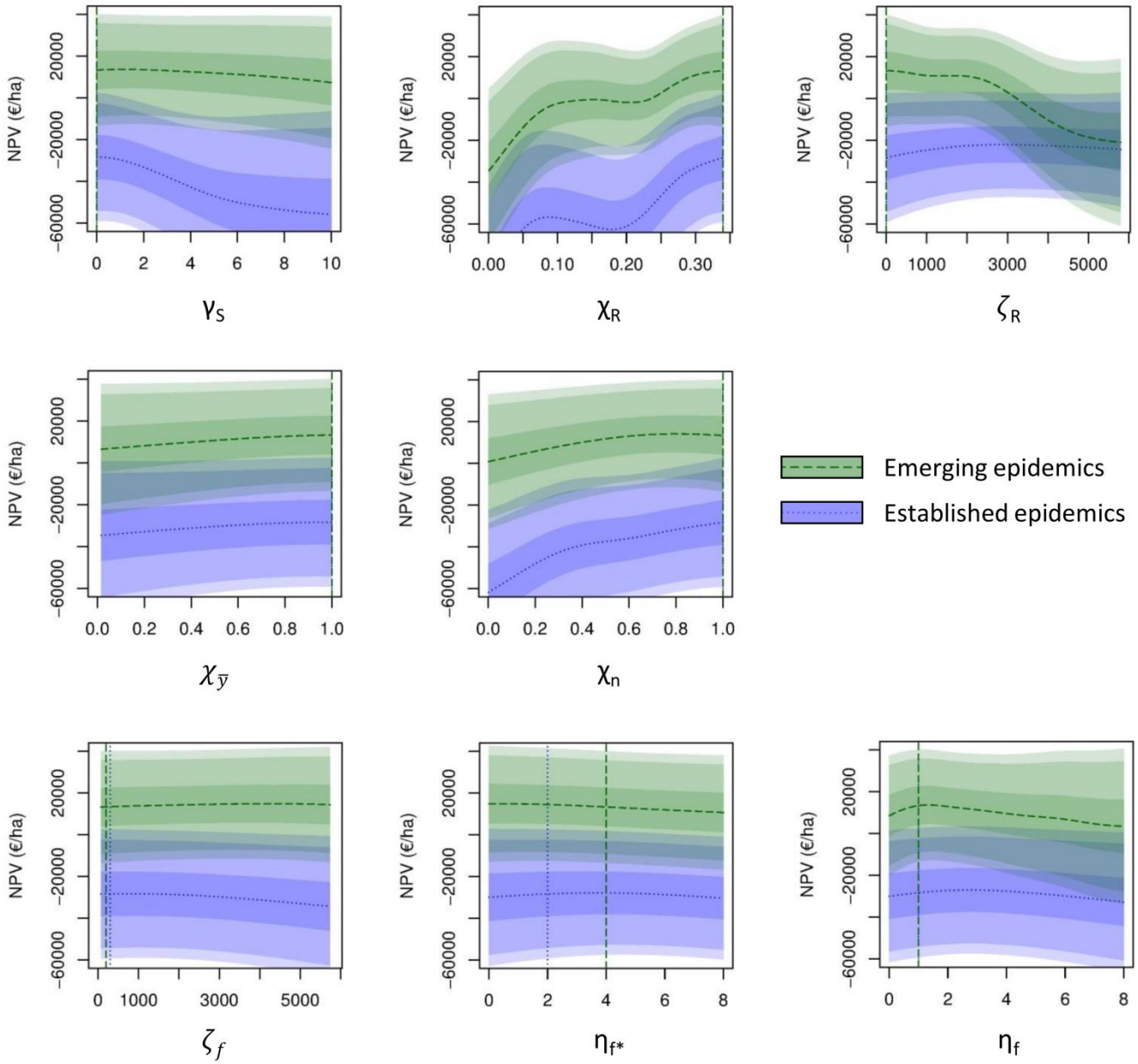


Fig 6. 2D view of kriging metamodels. These metamodels were obtained at the end of the optimization of the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ for the most aggregated landscape (H) for emerging (green) and established (blue) epidemics. Each plot represents the influence of a single management parameter on the NPV, setting the values of the other management parameters to their optimal value. Here, the parameters correspond to removals (top row), plantation bans (middle row) and surveillance (bottom row). Vertical dotted lines correspond to the optimized parameter values and breaks between different blue and green colors indicate the 50th, 80th, 90th, 95th and 99th percentiles (representing uncertainty in the kriging model). As an example, when the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ keeps the same value regardless of the management parameter value, this parameter have no influence on the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ (if the other parameters are setting to their optimal values).

Robustness of optimized strategies

In order to assess the robustness of the optimized strategies in different epidemic contexts, the best optimized strategy found for established epidemics (strategy obtained by mixing optimized strategies) was tested in the case of emerging epidemics. The $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ were lower than in simulations performed with the strategy optimized in the case of emerging epidemics for all landscape types (S3 Table). Then, we carried out simulations with the best strategy identified for emerging epidemics (strategy obtained by mixing optimized strategies) in the case of established epidemics. As previously, The $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ were lower than in simulations performed with the strategy optimized in the case of established epidemics except for the $\overline{\mathrm{NPV}}$ obtained with the landscapes M and L (this can be explained by the fact that the optimized strategy mixes various strategies, and is not optimal for all the criteria). Finally, the optimized strategies are globally less efficient if they are not applied in the epidemic context for which they have been optimized. However, they still remain more profitable than the French management strategy.

Then, we performed simulations with the best optimized strategies for emerging and established epidemics in more severe epidemic contexts (with doubled and tripled bounds of β). These strategies were still efficient and were more profitable than the French management, whether they are performed on emerging or strong epidemics (S3 Table).

Discussion

In this study, we showed how a generic *in silico* approach based on a global optimization algorithm can be used to optimize plant disease management. This approach was applied to sharka, for which a complex management strategy is enforced in France. We used a recently developed method (based on a kriging metamodel) on a spatiotemporal model simulating sharka dispersal and management, in order to sparingly explore the space of possible management strategies and to optimize an economic criterion. In particular, we attempted to optimize the mean of the NPV ($\overline{\mathrm{NPV}}$) and the mean of the 10% lower NPV ($\mathrm{NPV}_{\overline{10\%}}$) for 3 levels of patch aggregation and 2 types of epidemics (emerging and established epidemics). For each epidemic type, we identified an optimized strategy that is efficient for all landscape types and for both economic criteria ($\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$). These strategies are more efficient than the French management strategy and easier to implement in practice.

Relevance of the method

The approach presented in this study was adapted to our optimization problem. Indeed, we were able to obtain not only better NPVs than the French management strategy for all our criteria ($\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$, for 3 landscape types), but also than a previous work which aimed to improve sharka management using the results of a sensitivity analysis ([35], S3 Table). However, with its 21 management parameters, the strategy was a challenge to most optimization algorithms (especially since the model was stochastic). To succeed, we used a global optimization algorithm because previous works (Rimbaud et al. 2018, Picard et al. 2018) had shown that the underlying function of our model was multimodal (with several local maxima). The specific algorithm that we used explores the whole design space to avoid getting trapped around local optima, while using local intensification to locate the optimum more precisely. In addition, such approach is based on an approximation of the objective function (metamodel), which is a basic tool for handling complex models [47]. A major difficulty was model stochasticity caused by the variability of the epidemics. For this reason, during the optimization process the $\overline{\mathrm{NPV}}$ (or $\mathrm{NPV}_{\overline{10\%}}$, depending on the criterion to optimize) was calculated at each iteration. The final optimized strategy is the one that leads to the best $\overline{\mathrm{NPV}}$ (or $\mathrm{NPV}_{\overline{10\%}}$) among the performed iterations. Thus, an accurate estimation of the optimization criterion is necessary to prevent inadvertent selection of a suboptimal management strategy. For this reason, we performed iterations with 1000 replicates (our initial attempts with 100 replicates were not satisfactory, probably because the estimated $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ were not accurate enough). To improve the accuracy of the estimation of the optimization criterion, we might perform more than 1000 replicated simulations at each iteration, however, the calculation time could quickly become excessive. Indeed, to perform an optimization with 1000 iterations and 1000 simulation repetitions, about 45 days were necessary in this study.

Several methodological developments may be pursued in the future. First, in our setup we fixed the number of repeated simulations (to 1000) for all strategies. Intuitively, a substantial gain in efficiency could be achieved by adapting the number of replicates on the fly, as previously suggested [48,49] in order to avoid spending time on poor strategies and obtain more accurate estimates for the best ones. However, such approach was not followed since this is still an open question in the optimization community [28]. In addition, independent kriging models and optimization runs were carried out for each landscape and epidemic type. A more complex but more efficient solution might be to fit a single kriging model to all landscape and epidemic conditions by considering conditions as qualitative factors [50]. Finally, a multiobjective setup could be considered, either by optimizing jointly the $\overline{\mathrm{NPV}}$ (or $\mathrm{NPV}_{\overline{10\%}}$) for all landscape and epidemic types, or by optimizing jointly the $\overline{\mathrm{NPV}}$

and $\mathrm{NPV}_{\overline{10\%}}$ for each landscape and each epidemic condition. The first would allow us to analyze in details the trade-offs between average performance and risk-averse strategies, while the second would highlight parameters that differ depending on the landscape and epidemic conditions.

Practical implications

In terms of practical application, our results suggest that the French management strategy might be improved in order to optimize the NPV. Although landscape characteristics may influence epidemic spread [1,35,51], we identified optimized strategies efficient for all landscape types. This is particularly important for stakeholders because it can be difficult to delineate zones that differ by their level of landscape aggregation. In addition, these strategies are less complex than the French management since they do not include plantation bans and only require the removal of symptomatic trees and one surveillance zone (no particular surveillance for young orchards).

We showed that such simplification allows a significant reduction of surveillance costs and an increase of products due to fruit sales (Fig 5), resulting in higher NPVs. Such results represent significant economic savings considering the 11,000 ha of peach orchards cultivated in France [52]. Indeed, on average 55 million euros could be saved for landscape H for emerging epidemics and 28 million for landscape L (182 and 33 million, respectively, for established epidemics), and 124 million euros and 34 million for the $\mathrm{NPV}_{\overline{10\%}}$ (293 and 75 million for established epidemics) over a period of 30 years. In addition, if these optimized strategies are applied in other epidemic contexts, they are still more economically efficient than the French management strategy.

However, the optimized strategies can be less efficient if they are not used in the epidemic context for which they have been optimized. In practice, stakeholders might adapt the management regarding the epidemic conditions of a particular region. In addition, we attempted in this study to provide other relevant information to stakeholders, to enable them to choose the strategy to apply. Firstly, we optimized here an economic criterion balancing costs and benefits of a disease management strategy and not epidemiological criterions, as many studies do [11,12,14–18,53–55]. Then, we also accounted for the level of risk aversion of decision-makers by optimizing on $\mathrm{NPV}_{\overline{10\%}}$ (allowing to limit the proportion of epidemics causing substantial economic damage for a particular management strategy) because the strategy efficiency depends on the percentile of the criterion to optimize [4,12].

To go further, this approach might be applied on other diseases by changing epidemic and management parameters, although several model assumptions are specific to the sharka

pathosystem and should be modified. It could be interesting in particular for diseases that require collective action and for which it is impossible to test management strategies in field trials.

Acknowledgments

The authors thank Raynald Havard for their contribution to the definition of age and orchard duration at the initial stage of the model. This work was supported by the CIRAD-UMR AGAP HPC Data Center of the South Green Bioinformatics platform (http://www.southgreen.fr), and we are grateful to Sébastien Ravel for the associated help.

References

1. Gilligan C, Truscott JE, Stacey AJ. Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. J R Soc Interface. 2007; 4(16): 925–934. doi:10.1098/rsif.2007.1019.

2. Cunniffe NJ, Koskella B, Metcalf CJ, Parnell S, Gottwald TR, Gilligan CA. Thirteen challenges in modelling plant diseases. Epidemics. 2015; 10: 6–10. doi:10.1016/j.epidem.2014.06.002.

3. Parnell S, van den Bosch F, Gottwald T, Gilligan CA. Surveillance to inform control of emerging plant diseases: an epidemiological perspective. Annu Rev Phytopathol. 2017; 55: 591–610. doi:10.1146/annurev-phyto-080516-035334.

4. Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc Natl Acad Sci. 2016; 113(20): 5640–5645. doi:10.1073/pnas.1602153113.

5. Mushayabasa S, Tapedzesa G. Modeling the Effects of Multiple Intervention Strategies on Controlling Foot-and-Mouth Disease. BioMed Research International. 2015. doi:10.1155/2015/584234.

6. Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse MEJ, et al. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature. 2006; 440(7080): 83. doi:10.1038/nature04324.

7. Vial F, Cleaveland S, Rasmussen G, Haydon DT. Development of vaccination strategies for the management of rabies in African wild dogs. Biol Conserv. 2006; 131(2): 180–192. doi:10.1016/j.biocon.2006.04.005.

8. Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. J R Soc Interface. 2012; 9(76): 2814–2825. doi:10.1098/rsif.2012.0289.

9. Kompas T, Ha PV, Nguyen HTM, East I, Roche S, Garner G. Optimal surveillance against foot-and-mouth disease: the case of bulk milk testing in Australia. Aust J Agric Resour Econ. 2017; 61(4): 515–538. doi:10.1111/1467-8489.12224.

10. VanderWaal K, Enns EA, Picasso C, Alvarez J, Perez A, Fernandez F, et al. Optimal surveillance strategies for bovine tuberculosis in a low-prevalence country. Sci Rep. 2017; 7(1): 4140. doi:10.1038/s41598-017-04466-2.

11. Cunniffe NJ, Laranjeira FF, Neri FM, DeSimone RE, Gilligan CA. Cost-effective control of plant disease when epidemiological knowledge is incomplete: modelling Bahia Bark Scaling of Citrus. PLOS Comput Biol. 2014; 10: e1003753. doi:10(8).1371/journal.pcbi.1003753.

12. Cunniffe NJ, Stutt RO, DeSimone RE, Gottwald TR, Gilligan CA. Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. PLOS Comput Biol. 2015; 11(4): e1004211. doi:10.1371/journal.pcbi.1004211.

13. Hyatt-Twynam SR, Parnell S, Stutt RO, Gottwald TR, Gilligan CA, Cunniffe NJ. Risk-based management of invading plant disease. New Phytol. 2017; 214(3): 1317–1329. doi:10.1111/nph.14488.

14. Parnell S, Gottwald TR, van den Bosch F, Gilligan CA. Optimal strategies for the eradication of asiatic Citrus Canker in heterogeneous host landscapes. Phytopathology. 2009; 99(12): 1370–1376. doi:10.1094/PHYTO-99-12-1370.

15. Parnell S, Gottwald TR, Gilligan CA, Cunniffe NJ, van den Bosch F. The effect of landscape pattern on the optimal eradication zone of an invading epidemic. Phytopathology. 2010; 100(7): 638–644. doi:10.1094/PHYTO-100-7-0638.

16. Sisterson MS, Stenger DC. Roguing with replacement in perennial crops: conditions for successful disease management. Phytopathology. 2012; 103(2): 117–128. doi:10.1094/PHYTO-05-12-0101-R.

17. Parnell S, Gottwald TR, Riley T, van den Bosch F. A generic risk-based surveying method for invading plant pathogens. Ecol Appl. 2014; 24(4): 779–790. doi:10.1890/13-0704.1.

18. Jeger MJ, Chan MS. Theoretical aspects of epidemics: uses of analytical models to make strategic management decisions. Can J Plant Pathol. 1995; 17(2): 109–114. doi:10.1080/07060669509500701.

19. Quilot-Turion B, Ould-Sidi M, Kadrani A, Hilgert N, Génard M, Lescourret F. Optimization of parameters of the 'Virtual Fruit' model to design peach genotype for sustainable production systems. Eur J Agron. 2012; 42: 34–48. doi:10.1016/j.eja.2011.11.008.

20. Tsoukalas I, Makropoulos C. Multiobjective optimisation on a budget: exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. Environ Model Softw. 2015; 69: 396–413. doi:10.1016/j.envsoft.2014.09.023.

21. Cools J, Broekx S, Vandenberghe V, Sels H, Meynaerts E, Vercaemst P, et al. Coupling a hydrological water quality model and an economic optimization model to set up a cost-effective emission reduction scenario for nitrogen. Environ Model Softw. 2011; 26(1): 44–51. doi:10.1016/j.envsoft.2010.04.017.

22. Holzkämper A, Klein T, Seppelt R, Fuhrer J. Assessing the propagation of uncertainties in multi-objective optimization for agro-ecosystem adaptation to climate change. Environ Model Softw. 2015; 66: 27–35. doi:10.1016/j.envsoft.2014.12.012.

23. Semenov MA, Stratonovitch P, Alghabari F, Gooding MJ. Adapting wheat in Europe for climate change. J Cereal Sci. 2014; 59(3): 245–256. doi:10.1016/j.jcs.2014.01.006.

24. Grechi I, Ould-Sidi M, Hilgert N, Senoussi R, Sauphanor B, Lescourret F. Designing integrated management scenarios using simulation-based and multi-objective optimization: application to the peach tree–Myzus persicae aphid system. Ecol Model. 2012; 246: 47–59. doi:10.1016/j.ecolmodel.2012.07.023.

25. Bergez JE, Colbach N, Crespo O, Garcia F, Jeuffroy MH, Justes E, et al. Designing crop management systems by simulation. Eur J Agron. 2010; 32(1): 3–9. doi:10.1016/j.eja.2009.06.001.

26. Ould-Sidi M, Lescourret F. Model-based design of integrated production systems: a review. Agron Sustain Dev. 2011; 31(3): 571. doi:10.1007/s13593-011-0002-8.

27. Picheny V, Trépos R, Casadebaig P. Optimization of black-box models with uncertain climatic inputs—Application to sunflower ideotype design. PLOS ONE. 2017; 12(5): e0176815. doi:10.1371/journal.pone.0176815.

28. Jalali H, Van Nieuwenhuyse I, Picheny V. Comparison of Kriging-based algorithms for simulation optimization with heterogeneous noise. Eur J Oper Res. 2017; 261(1): 279–301. doi:10.1016/j.ejor.2017.01.035.

29. Jalali H, Nieuwenhuyse I. Simulation optimization in inventory replenishment: a classification. IIE Trans. 2015; 47(11): 1217–1235. doi:10.1080/0740817X.2015.1019162.

30. Saif A, Elhedhli S. Cold supply chain design with environmental considerations: a simulation-optimization approach. Eur J Oper Res. 2016; 251(1): 274–287. doi:10.1016/j.ejor.2015.10.056.

31. Cambra M, Capote N, Myrta A, Llácer G. *Plum pox virus* and the estimated costs associated with sharka disease. EPPO Bull. 2006; 36(2): 202–204. doi:10.1111/j.1365-2338.2006.01027.x.

32. Rimbaud L, Dallot S, Gottwald T, Decroocq V, Jacquot E, Soubeyrand S, et al. Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annu Rev Phytopathol. 2015; 53: 357–378. doi:10.1146/annurev-phyto-080614-120140.

33. JORF. Arrêté du 17 mars 2011 relatif à la lutte contre le Plum Pox Virus, agent causal de la maladie de la Sharka, sur les végétaux sensibles du genre Prunus. JORF n°0067 du 20 mars 2011. République Française. Paris, France. NOR: AGRG1105295A. 2011; Available from: http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023728913&dateTexte=&categorieLien=id.

34. Rimbaud L, Dallot S, Bruchou C, Thoyer S, Jacquot E, Soubeyrand S, et al. Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. bioRxiv. 2018; 315747. doi:10.1101/315747.

35. Picard C, Soubeyrand S, Jacquot E, Thébaud G. Analyzing the influence of landscape aggregation on disease spread to improve management strategies. In revision.

36. Pleydell DRJ, Soubeyrand S, Dallot S, Labonne G, Chadœuf J, Jacquot E, et al. Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLOS Comput Biol. 2018; 14(4): e1006085. doi:10.1371/journal.pcbi.1006085.

37. Rimbaud L, Bruchou C, Dallot S, Pleydell DRJ, Jacquot E, Soubeyrand S, et al. Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. R Soc Open Sci. 2018; 5(1): 171435. doi:10.1098/rsos.171435.

38. Picard C, Rimbaud L, Hendrikx P, Soubeyrand S, Jacquot E, Thébaud G. PESO: a modelling framework to help improve management strategies for epidemics – application to sharka. EPPO Bull. 2017; 47(2): 231-236. doi:10.1111/epp.12375.

39. Rockafellar RT, Uryasev S. Optimization of conditional value-at-risk. Journal of risk. 2000; 2: 21-42. Available from: http://www.pacca.info/public/files/docs/public/finance/Active%20Risk%20Management/Uryasev%20Rockafellar-%20Optimization%20CVaR.pdf.

40. Labopin-Richard T, Gamboa F, Garivier A, Iooss B. Bregman superquantiles. Estimation methods and applications. Dependence Modeling. 2014; 4(1). doi:10.1515/demo-2016-0004.

41. Rockafellar RT, Royset JO, Miranda SI. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. Eur J Oper Res. 2014; 234(1): 140–154. doi:10.1016/j.ejor.2013.10.046.

42. Picheny V, Ginsbourger D. Noisy kriging-based optimization methods: a unified implementation within the DiceOptim package. Comput Stat Data Anal. 2014; 71: 1035–1053. doi:10.1016/j.csda.2013.03.018.

43. Forrester AIJ, Keane AJ, Bressloff NW. Design and analysis of “noisy” computer experiments. AIAA J. 2006; 44(10): 2331–2339. doi:10.2514/1.20068.

44. Fang K, Li R, Sudjianto A. Design and modeling for computer experiments. CRC Press; 2005.

45. Cressie N. Statistics for spatial data. John Wiley & Sons; 2015.

46. Richet Y, Deville Y, Chevalier C. DiceView: plot methods for computer experiments design and surrogate. 2012. Available from: http://cran.r-project.org/web/packages/DiceView.

47. Kleijnen JPC. Kriging metamodeling in simulation: a review. Eur J Oper Res. 2009; 192(3): 707–716. doi:10.1016/j.ejor.2007.10.013.

48. Picheny V, Ginsbourger D, Richet Y, Caplin G. Quantile-based optimization of noisy computer experiments with tunable precision. Technometrics. 2013; 55(1): 2–13. doi:10.1080/00401706.2012.707580.

49. Binois M, Gramacy RB, Ludkovski M. Practical heteroskedastic Gaussian process modeling for large simulation experiments. J. Comput. Graph. Statist. 2016; 1-41. doi: 10.1080/10618600.2018.1458625.

50. Qian PZG, Wu H, Wu CFJ. Gaussian process models for computer experiments With Qualitative and Quantitative Factors. Technometrics. 2008; 50(3): 383–396. doi:10.1198/004017008000000262.

51. Papaïx J, Adamczyk-Chauvat K, Bouvier A, Kiêu K, Touzeau S, Lannou C, et al. Pathogen population dynamics in agricultural landscapes: the Ddal modelling framework. Infect Genet Evol. 2014; 27: 509–520. doi:10.1016/j.meegid.2014.01.022.

52. Agreste. Inventaire des vergers 2013. 2013. Available from: http://agreste.agriculture.gouv.fr/enquetes/productions-vegetales-528/vergers-et-fruits/

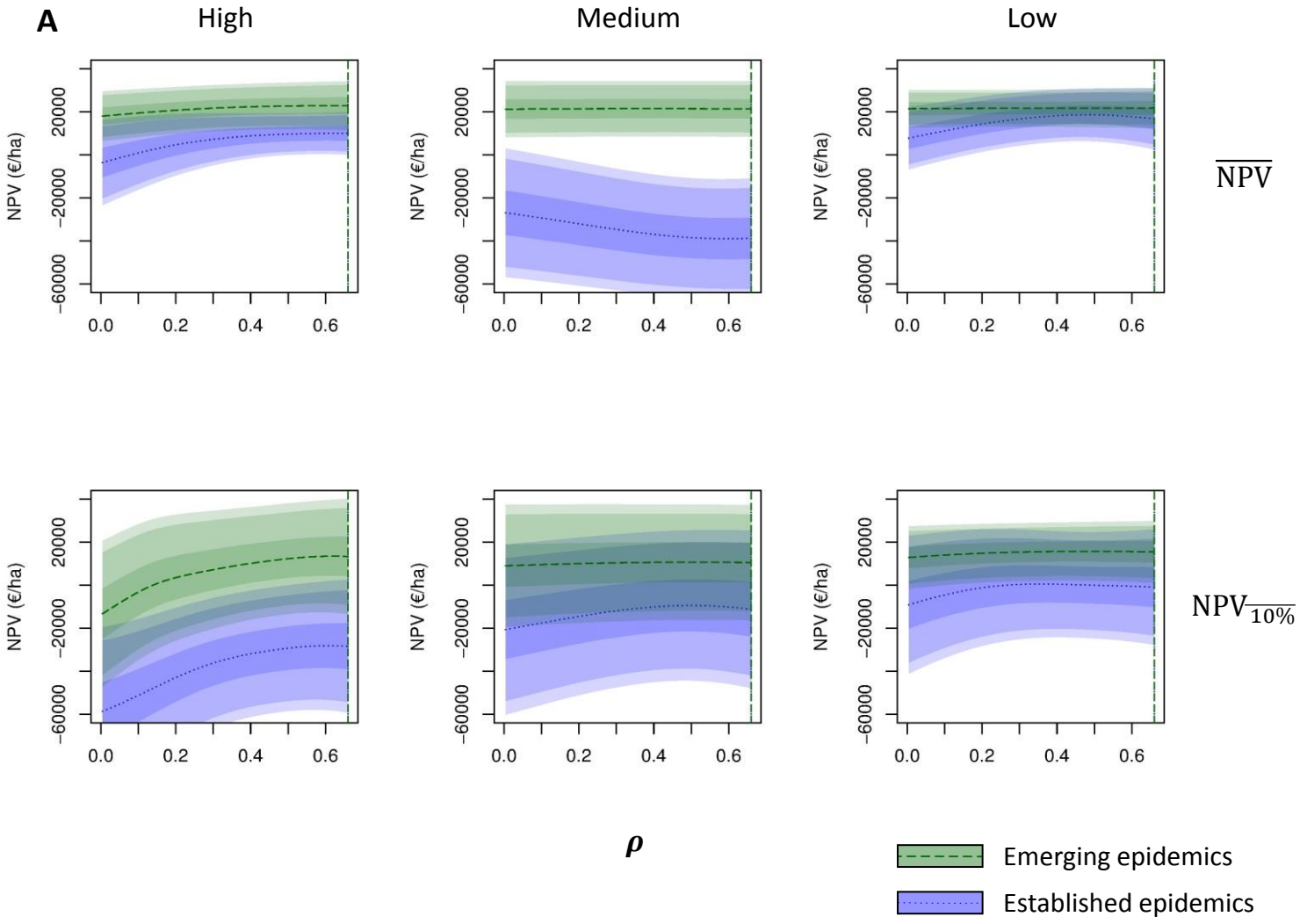
53. Filipe JAN, Cobb RC, Meentemeyer RK, Lee CA, Valachovic YS, Cook AR, et al. Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden Oak Death in Northern Californian forests. PLOS Comput Biol. 2012; 8(1): e1002328. doi:10.1371/journal.pcbi.1002328.

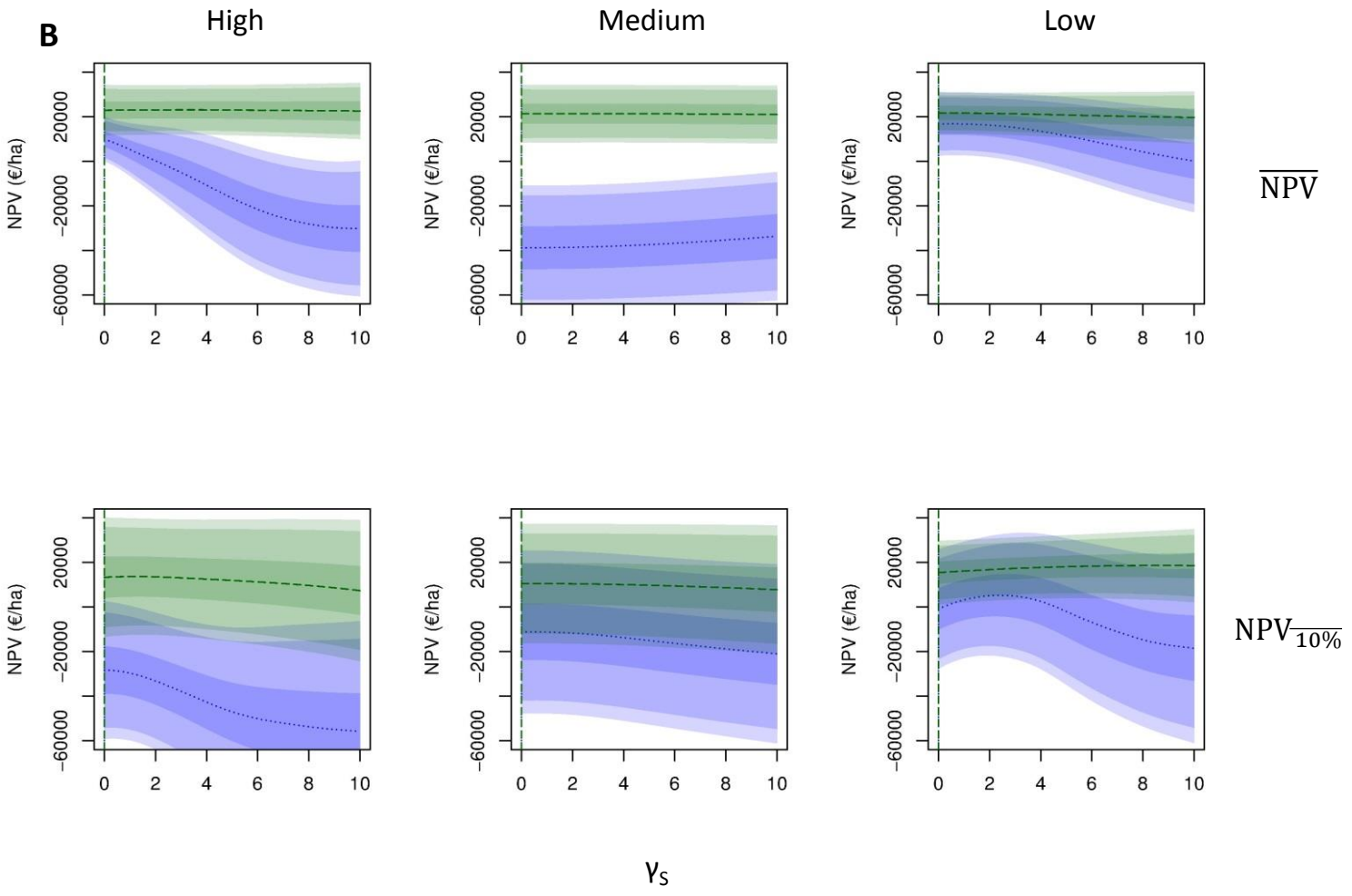
54. Parnell S, Gottwald TR, Gilks WR, van den Bosch F. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. J Theor Biol. 2012; 305: 30–36. doi:10.1016/j.jtbi.2012.03.009.

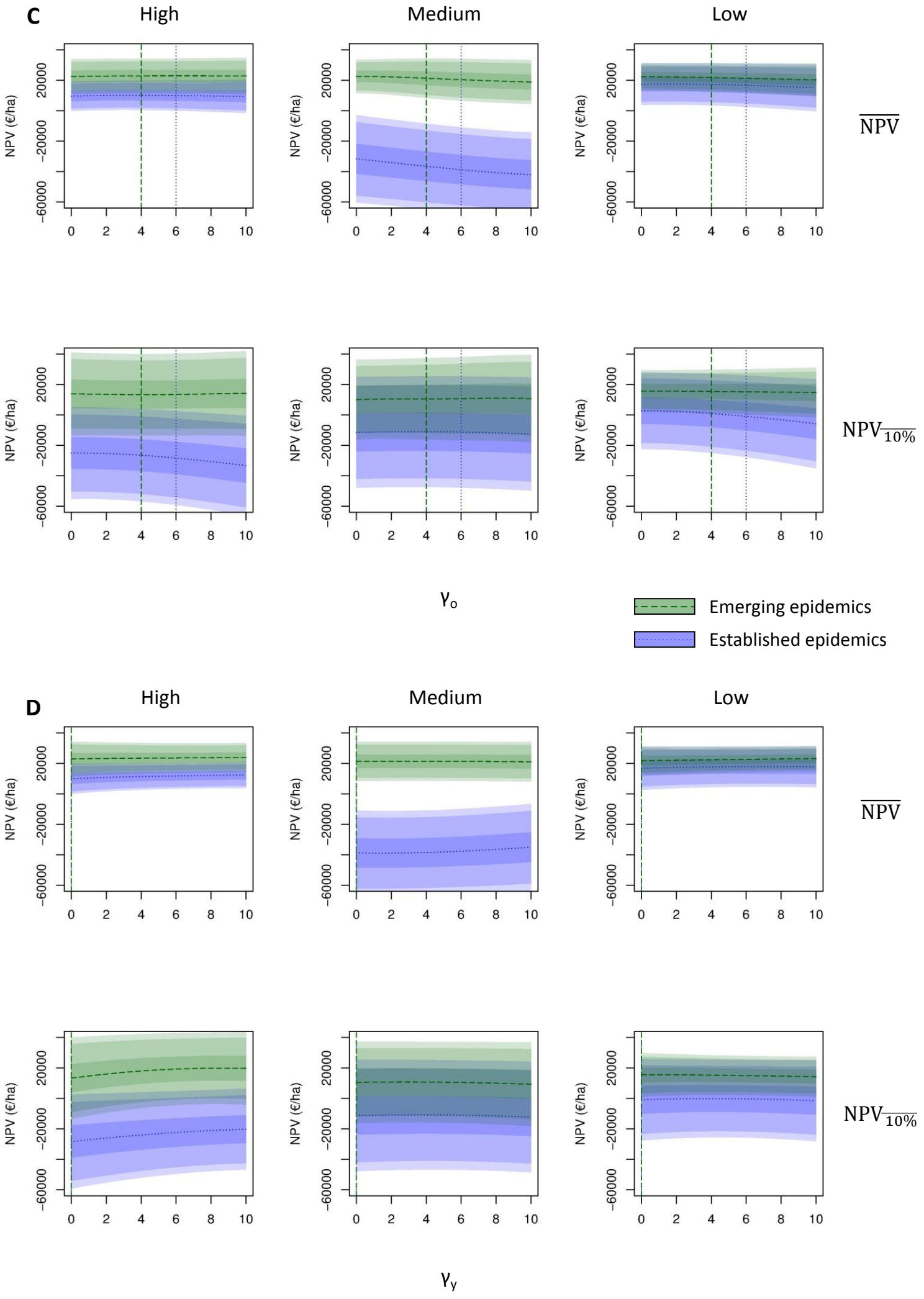
55. Chan M, Jeger MJ. An analytical model of plant virus disease dynamics with roguing and replanting. J Appl Ecol. 1994; 31: 413–427. doi:10.2307/2404439.

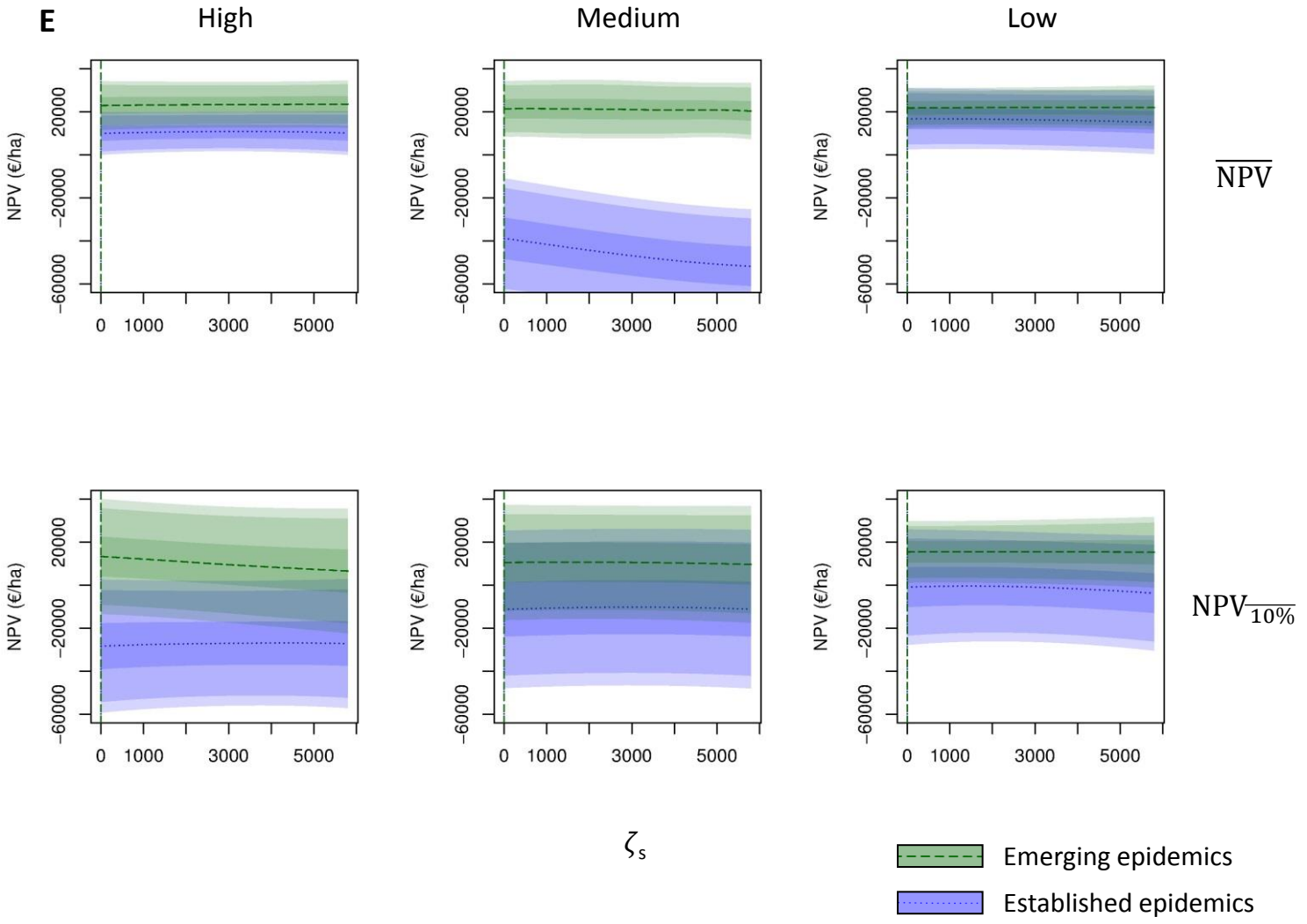
Supporting information

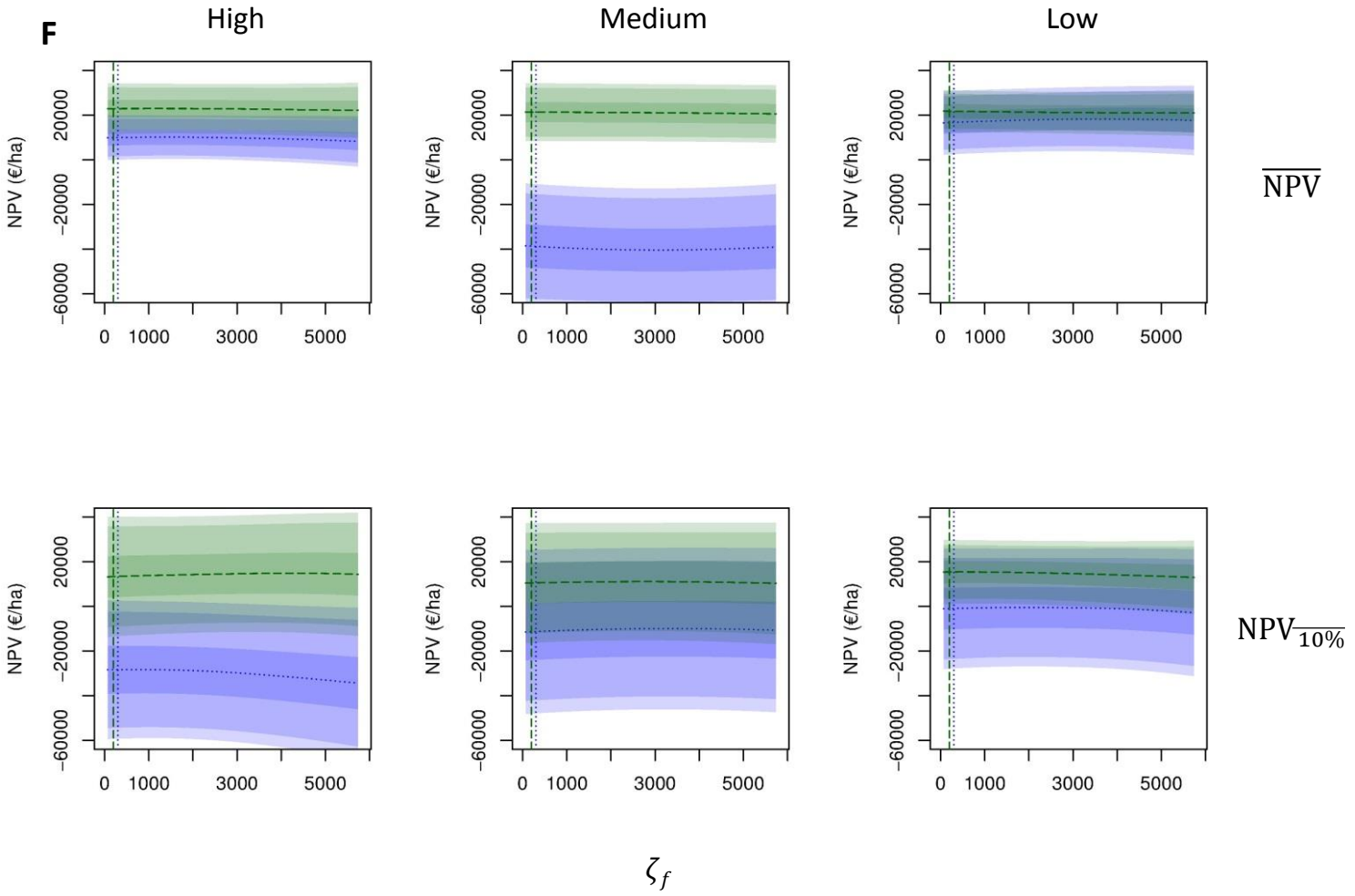
S1 Fig. 2D view of the kriging metamodels. These metamodels were obtained at the end of the optimization of $\overline{\mathbf{NPV}}$ (top row) and $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ (bottom row) for the 3 levels of landscape aggregation (from left to right) in case of emerging (green) and established (blue) epidemics. Each plot represents the influence of a single management parameter on the NPV, setting the values of the other management parameters to their optimal value. Here, the 21 management parameters are represented (A: $\boldsymbol{\rho}$, B: γ_S, C: γ_o, D: γ_y, E: ζ_s, F: ζ_f, G: ζ_{eO}, H: ζ_n, I: ζ_R, J: ζ_{eR}, K: $1/\eta 0$, L: η_s, M: η_f, N: η_{f*}, O: η_y, P: η_{y*}, Q: χ_o, R: χ_y, S: χ_{y*}, T: χ_n, U: χ_R). Dotted lines correspond to the optimized parameter values and different shades correspond to the 50th, 95th and 95th percentiles (representing the uncertainty of the kriging model). As an example, when the $\overline{\mathbf{NPV}}$ (or $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$) keeps the same value regardless of the management parameter value, this parameter have no influence on the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ (if the other parameters are setting to their optimal values).

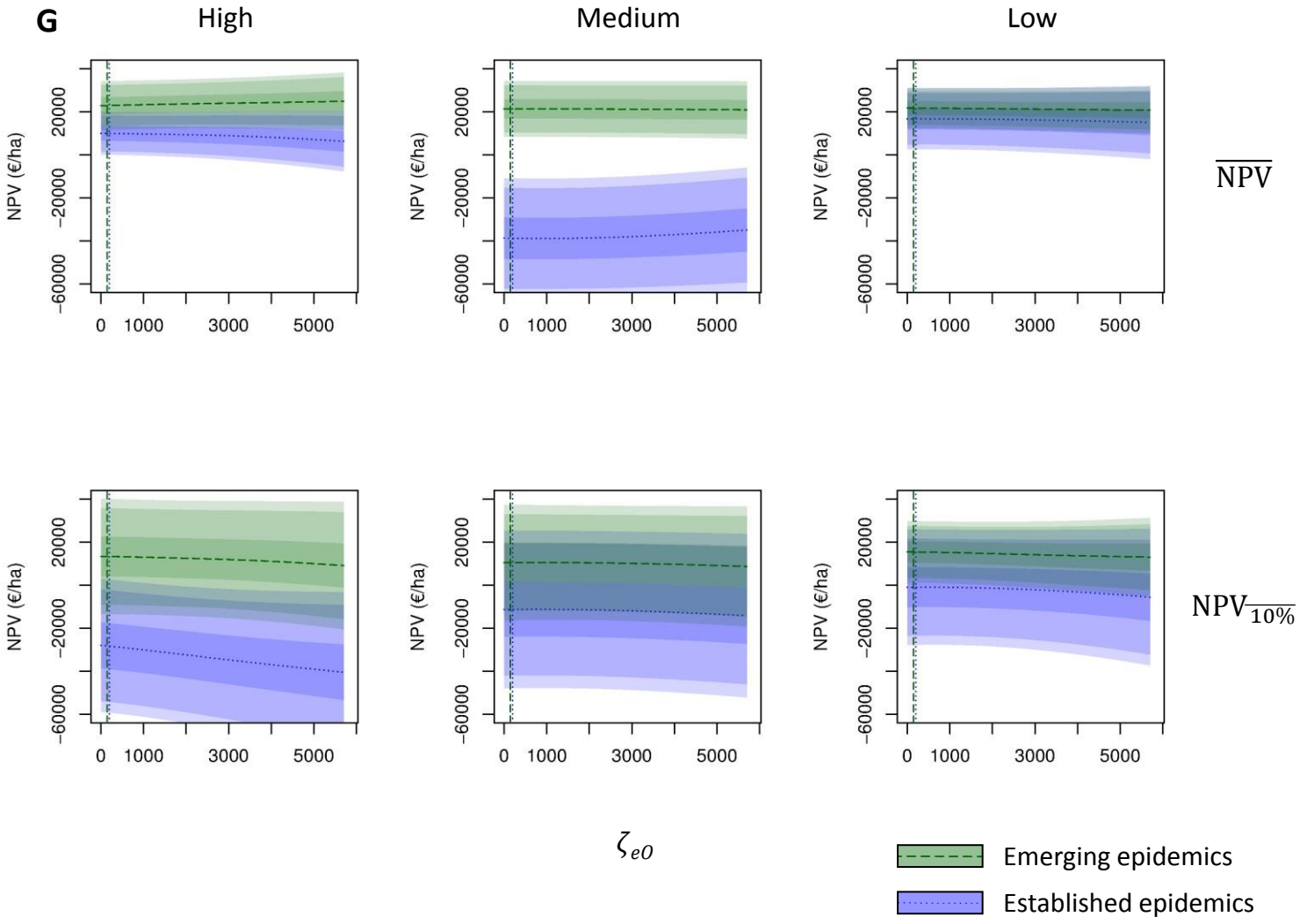

A
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.0
0.2
0.4
0.6
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ρ
Emerging epidemics
Established epidemics

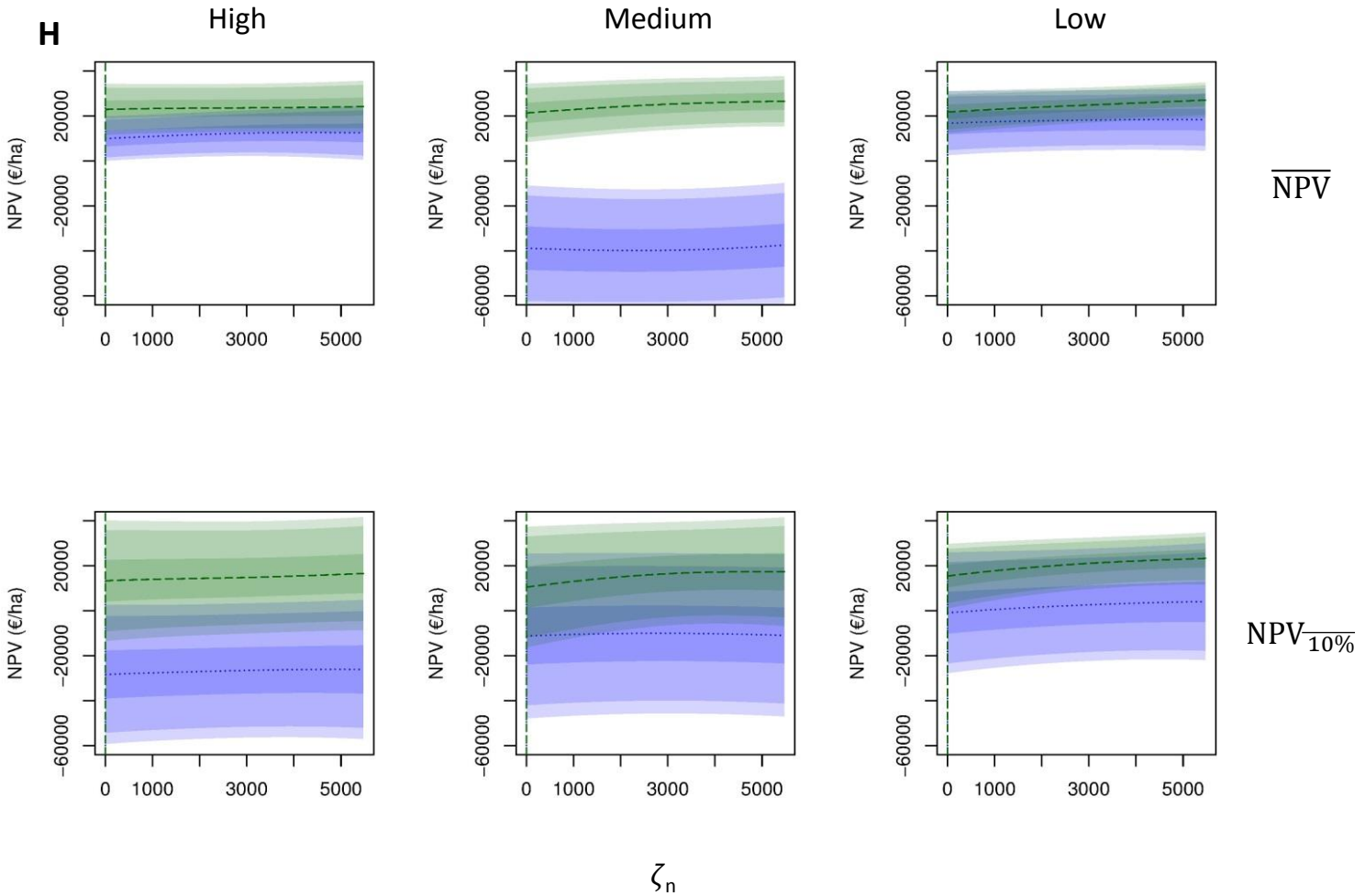

B
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
10
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
γ_S

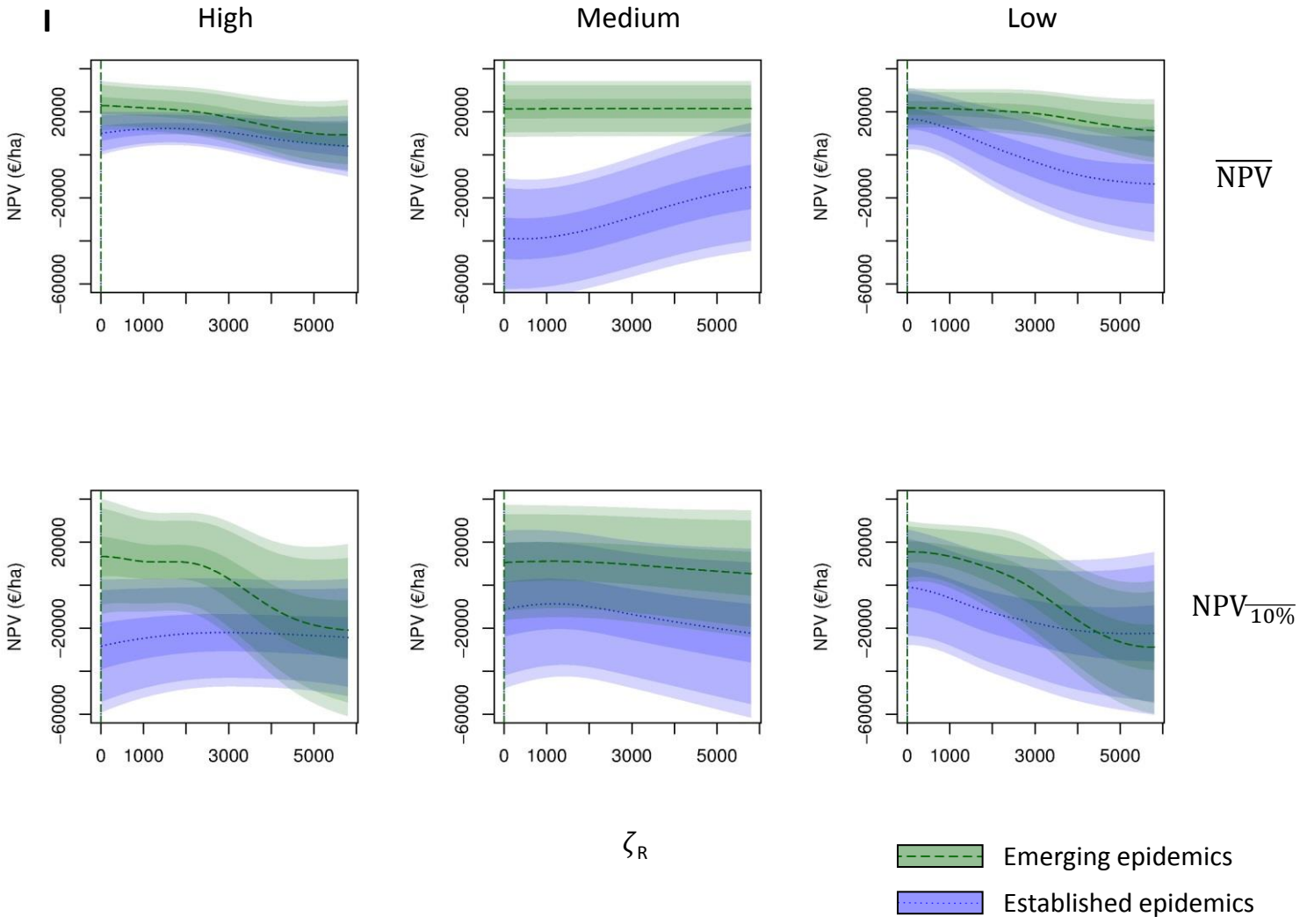

C
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
10
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
γ_o
Emerging epidemics
Established epidemics
D
High
Medium
Low
NPV (€/ha)
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
γ_y

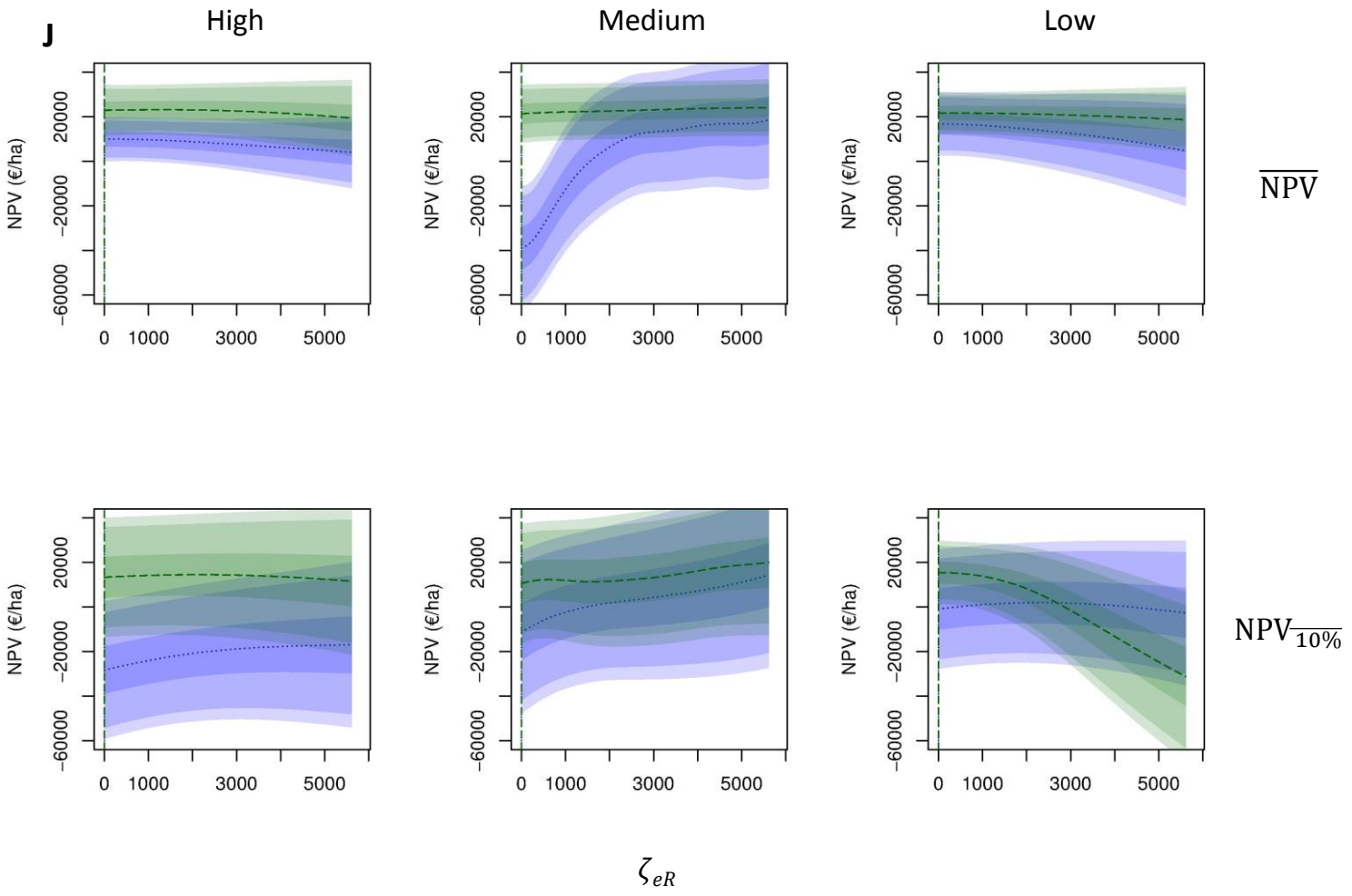

E
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_s
Emerging epidemics
Established epidemics

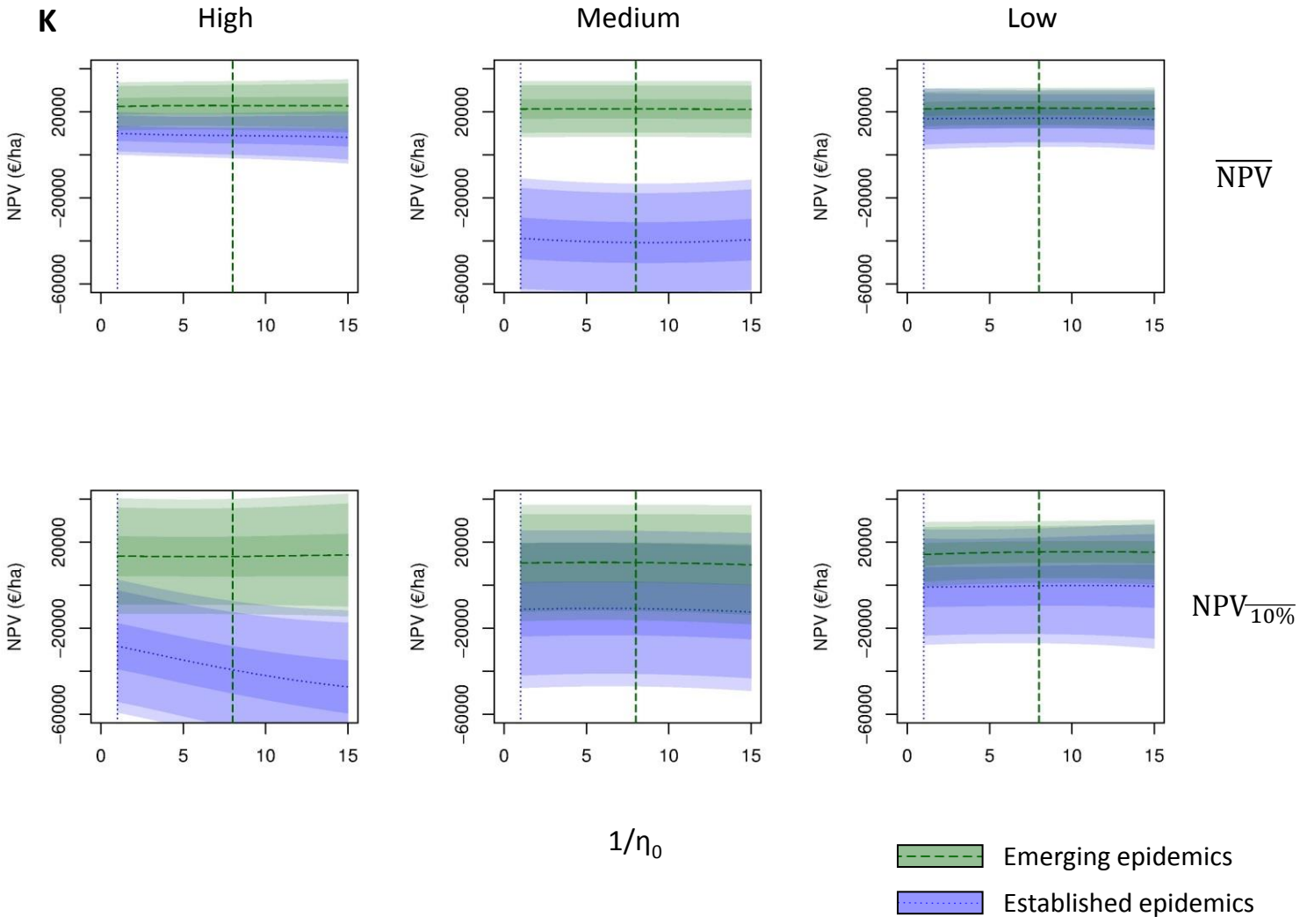

F
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_f

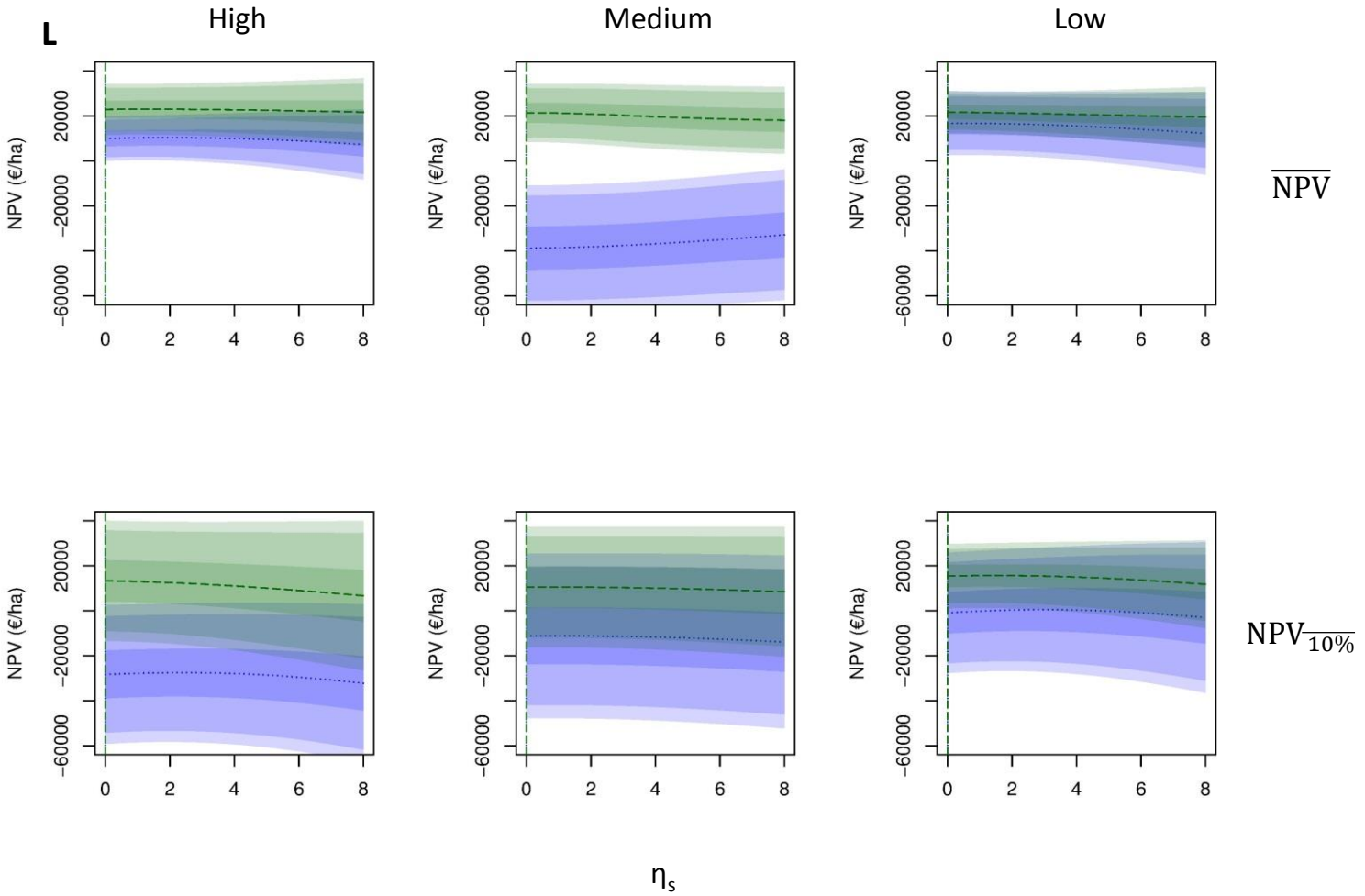

G
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_{e0}
Emerging epidemics
Established epidemics

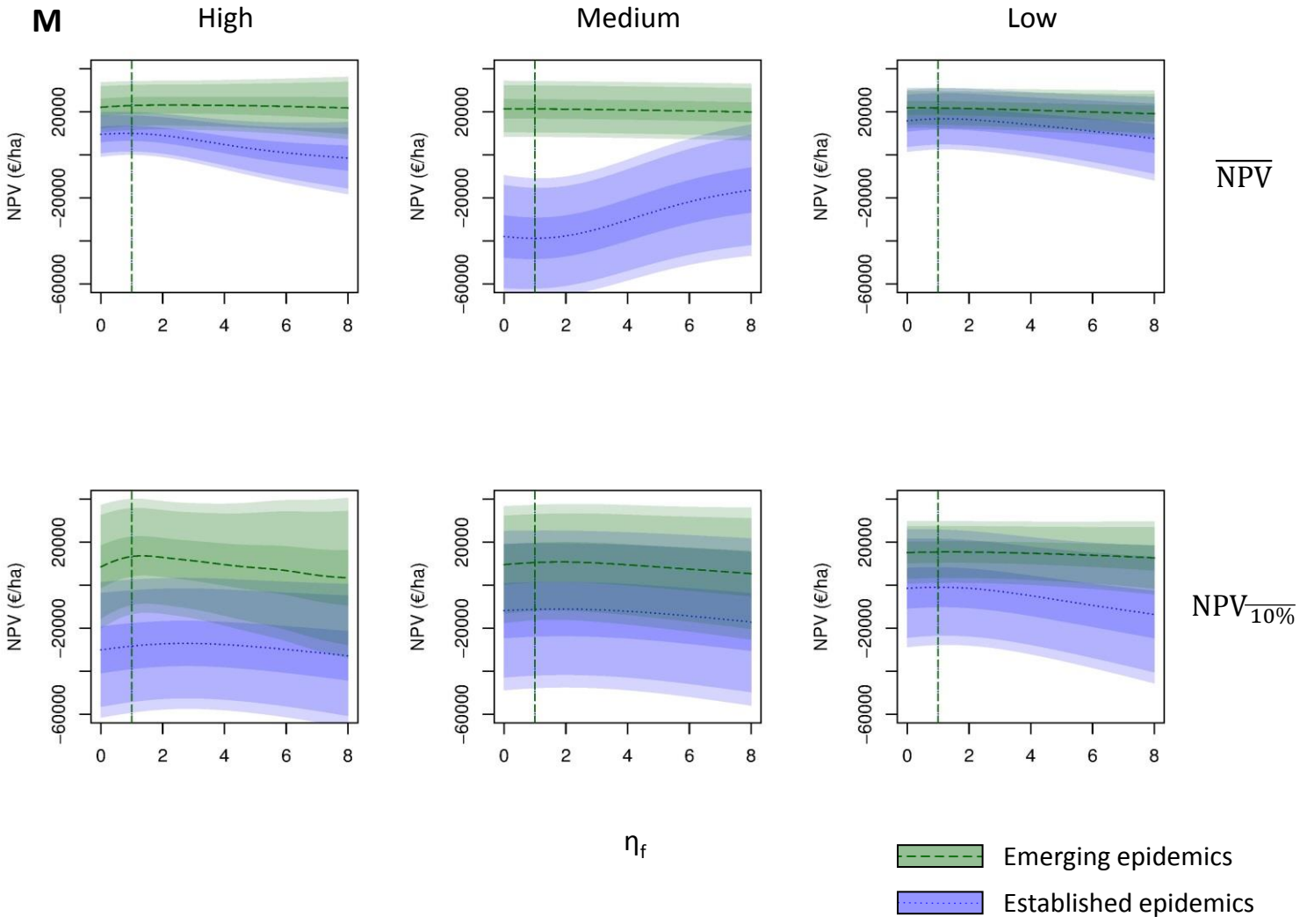

H
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_{n}

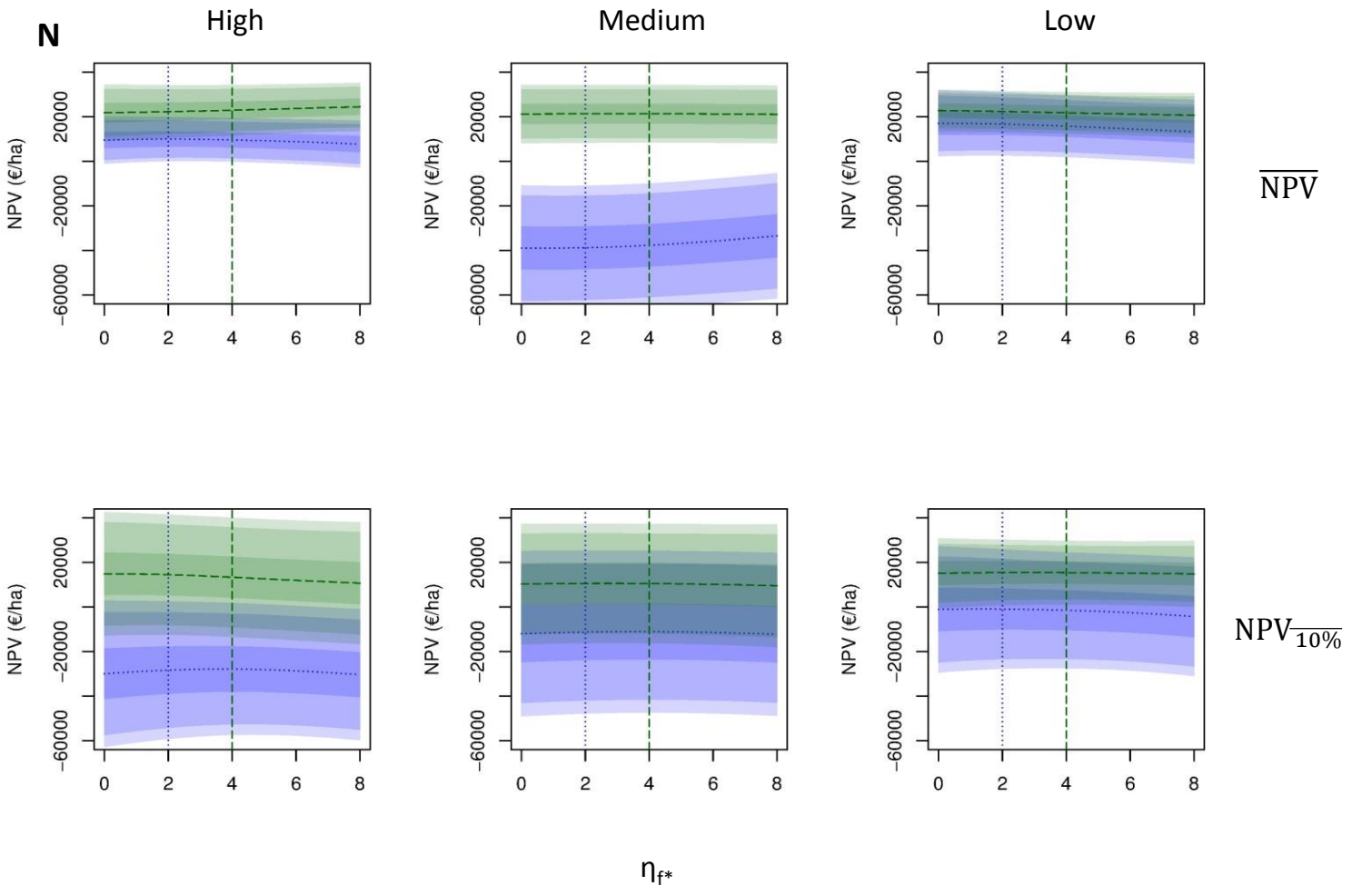

I
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_R
Emerging epidemics
Established epidemics

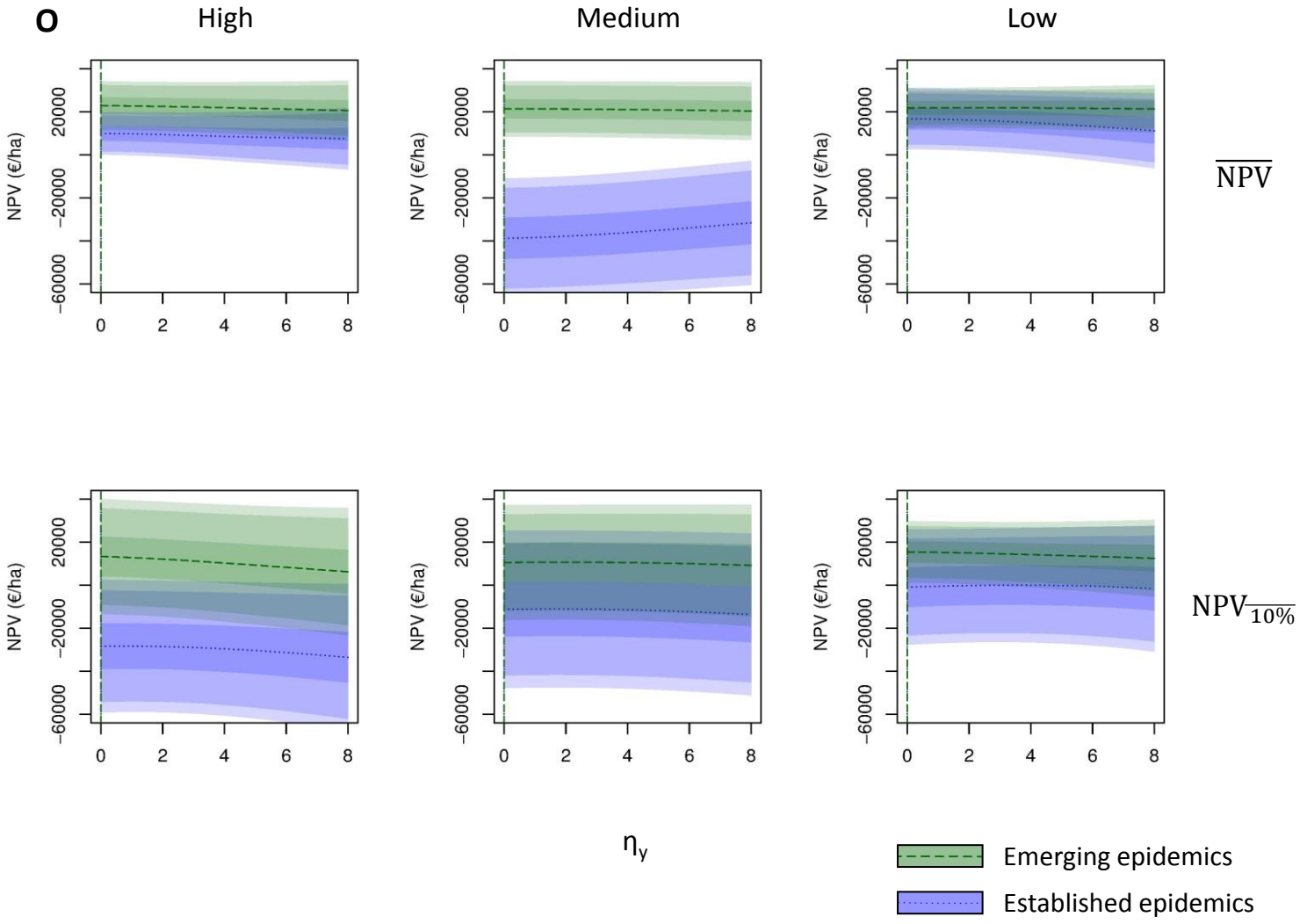

J
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_{eR}

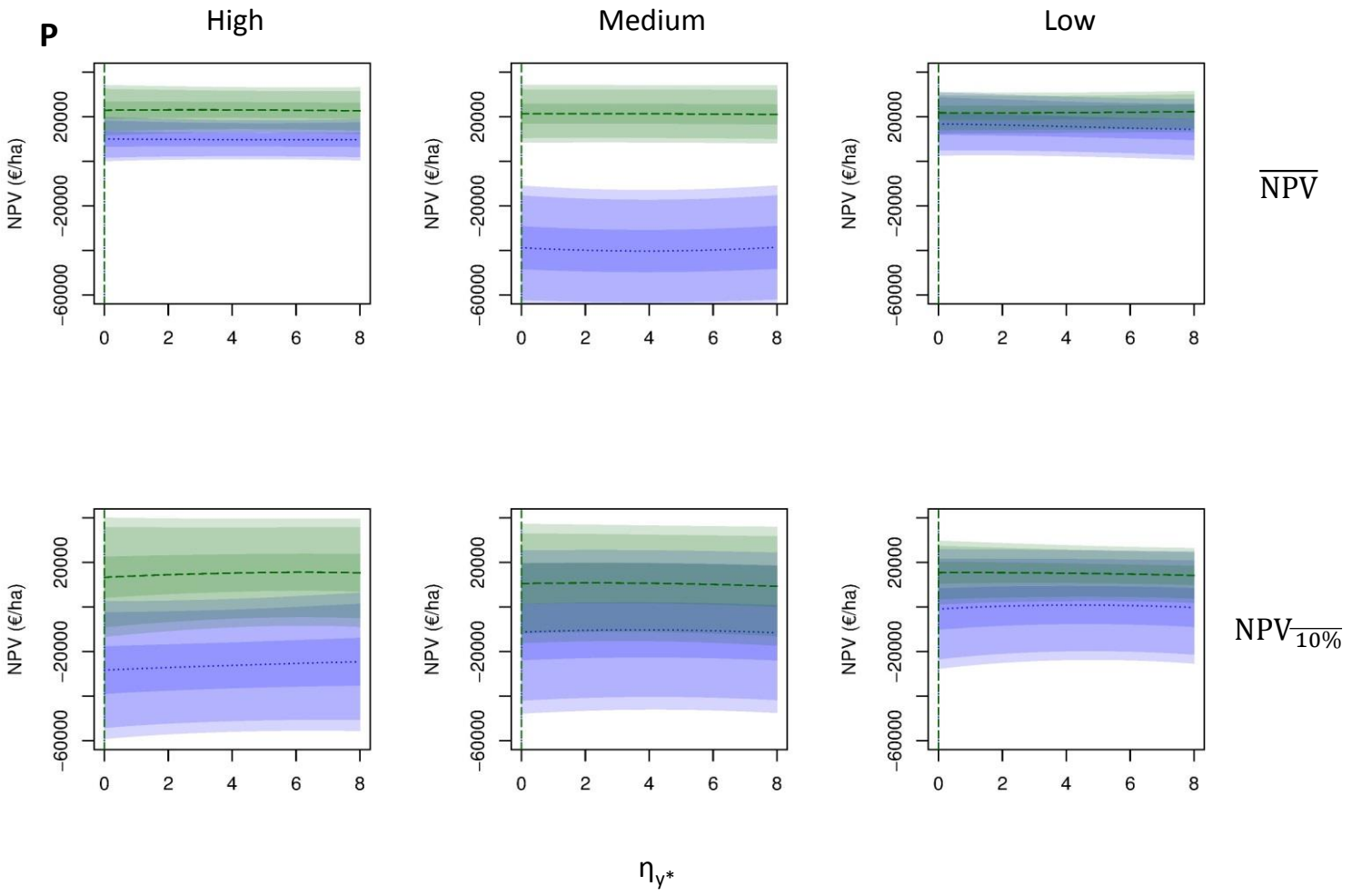

K
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
5
10
15
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
$1/\eta_0$
Emerging epidemics
Established epidemics

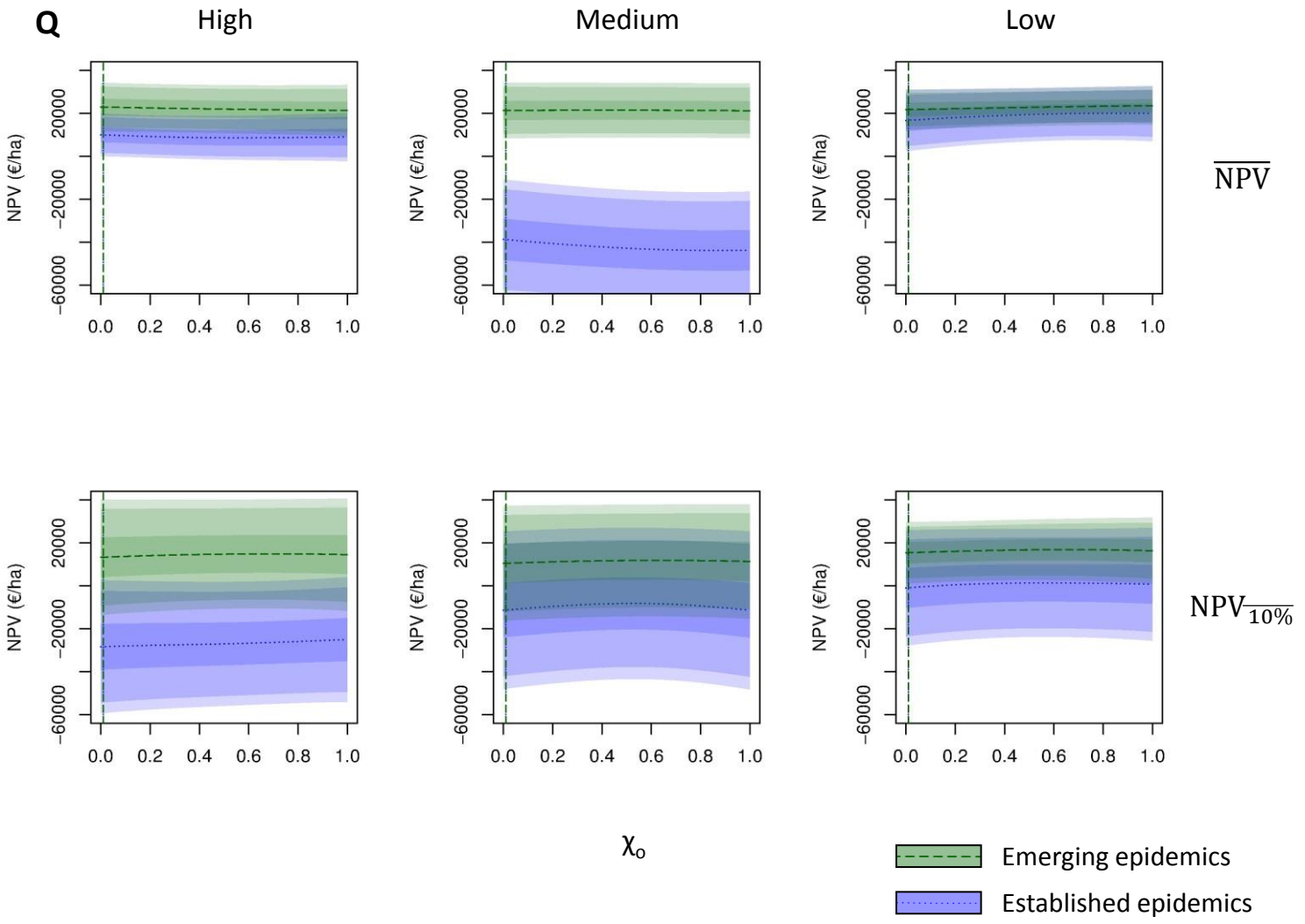

L
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
η_s

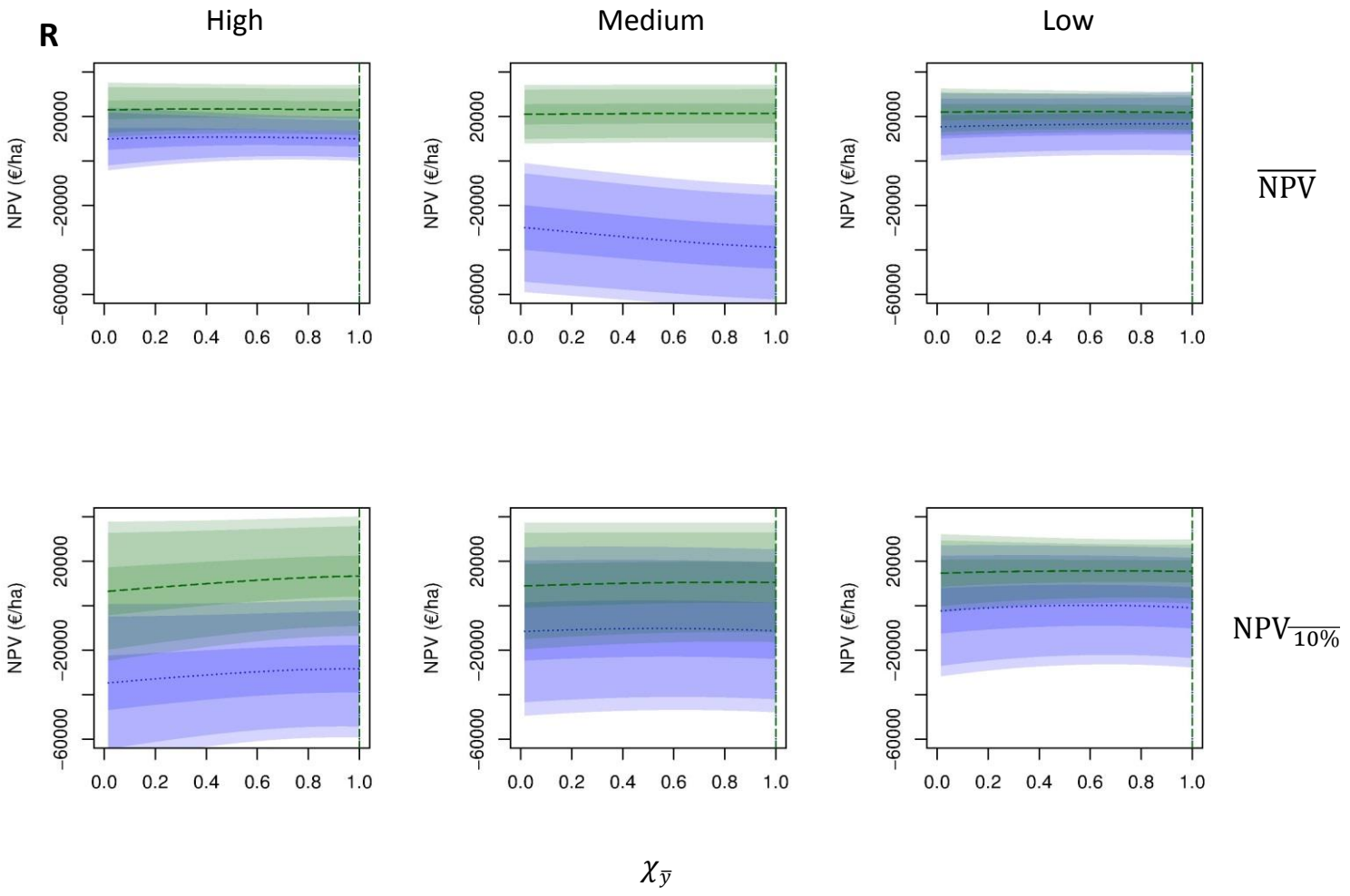

M
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
η_f
Emerging epidemics
Established epidemics

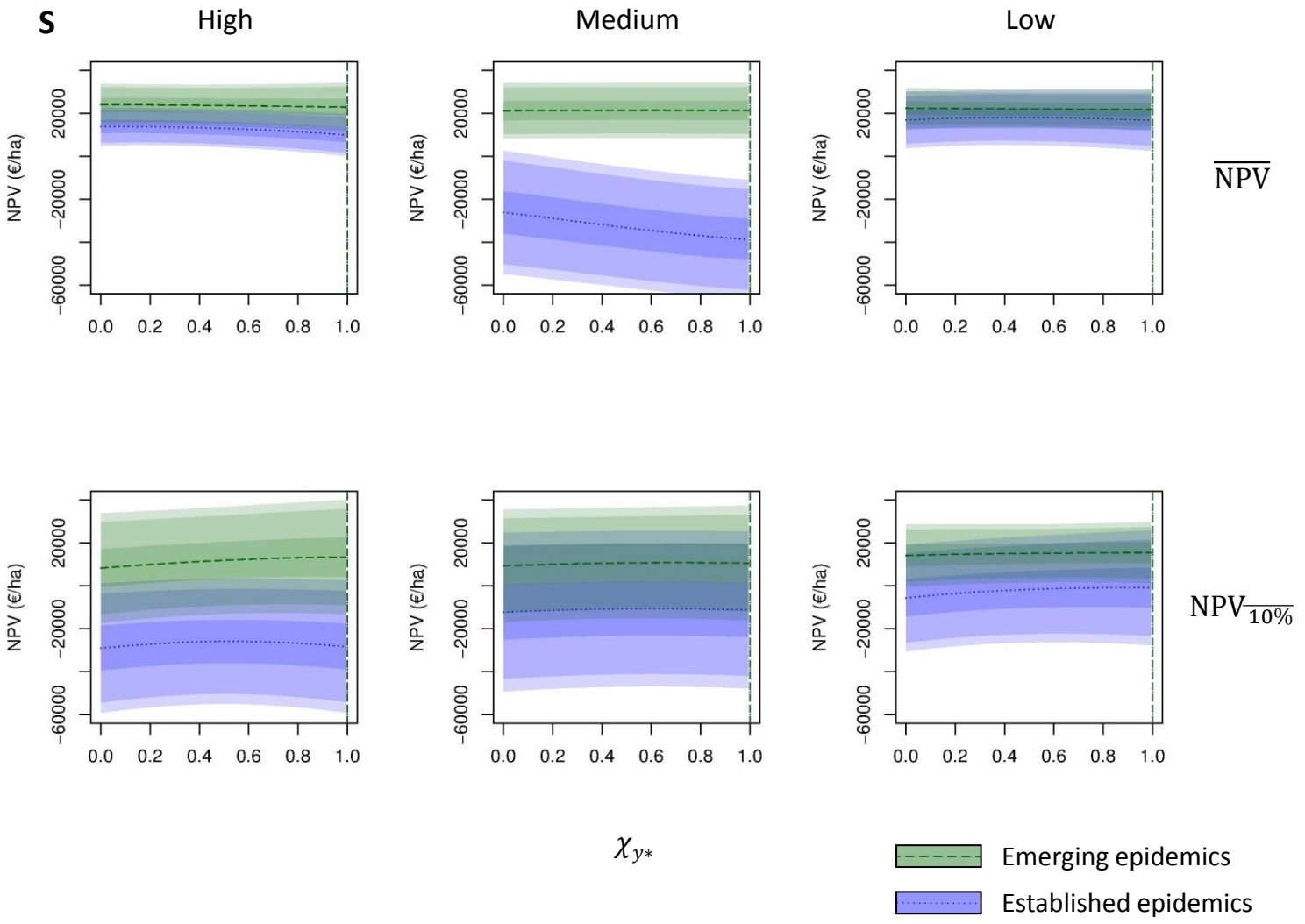

N
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
η_{f^*}

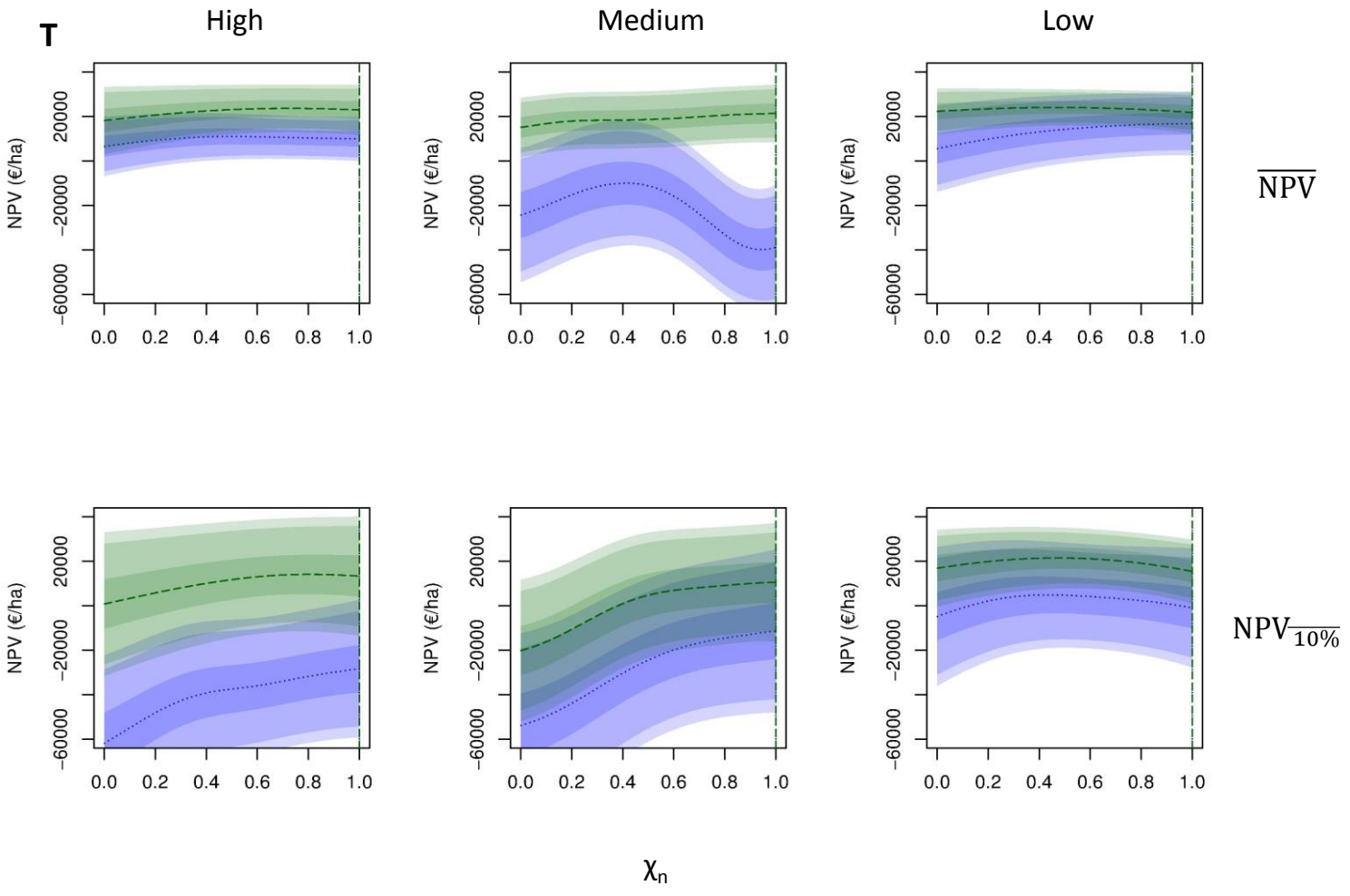

O
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
η_y
Emerging epidemics
Established epidemics

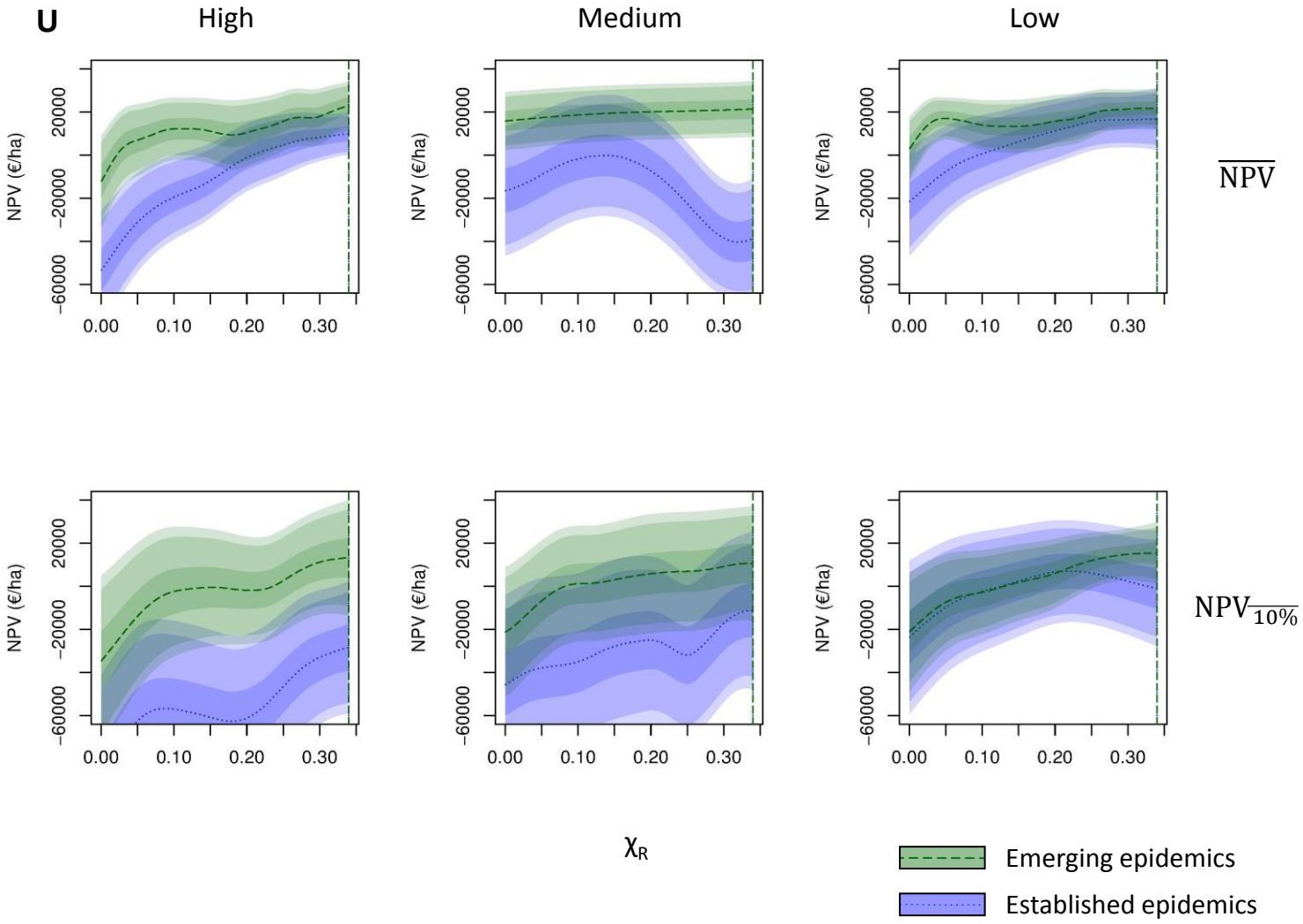

P
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0
2
4
6
8
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
η_{y^*}


Q
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.0
0.2
0.4
0.6
0.8
1.0
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
χ_o
Emerging epidemics
Established epidemics


R
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.0
0.2
0.4
0.6
0.8
1.0
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
$\chi_{\bar{y}}$


S
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.0
0.2
0.4
0.6
0.8
1.0
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
χ_{y*}
Emerging epidemics
Established epidemics


T
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.0
0.2
0.4
0.6
0.8
1.0
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
χ_n


U
High
Medium
Low
NPV (€/ha)
20000
−20000
−60000
0.00
0.10
0.20
0.30
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
χ_R
Emerging epidemics
Established epidemics

S1 Table. Management parameters implemented in the model. The values correspond to the French management strategy and to optimized strategies for emerging and established epidemics (after a simplification step).

Management parameters		French management	Emerging epidemics							Established epidemics						
			Optimization criterions						Strategy obtained by mixing some optimized strategies	Optimization criterions						Strategy obtained by mixing some optimized strategies
			$\overline{NPV}$			$NPV_{\overline{10\%}}$				$\overline{NPV}$			$NPV_{\overline{10\%}}$			
			H	M	L	H	M	L		H	M	L	H	M	L	
ρ	Probability of detection of a symptomatic tree	**0.66**	0.66	0.6	0.44	0.51	0.63	0.24	**0.66**	0.63	0.65	0.46	0.65	0.64	0.66	**0.66**
γ_S	Delay before replantation of a removed orchard (years)	**0**	3	2	0	0	2	0	**0**	0	0	1	0	0	0	**0**
γ_o	Duration of observation zones (years)	**3**	4	10	5	5	1	2	**4**	9	3	6	8	4	1	**6**
γ_y	Duration of young orchards (years)	**3**	0	0	0	3	0	1	**0**	4	2	0	9	3	5	**0**
ζ_s	Radius of security zones (m)	**2500**	0	0	0	0	0	0	**0**	4042	0	426	167	5008	1990	**0**
ζ_f	Radius of focal zones (m)	**1500**	196	137	242	186	67	304	**200**	1172	5230	188	120	1	252	**300**
ζ_{eO}	Radius of observation epicenter (m)	**564**	152	0	0	0	0	0	**150**	85	82	0	16	21	16	**200**
ζ_n	Radius of the close neighborhood (m)	**200**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**
ζ_R	Radius of the removal zone (m)	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**
ζ_{eR}	Radius of the removal epicenter (m)	**0**	0	0	0	0	0	0	**0**	0	0	0	0	0	0	**0**
$1/\eta_0$	Maximal period between 2 observations (year)	**6**	6	10	12	13	7	15	**8**	1	2	7	1	1	3	**1**
η_s	Observation frequency in security zones ($year^{-1}$)	**1**	0	0	0	0	0	0	**0**	1	0	1	3	1	1	**0**
η_f	Observation frequency in focal zones ($year^{-1}$)	**2**	1	1	1	2	2	3	**1**	1	0	2	3	1	2	**1**
η_{f*}	Modified observation frequency in focal zones ($year^{-1}$)	**3**	5	5	0	0	0	0	**5**	3	3	0	4	8	8	**3**
η_y	Observation frequency in young orchards ($year^{-1}$)	**2**	0	0	0	2	0	1	**0**	0	0	0	0	0	0	**0**
η_{y*}	Modified observation frequency in young orchards ($year^{-1}$)	**3**	0	1	0	0	0	0	**0**	7	8	0	8	2	7	**0**
χ_o	Contamination threshold in the observation epicenter, above which the observation frequency in focal zone is modified	**0.02**	0.01	1	1	1	1	1	**0.01**	0.01	0.3	1	0.57	0.4	0.52	**0.01**
χ_y	Contamination threshold in the environment around young orchards, above which the plantation of orchards is forbidden	**0.02**	1	1	1	1	1	1	**1**	1	1	1	1	1	1	**1**
χ_{y*}	Contamination threshold in the environment, above which the observation frequency in young orchards is modified	**0.01**	1	1	1	1	1	1	**1**	0.04	0.16	1	0.88	0.18	1	**1**
χ_n	Contamination threshold on an orchard in the neighborhood, above which the plantation of orchards is forbidden	**0.05**	1	1	1	1	1	1	**1**	1	1	1	1	1	1	**1**
χ_R	Contamination threshold in the removal epicenter, above which orchards inside the removal zone are removed	**0.1**	0.34	0.34	0.34	0.34	0.34	0.34	**0.34**	0.34	0.34	0.34	0.34	0.34	0.34	**0.34**

S2 Table. $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ obtained after 10.000 simulations of PPV dispersal and its management. Simulations were carried out for emerging and established epidemics on landscapes varying by their level of patch aggregation (high (H), medium (M) and low (L) aggregation), with strategies optimized for all the landscapes after 700 iterations, after 1000 iterations (local optimization) and after a simplification step. Optimized strategies identified by mixing optimized strategies are also presented. Values in bold represent the $\overline{\mathrm{NPV}}$ (or the $\mathrm{NPV}_{\overline{10\%}}$) corresponding to simulations performed on a landscape with the management optimized for the $\overline{\mathrm{NPV}}$ (or the $\mathrm{NPV}_{\overline{10\%}}$) for the same landscape (e.g. the $\overline{\mathrm{NPV}}$ of simulations performed on landscape H with management parameters optimized for the $\overline{\mathrm{NPV}}$ on the landscape H).

		Emerging epidemics						**Established epidemics**					
		H		**M**		**L**		**H**		**M**		**L**	
	Optimized strategies	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$
At iteration 700	$\overline{\mathrm{NPV}}$ for landscape H	**26297**	20744	27128	23222	27265	23462	**16996**	1028	21901	13668	22800	17304
	$\overline{\mathrm{NPV}}$ for landscape M	26448	20395	**27510**	23386	27663	23817	13774	-10884	**21832**	11403	23218	16725
	$\overline{\mathrm{NPV}}$ for landscape L	25827	18697	27272	22170	**27479**	22955	9090	-25841	21378	8194	**23207**	15385
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape H	25844	**21361**	26618	22909	26899	23402	15419	**-893**	20472	12013	21769	15881
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape M	26296	21107	27283	**23193**	27496	23612	14613	-8226	21532	**11904**	22641	16676
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape L	25993	20083	27232	23201	27463	**23648**	13265	-11807	21706	10946	23091	**16228**
At iteration 1000	$\overline{\mathrm{NPV}}$ for landscape H	**26846**	22388	27516	23692	27590	23871	**17074**	866	21511	13329	22459	16790
	$\overline{\mathrm{NPV}}$ for landscape M	26438	20393	**27499**	23192	27559	23437	9639	-22495	**22278**	11726	23603	17231
	$\overline{\mathrm{NPV}}$ for landscape L	25960	17670	27500	23122	**27700**	23826	13700	-9474	21662	9929	**23271**	15297
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape H	26233	**21783**	26952	23309	27154	23638	15634	**362**	20531	12631	21735	15877
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape M	26546	20732	27492	**23169**	27592	23681	15362	-6182	21886	**13227**	22827	17215
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape L	26077	19973	27442	23321	27650	**23841**	16137	-3416	22202	13006	23453	**17214**
Simplified strategies	$\overline{\mathrm{NPV}}$ for landscape H	**26889**	22580	27559	23906	27668	24087	**17295**	1020	21620	13681	22501	16982
	$\overline{\mathrm{NPV}}$ for landscape M	26770	21743	**27616**	23649	27686	23884	9602	-22148	**22476**	12782	23628	17436
	$\overline{\mathrm{NPV}}$ for landscape L	26432	19557	27630	23695	**27763**	24048	14272	-9470	21862	10905	**23482**	16348
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape H	26204	**21847**	26965	23142	27137	23615	15720	**297**	20471	12368	21751	15905
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape M	26737	21635	27577	**23485**	27665	23944	15404	-6624	21869	**13217**	22839	17397
	$\mathrm{NPV}_{\overline{10\%}}$ for landscape L	26323	20900	27480	23565	27737	**24116**	16853	-498	22323	13507	23454	**17435**
Strategy obtained by mixing optimized strategies		27045	22897	27655	24025	27778	24285	17455	1907	21974	14397	22881	17615

S3 table. $\overline{\mathbf{NPV}}$ and $\mathbf{NPV}_{\overline{10\%}}$ (€/ha) obtained after 10,000 simulations of PPV dispersal and its management. Simulations are carried out for emerging and established epidemics on landscapes varying by their level of patch aggregation (high (H), medium (M) and low aggregation (L)), with various disease management strategies (the strategies optimized for emerging and established epidemics correspond to the strategies obtained by mixing optimized strategies). In addition, simulations were performed for different variation ranges of the transmission coefficient (β).

Transmission coefficient (β)	Management strategy	Emerging epidemics						Established epidemics					
		H		M		L		H		M		L	
		$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$
[1.25;1.39]	**French management strategy**	22073	11698	24473	18950	25240	21213	973	-24587	15997	1369	19865	10866
	Without disease without management	28562	25351	28593	25335	28607	25298	27639	24390	27624	24392	27666	24419
	Without management	18814	-12684	26372	15788	27049	20134	-38235	-108165	8763	-32919	16236	-9782
	Strategy optimized for emerging epidemics	**27045**	**22897**	**27655**	**24025**	**27778**	**24285**	16725	-1385	**22661**	12365	**23978**	16785
	Strategy optimized for of established epidemics	25451	21940	25574	22188	25562	22186	**17455**	**1907**	21974	**14397**	22881	**17615**
	Optimized strategy from (Picard et al., 2018)	26414	21139	27164	23281	27164	23567	8706	-28952	20015	2812	21681	10373
[2.50;2.78]	**French management strategy**	21239	8537	24347	18083	25221	20887	-6082	-35802	12384	-5841	17305	4545
	Without management	692	-57030	22194	-5728	23875	5467	-91794	-172227	-15276	-89433	-4129	-59270
	Strategy optimized for emerging epidemics	**26638**	**21925**	**27503**	**23665**	**27611**	**23894**	9385	-16934	20199	4648	21739	10363
	Strategy optimized for established epidemics	25330	21556	25535	22107	25504	22210	**11614**	**-9985**	**20448**	**9203**	**21823**	**14562**
[3.75;4.17]	**French management strategy**	21244	8652	24345	18196	25195	20909	-6107	-35951	12302	-6412	17378	4825
	Without management	424	-56987	22133	-5919	23800	5380	-92355	-173160	-15479	-90863	-4041	-58542
	Strategy optimized for emerging epidemics	**26671**	**21847**	**27505**	**23681**	**27612**	**23909**	9140	-17979	20163	4765	**21782**	10659
	Strategy optimized for established epidemics	25315	21332	25559	22144	25546	22201	**11543**	**-10062**	**20403**	**9081**	21746	**14360**

Initialization of orchard ages and durations in the simulation model

We present here how orchard ages and durations are initialized in the simulation model used in this article. The general idea is to sample orchard age (a discrete number of years, identical for all trees in the orchard) from the stable age distribution (i.e. at the steady state, for a standard turnover of the orchards).

1 Notations

Each orchard is associated with a single patch z. The orchard is at age 0 during the time step following its plantation, and will live until the end of time step r (i.e. during r+1 time steps). The Boolean variable $S_{z,t}$ defines the state of patch z at time t: if it is occupied by an orchard, $S_{z,t}$=1; otherwise, $S_{z,t}$=0. This orchard is at age $A_{z,t}$ and its total lifespan (past and future) is $X_{z,t}$.

2 Lifespan Distribution

It is assumed that the total lifespan at birth (i.e. plantation) for a given orchard follows a Poisson distribution with parameter λ. Therefore, the total lifespan of an orchard at age $A_{z,t} = i$ present on site z at time t follows a left-truncated (up to i-1) Poisson distribution with parameter λ.
Thus, $\forall i \in \mathbb{N}$ and $\forall r \geq i$,

$$P(X_{z,t} = r \mid A_{z,t} = i) = \frac{e^{-\lambda}\lambda^r}{r!} \Big/ \sum_{k=i}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \tag{1}$$

and in particular:

$$P(X_{z,t} = r \mid A_{z,t} = 0) = \frac{e^{-\lambda}\lambda^r}{r!}. \tag{2}$$

3 Age Distribution

The stable age distribution is defined as the probability that the orchard sampled on site z at time t has age i, which obviously only concerns the patches occupied at time t. For these patches, this corresponds to the probability that the orchard sampled on site z at time t was planted at time t-i. Thus,

$$P(A_{z,t} = i) = P(A_{z,t} = i \mid S_{z,t} = 1) = P(A_{z,t-i} = 0 \mid S_{z,t} = 1). \tag{3}$$

$\{A_{z,t-i} = 0\}_{i\in\mathbb{N}}$ being a partition of the sample space, Bayes' theorem gives:

$$P(A_{z,t-i} = 0 \mid S_{z,t} = 1) = \frac{P(A_{z,t-i} = 0).P(S_{z,t} = 1 \mid A_{z,t-i} = 0)}{\sum_{j=0}^{\infty}\left[P(A_{z,t-j} = 0).P(S_{z,t} = 1 \mid A_{z,t-j} = 0)\right]}. \tag{4}$$

Now note that $P(S_{z,t} = 1 \mid A_{z,t-k} = 0) = P(X_{z,t-k} \geq k \mid A_{z,t-k} = 0)$. In addition, at the steady state, $\forall k \in \mathbb{N}$, $P(A_{z,t-k} = 0)$ is a constant. Therefore,

$$P(A_{z,t} = i) = \frac{P(X_{z,t-i} \geq i \mid A_{z,t-i} = 0)}{\sum_{j=0}^{\infty} P(X_{z,t-j} \geq j \mid A_{z,t-j} = 0)} = \frac{\sum_{k=i}^{\infty} P(X_{z,t-i} = k \mid A_{z,t-i} = 0)}{\sum_{j=0}^{\infty}\sum_{k=j}^{\infty} P(X_{z,t-j} = k \mid A_{z,t-j} = 0)}. \tag{5}$$

The numerator equals $\sum_{k=i}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}$ (see eq. 2). Similarly, the denominator simplifies into:

$$\sum_{k=0}^{\infty}\sum_{j=0}^{k} \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{\infty}\left[(k+1)\frac{e^{-\lambda}\lambda^k}{k!}\right] = \left(\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\right) + \sum_{k=0}^{\infty}\frac{k.e^{-\lambda}\lambda^k}{k!} = 1 + \lambda\sum_{l=0}^{\infty}\frac{e^{-\lambda}\lambda^l}{l!} = \lambda + 1. \tag{6}$$

The stable age distribution is therefore defined by:

$$P(A_{z,t} = i) = \frac{1}{\lambda+1}\sum_{k=i}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}. \tag{7}$$

At the steady state, the expected age of the orchards is:

$$\begin{aligned}\mathbb{E}(A_{z,t}) &= \sum_{i=0}^{\infty}[i.P(A_{z,t}=i)] = \frac{1}{\lambda+1}\sum_{i=0}^{\infty}\sum_{k=i}^{\infty}\left(i\frac{e^{-\lambda}\lambda^k}{k!}\right) = \frac{1}{\lambda+1}\sum_{k=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^k}{k!}\sum_{i=0}^{k}i\right) = \frac{1}{\lambda+1}\sum_{k=0}^{\infty}\left[\frac{k(k+1)}{2}.\frac{e^{-\lambda}\lambda^k}{k!}\right] \\ &= \frac{\lambda}{2(\lambda+1)}\sum_{k=1}^{\infty}\frac{(k+1)e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \frac{\lambda}{2(\lambda+1)}\sum_{l=0}^{\infty}(l+2)\frac{e^{-\lambda}\lambda^l}{l!} = \frac{\lambda}{2(\lambda+1)}\left[2\left(\sum_{l=0}^{\infty}\frac{e^{-\lambda}\lambda^l}{l!}\right)+\sum_{l=0}^{\infty}\frac{l.e^{-\lambda}\lambda^l}{l!}\right].\end{aligned}$$

This last term being equal to λ (see eq. 6), we obtain:

$$\mathbb{E}(A_{z,t}) = \frac{\lambda(\lambda+2)}{2(\lambda+1)}. \tag{8}$$

At the stationary state, the variance of orchard age is:
$Var(A_{z,t}) = \mathbb{E}(A_{z,t}^2) - \mathbb{E}^2(A_{z,t}) = \mathbb{E}(A_{z,t}^2 - A_{z,t}) + \mathbb{E}(A_{z,t}) - \mathbb{E}^2(A_{z,t})$.

And, according to (eq. 8), $\mathbb{E}(A_{z,t}) - \mathbb{E}^2(A_{z,t}) = \frac{\lambda(\lambda+2)(2-\lambda^2)}{4(\lambda+1)^2}$.

$$\begin{aligned}\text{Furthermore, } \mathbb{E}(A_{z,t}^2 - A_{z,t}) &= \sum_{i=0}^{\infty}\left(\frac{i^2-i}{\lambda+1}\sum_{k=i}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\right) = \frac{1}{\lambda+1}\sum_{k=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^k}{k!}\left[\left(\sum_{i=0}^{k}i^2\right)-\sum_{i=0}^{k}i\right]\right) \\ &= \frac{1}{3(\lambda+1)}\sum_{k=2}^{\infty}\left[(k-1)k(k+1)\frac{e^{-\lambda}\lambda^k}{k!}\right] = \frac{\lambda^2}{3(\lambda+1)}\sum_{l=0}^{\infty}\left[(l+3)\frac{e^{-\lambda}\lambda^l}{l!}\right].\end{aligned}$$

This last sum being equal to λ+3 (see eq. 6), we finally get:

$$Var(A_{z,t}) = \frac{\lambda^2(\lambda+3)}{3(\lambda+1)} + \frac{\lambda(\lambda+2)(2-\lambda^2)}{4(\lambda+1)^2}. \tag{9}$$

4 Implementation

To initialize the simulation model, we randomly sample for each orchard:
- an age in the stable age distribution: $P(A_{z,t}=i) = \frac{1}{\lambda+1}\sum_{k=i}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}$;
- a lifespan in the stationary lifespan distribution conditional on the previously sampled age:
$P(X_{z,t}=r \mid A_{z,t}=i) = \frac{e^{-\lambda}\lambda^r}{r!}\Big/\sum_{k=i}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}$.

Simulations were performed for lambda=15, as in previous work (Rimbaud et al., 2018).

S2 Text. Details of NPV calculation.

In the model, for year a, the gross margin (GM_a) generated by a set of orchards (i in {1, … , I}) in a landscape is calculated as the benefit engendered by fruit sales, minus all costs due to *Prunus* cultivation and management actions (Rimbaud et al., 2018):

$$\boldsymbol{GM_a = \sum_{i=1}^{I} \left(y_{i,a}.(p - c_h).\frac{S_{i,a}+E_{i,a}}{N_{i,a}}.A_i - c_F.A_i - \mathbb{I}_{i,a}^{R}.c_R.A_i - \mathbb{I}_{i,a}^{S}.c_S.A_i - c_R^T.R_{i,a}^{+} - c_A.O_{i,a} - c_o.O_{i,a}.A_i \right),}$$

with the following parameters:

Economic parameters		**Reference value (Rimbaud et al., 2018)**
A_i	Orchard area (ha)	
$\frac{S_{i,a}+E_{i,a}}{N_{i,a}}$	Proportion of uninfected trees in the orchard	
$O_{i,a}$	Number of observations in year a	
$R_{i,a}^{+}$	Number of newly (individually) removed trees due to PPV detection	
$\mathbb{I}_{i,a}^{R}$	Boolean which equals 1 if the orchard is removed, and 0 otherwise	
$\mathbb{I}_{i,a}^{S}$	Boolean which equals 1 if the orchard is planted, and 0 otherwise	
$y_{i,a}$	Relative age-dependent yield of trees in S or E states	0.00 until 2 years 0.50 at 3 years 0.65 at 4 years 0.85 at 5 years 1.00 from 6 to 15 years 0.80 from 16 years
c_S	Planting cost for one orchard (€.ha^{-1})	14,000
c_R	Removal cost for one orchard (€.ha^{-1})	1,000
c_R^T	Removal cost for one individual tree (€)	15
c_F	Yearly fixed cost associated with *Prunus* cultivation (€.ha^{-1})	13,600
c_O	Cost of one observation (€.ha^{-1})	160 (with ρ=0.66) *
c_A	Cost of the access to an orchard to survey (€.ha^{-1})	40
c_h	Cost of harvest	
p	Maximal yearly benefit generated by fruit harvest (€.ha^{-1})	37,250
τ_a	Discount rate	0.04

* The cost of one orchard observation is described by a simple linear function of the detection probability: $c_O = 182 \times \rho$, to account for the effect of partial observation of orchards (e.g. surveillance of every other row only), which reduces both the probability of detection and the cost of observation.

Using a discount rate τ_a=4%, the net present value (NPV) of the landscape between years a_m and a_f is:

$$NPV = \sum_{a=a_m}^{a_f} \frac{GM_a}{(1+\tau_a)^{(a-a_m)}}.$$

Résultats clés de l'Article 6

OPTIMISATION DE LA GESTION DES MALADIES DES PLANTES A L'ECHELLE DU PAYSAGE

- **Un algorithme d'optimisation pour améliorer la gestion des maladies**
  - L'algorithme utilisé permet d'optimiser conjointement un grand nombre de paramètres (21 dans notre cas), grâce à sa capacité à explorer de manière parcimonieuse l'espace des paramètres possibles.
  - Les défis de ce travail ont été de redéfinir par distorsion les paramètres de gestion et de prendre en compte la stochasticité du modèle ainsi que la coexistence de paramètres discrets et continus.

- **Optimisation *in silico* de la stratégie de gestion de la sharka**
  - Des stratégies optimisées dans le cas d'épidémies émergentes et installées ont été identifiées. Elles sont efficaces pour les 3 niveaux d'agrégation du paysage.
  - D'après le modèle de simulation, ces stratégies sont plus efficaces économiquement que la stratégie de gestion française et plus simples à mettre en place en pratique (elles n'incluent pas d'interdiction de plantation, ni d'arrachages de vergers entiers, et requièrent moins de surveillance des vergers).

3. Optimisation de la répartition de variétés résistantes dans un paysage

Pour gérer les maladies, des variétés résistantes sont aujourd'hui créées et implantées dans le paysage. Néanmoins, l'introduction de résistances dans l'ensemble d'une gamme variétale peut prendre du temps, notamment pour les plantes pérennes. Le remplacement des variétés sensibles à une maladie par des variétés résistantes peut alors difficilement se faire la même année : il se fait généralement de manière progressive au cours du temps. De plus, pour que cette gestion soit durable, toutes les plantes sensibles des parcelles cultivées ne doivent pas être remplacées par des résistantes afin que la résistance ne soit pas contournée. Par conséquent, nous avons travaillé sur l'optimisation de la répartition des variétés résistantes dans le temps et l'espace.

Pour évaluer l'influence du déploiement de variétés résistantes sur la productivité, nous avons simulé des épidémies de sharka (émergentes ou installées) en testant différents scénarios de répartition des variétés résistantes et 3 scénarios de gestion pour 3 types de paysages (différant par leur niveau d'agrégation). De plus, le déploiement optimal des cultivars résistants peut modifier une stratégie de gestion optimale, c'est pourquoi, nous avons de nouveau optimisé la stratégie de gestion de la sharka avec l'algorithme présenté précédemment dans le cas où des variétés sensibles sont remplacées par des résistantes. Les résultats de cette étude sont présentés dans l'article 7.

ARTICLE 7

Optimization of the spatiotemporal deployment of resistant cultivars and disease control options

Coralie Picard, Victor Picheny, Samuel Soubeyrand and Gaël Thébaud

Optimization of the spatio-temporal deployment of resistant cultivars and disease control options

Coralie Picard, Victor Picheny, Samuel Soubeyrand and Gaël Thébaud

First, third and fourth authors: BGPI, INRA, Montpellier SupAgro, Univ. Montpellier, Cirad, TA A-54/K, 34398, Montpellier Cedex 5, France; second author: MIAT, INRA, Castanet-Tolosan, France.

Corresponding author: G. Thébaud
E-mail address: gael.thebaud@inra.fr

ABSTRACT

In order to control plant diseases, which cause significant damage in agricultural crops, various strategies are developed as the use of resistant cultivars. However, because generating resistant cultivars can take numerous years, in particular when the objective is to obtain a varietal range of a same species, all susceptible crops cannot be replaced by resistant ones at the same time, especially for perennial crops. Here, we study how the resistant cultivars should be allocated in the landscape over time to minimize the economic damage. Particularly, we assess the influence of the deployment of resistant varieties both with and without the application of another management strategy. To this end, we used the example of sharka disease, one of the most damaging pathogen of genus *Prunus*. For now, a management strategy based on tree removals, plantation bans and orchard surveillance is applied in France to control this pathogen, and a previous study already showed how it was possible to improve its efficiency. Using a SEIR model, we tested several allocations of resistant orchards for two epidemic cases (emerging and established epidemics) and various aggregations of patches in the landscape. We showed that the most promising deployment of resistant orchards without management was mixing uniformly susceptible and resistant orchards. However, with the application of a management strategy, such deployment does not influence the productivity, which is particularly interesting in practice. In addition, to test whether the optimal management strategy might change when susceptible orchards are progressively replaced by resistant one, we optimized this strategy. Although we identified a strategy which allows improving slightly the productivity, our results indicate that a strategy optimized without the deployment of resistant cultivar can still be efficient in the context where resistant orchards are introduced in the landscape, which is also important for stakeholders.

Keywords: resistant varieties, optimization, management, landscape, SEIR, spatiotemporal model, sharka, virus

1. Introduction

While agriculture has to adapt to the rapidly growing global population and has to reduce pesticide use, plant diseases play a major limiting role in agricultural production. However, management of diseases in cropping systems is often highly challenging since they result of complex interactions between epidemiological processes, human interventions and the organization of patches in the landscape. Management strategies such as the use of chemicals, the removal of infected plants or some changes in cultural practices have proved their efficiency but may remain unsatisfactory. The development of resistant cultivars is another alternative to limit pathogen damage and to reduce the use of phytosanitary products. Nevertheless, the development of a resistant cultivar and the varietal range of a same species can take several years. Thus, the replacement of susceptible cultivars by resistant ones is done progressively over time. In addition, it has been shown that replacing all susceptible plants by resistant is not a sustainable strategy over time (Papaïx et al. 2017). Therefore, one can argue about whether there is an optimal way in the deployment of the resistant varieties over time and space.

In order to study this question, models are helpful thanks to their ability to test several scenarios of epidemic spread and allocation of resistant varieties. Several studies have shown that with a limited number of resistant varieties, the most efficient spatial pattern to minimize incidence was the mixture of resistant and susceptible plants (Holt and Chancellor 1999; Mundt 2002; Mundt and Brophy 1988; Papaïx et al. 2014a; Papaïx et al. 2014b; Skelsey et al. 2010). However, the optimal strategy for deploying resistance can depend on the pathogen dispersal function (short or long distance, Sapoukhina et al. 2010). Indeed, if a disease spreads by short-range dispersal, random mixtures can be used to slow down the epidemic spread because the resistant cultivars create an obstacle to the epidemic spread. In the case of long-range dispersal, heterogeneous patterns including a minimum distance between sensitive units must be used. This last point highlights the importance of patch size and shape for disease dispersal (Mikaberidze et al. 2016). In addition, the application of management strategies may influence the optimal deployment of resistant varieties since both aim to improve crop productivity. However, the epidemiological modeling studies only focused on the deployment of resistant varieties and do not model the application of management strategies at the same time.

In this work, we assess the influence of the deployment of resistant varieties on productivity using a model which enables to simulate both disease dispersal and management strategies, on landscapes with various levels of patch aggregation (Picard et al. in prep; Picard et al. in revision; Pleydell et al.

2018; Rimbaud et al. 2018a; Rimbaud et al. 2018b). We apply this approach to *Plum pox virus* (PPV), a quarantine pathogen which causes the most devastating disease of prunus trees (affecting mainly plum, apricot, and peach production, Cambra et al. 2006; García et al. 2014). This disease, transmitted between hosts through aphids, cause significant economic losses because the associated symptoms make fruit unfit for consumption. To reduce such damage, a strategy based on orchard surveillance, plantation bans and removal of symptomatic trees is applied in France (JORF 2011; Rimbaud et al. 2015). Another approach to control PPV spread is the development of resistant tree varieties: several breeding programs for resistance to PPV in various species in the genus *Prunus* are ongoing (Hartmann and Neumüller 2006; Polák et al. 2017; Zuriaga et al. 2018). For instance, several research laboratories have reported resistance in apricot varieties (Dondini et al. 2011; Pilařová et al. 2010; Vera Ruiz et al. 2011) and a range of resistant apricot trees was developed in 2013 (Mariette et al. 2016). As regards peach trees, no resistant cultivar is currently commercially available but promising studies may suggest that they soon will be (Cirilli et al. 2017; Pascal et al. 2002). In this context, we wonder how these resistant cultivars should be deployed in the landscape to limit the virus damage. To this end, we simulated various allocations of resistant cultivars and three management scenarios (without disease management, with the French management strategy and with an optimized strategy previously identified, Picard et al. in prep). In addition, to be as realistic as possible, we assume that only 50% of the susceptible orchards in the landscape (which are predefined) can be replaced by resistant ones, and we tested 2 hypotheses. The first one assumes that, among the 50% predefined susceptible orchards, all the non-productive orchards can be replaced by resistant cultivars. However, replacing all peach orchards by resistant cultivars can only happen if a range of resistant cultivar can guarantee a large production period. Generally, different resistant cultivars of a same species are not available at the same time since their creation and their acceptation to the official varieties catalogue may take several years. Thus, we also tested the hypothesis that the production of resistant cultivar is limited, which means that, among the 50% predefined susceptible orchards, all the non-productive trees cannot be replaced by resistant cultivars at the same time. In addition, a management strategy may influence the optimal deployment of resistant cultivars, but the opposite is also true. Therefore, we optimized sharka management strategy using a numerical algorithm by taking into account the replacement of susceptible orchards by resistant ones.

2. Materials and methods

2.1. Simulation of sharka spread and management

In order to simulate outbreaks, we used a stochastic, spatially explicit, SEIR (susceptible-exposed-infectious-removed) model initially proposed by Pleydell et al. (2018) and further developed by Picard et al. (in prep); Picard et al. (in revision) and Rimbaud et al. (2018a, 2018b). This orchard-based model simulates, with a discrete time step of 1 week, disease spread and management on landscapes composed of patches on which peach trees are grown. These patches vary by their aggregation level: the model includes patches with a high (H), medium (M) and low (L) level of aggregation. The simulation model accounts for epidemic stochasticity through 6 epidemiological parameters. Depending on their variation ranges, these parameters can represent either an emerging or an established epidemic. In addition, a management strategy based on French and US sharka management in prunus orchards is implemented in the model. It includes 21 parameters representing orchard surveillance, plantation bans and tree removals. This strategy is applied during 30 years after several years of epidemic simulation to allow time for the virus to spread.

The model output is the net present value (NPV), an economic criterion which balances benefits generated by prunus cultivation and the costs associated with production and disease management actions (observation, removal and replantation) (Picard et al. in prep; Rimbaud et al. 2018a). The NPV is calculated for the whole 30-year management period, for both emerging and established epidemics.

2.2. Simulation of the allocation of resistant varieties

The model developed by Picard et al. (in prep) includes 90 simulated landscapes (30 for each aggregation level) composed of 400 patches. In this study, all orchards planted on patches were susceptible to sharka disease. Here, the simulation model was modified to enable the replacement of susceptible orchards by resistant ones if they are removed during the simulation (because of sharka or because they are too old to be productive enough). Replacements by resistant cultivars can take place only during the 30 years of the simulation for which the management strategy is applied.

To this end, patches were assigned to either the resistance zone (where a removed orchard can be replaced by a resistant or a susceptible one) or to the susceptible zone (where a removed orchard is

always replaced by a susceptible one). For each of the 90 landscapes, these zones were defined in several ways (Fig. 1). Then, to simulate a situation where resistant cultivars would be available only progressively, we included in the model the possibility to have a (time-varying) threshold controlling the number of orchards that can be replaced by resistant orchards each year. Starting from 1 at the first year of management, this threshold doubles every 2 years. Note that, it is rare that the number of orchards to replace, which are located in the resistance zone, exceed 35 orchards/years in the case of the epidemic spread fast (for established epidemics on the most aggregated landscape (H) and without management). In simulations without replacement threshold, all removed orchards located in the resistance zone are replaced by resistant orchards. Otherwise, some removed orchards in the resistance zone are chosen randomly and independently to be replaced by resistant orchards (depending on the threshold); others are replaced by susceptible varieties.

All in all, we used 10 different ways to allocate the resistant varieties in a given landscape: 5 possibilities to assign resistant and susceptible zones (2^2, 4^2, 10^2, R and U) x 2 replacement scenarios (with and without replacement threshold). We carried out epidemic simulations of these scenarios on each landscape (334 simulations for each landscape to obtain around 10,000 simulations for each aggregation level), in cases of emerging and established epidemics, and with 3 different management strategies (without disease management, with the French management strategy, and with the optimized strategy from Picard et al. in prep). Two criteria were analyzed: the mean NPV (noted $\overline{\mathrm{NPV}}$) and the mean of the 10% "worst" NPVs among the 10,000 simulations (noted $\mathrm{NPV}_{\overline{10\%}}$). This last criterion was chosen to reduce the likelihood of significant losses.

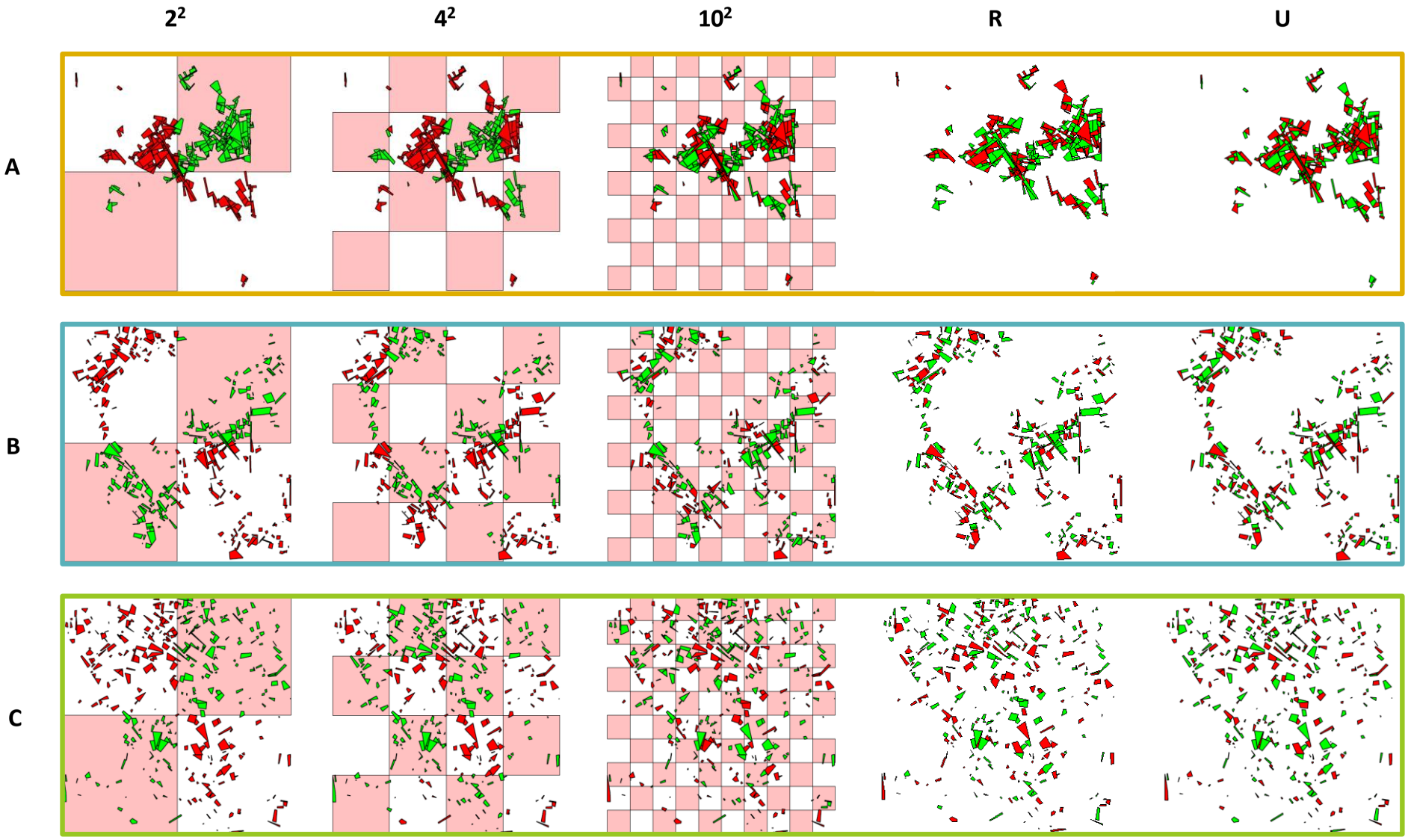


Figure 1: Spatial allocations of resistant cultivars on landscapes with high (A), medium (B), low (C) levels of patch aggregation. Green patches are located in the resistance zone, where resistant or susceptible orchards can be planted. Red patches are located in the susceptible zone, where only susceptible orchards can be planted. From left to right, resistance and susceptible zones are allocated based on a regular squared grid of 4 (2^2), 16 (4^2) and 100 (10^2) squares, uniformly random (R) and uniformly (U, using a minimum spanning tree algorithm between patch centroids: for each pair of neighboring patches in this tree, one patch was assigned to the resistance zone and the other one to the susceptible zone).

2.3. Optimization of the management strategy in the presence of resistant cultivars

We used an algorithm adapted from the R packages DiceKriging and DiceOptim (Picheny and Ginsbourger 2014) to optimize sharka management as in the study of Picard et al. (in prep). Here, we optimized disease management in the presence of a uniform replacement (without replacement threshold) of resistant varieties for established epidemic, for the 3 levels of patch aggregation on $\overline{\mathrm{NPV}}$. Then, we performed 10,000 simulations with these optimized strategies, using epidemic parameters corresponding to established epidemics.

3. Results

3.1. Simulations of various allocations of resistant varieties

Our results showed that NPVs were higher when we don't apply a replacement threshold, which was expected because we add resistant cultivars faster in time. In addition, for the simulations without management, we have a bigger gap of $\mathrm{NPV}_{\overline{10\%}}$ between the scenario 2^2 and the scenario U for simulations performed without replacement threshold than with the threshold. It is probably due to the fact that we reach the final allocation of resistant and susceptible cultivars faster with than without threshold.

Epidemic simulations were first performed without applying any management strategy. In such a situation, we showed that the deployment of resistant orchards in the landscape can influence the NPV for landscapes H and M (Fig. 2 and Supplementary Table S1 and S2). Globally, the NPVs were higher by mixing resistant and susceptible orchards (scenario U), although other scenarios are not significantly different. Regarding the landscape L, the allocation of resistant orchards did not influence the NPV. Then, simulations were performed with the French management strategy. The results showed that the allocation of resistant orchards in the landscape does not influence the NPV in all cases except for simulations on the most aggregated landscape (H) for established epidemics without replacement threshold. However, even in this case, NPV obtained for R, 2^2 and 4^2 scenarios are very close to the NPV obtained for 10^2 and U allocations from which they differ significantly (Supplementary Table S1 and S2). Finally, when simulating disease spread under a previously

identified optimized strategy (Picard et al. in prep), the allocation scenario of resistant orchards in the landscape did not influence the NPV for emerging or established epidemics. To summarize, without management strategy, the uniform allocation of resistant cultivars leads to higher NPVs, and regarding the other scenarios (without management strategy for landscape L, and with the French strategy and the optimized strategy for all landscapes), how to allocate of resistant varieties do not influence the NPV results.

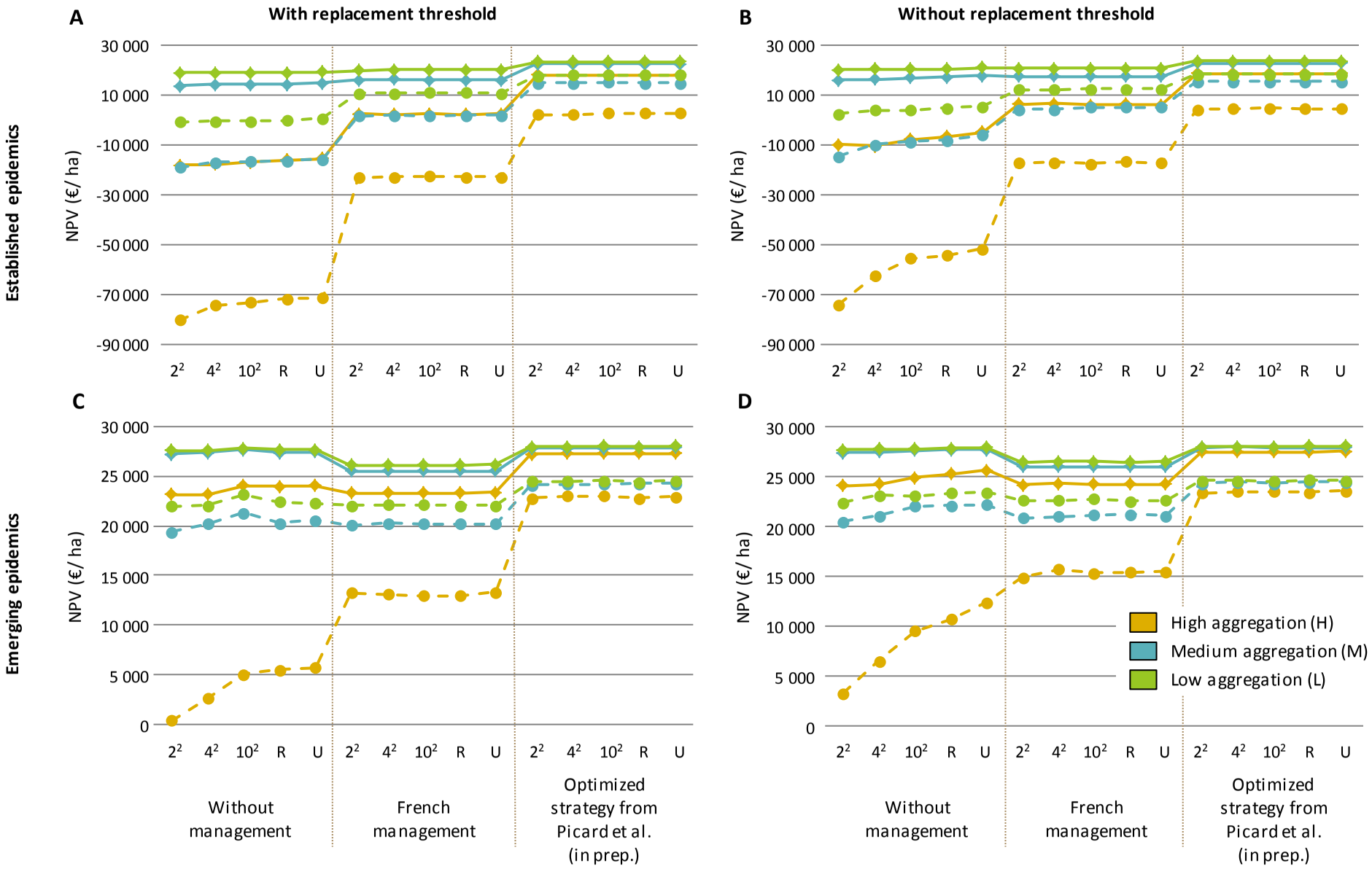


Figure 2: $\overline{\mathbf{NPV}}$ (solid lines) and $\mathbf{NPV}_{\overline{10\%}}$ (dotted lines) obtained after 10,000 simulations of PPV dispersal and management. Simulations were performed for 5 scenarios of allocation of the resistant cultivars in the landscape: 2^4. distribution in 4 squares 4^2. in 16 squares 10^2. in 100 squares R. random distribution U. uniform distribution.

3.2. An optimization case

To test whether the optimal disease management strategy might change when resistant cultivars are progressively introduced in the landscape, an algorithm was used to optimize sharka management with the replacement of removed orchards in a uniform way (without replacement threshold) in the case of established epidemics. We chose this scenario because without applying a threshold, the uniform allocation of resistant varieties led to the best results among all the performed simulations (Fig. 2). An optimization with such scenario is thus probably the one which will lead to the best NPV improvement.

We observed an interaction between the management strategy and the allocation of resistant orchards. Indeed, the optimized strategy obtained in this context had better $\overline{\mathrm{NPV}}$ than the best strategy identified when all cultivars are susceptible (Table 1). Nevertheless, the difference observed between the results of these different strategies is not significant. For instance, we found that the management strategy optimized for the $\overline{\mathrm{NPV}}$ of landscape H lead to results differing by only 0.01% with the previously identified strategy (Picard et al. in prep). This can be explained by the similarity between these 2 strategies (Fig. 3). They differ essentially in the surveillance process: the surveys are more localized around the detected infected tree for the strategy found in this study than in the previous one.

Table 1: $\overline{\mathrm{NPV}}$ (€/ha) obtained after 10,000 simulations of PPV dispersal and management. Simulations were carried out with the replacement of removed orchards in a uniform way for established epidemics. Values in bold represent the $\overline{\mathrm{NPV}}$ (or the $\mathrm{NPV}_{\overline{10\%}}$) corresponding to simulations performed on a landscape with the management optimized for the $\overline{\mathrm{NPV}}$ (or the $\mathrm{NPV}_{\overline{10\%}}$) for the same landscape (e.g. the $\overline{\mathrm{NPV}}$ of simulations performed on landscape H with management parameters optimized for the $\overline{\mathrm{NPV}}$ on the landscape H).

	Landscape H		**Landscape M**		**Landscape L**	
Optimized strategies optimized for :	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$	$\overline{\mathbf{NPV}}$	$\mathbf{NPV}_{\overline{10\%}}$
$\overline{\mathrm{NPV}}$ of landscape H	**18974**	4977	22983	15557	23917	18544
$\overline{\mathrm{NPV}}$ of landscape M	17166	-1698	**23236**	14231	24375	18323
$\overline{\mathrm{NPV}}$ of landscape L	17625	-782	23422	14381	**24547**	18441
Without disease without management	-4802	-51587	17924	-6013	21087	5534
French management strategy	6259	-17129	17754	5264	20881	12587
Optimized strategy from Picard et al. (in prep)	18959	4666	23014	15698	23820	18707

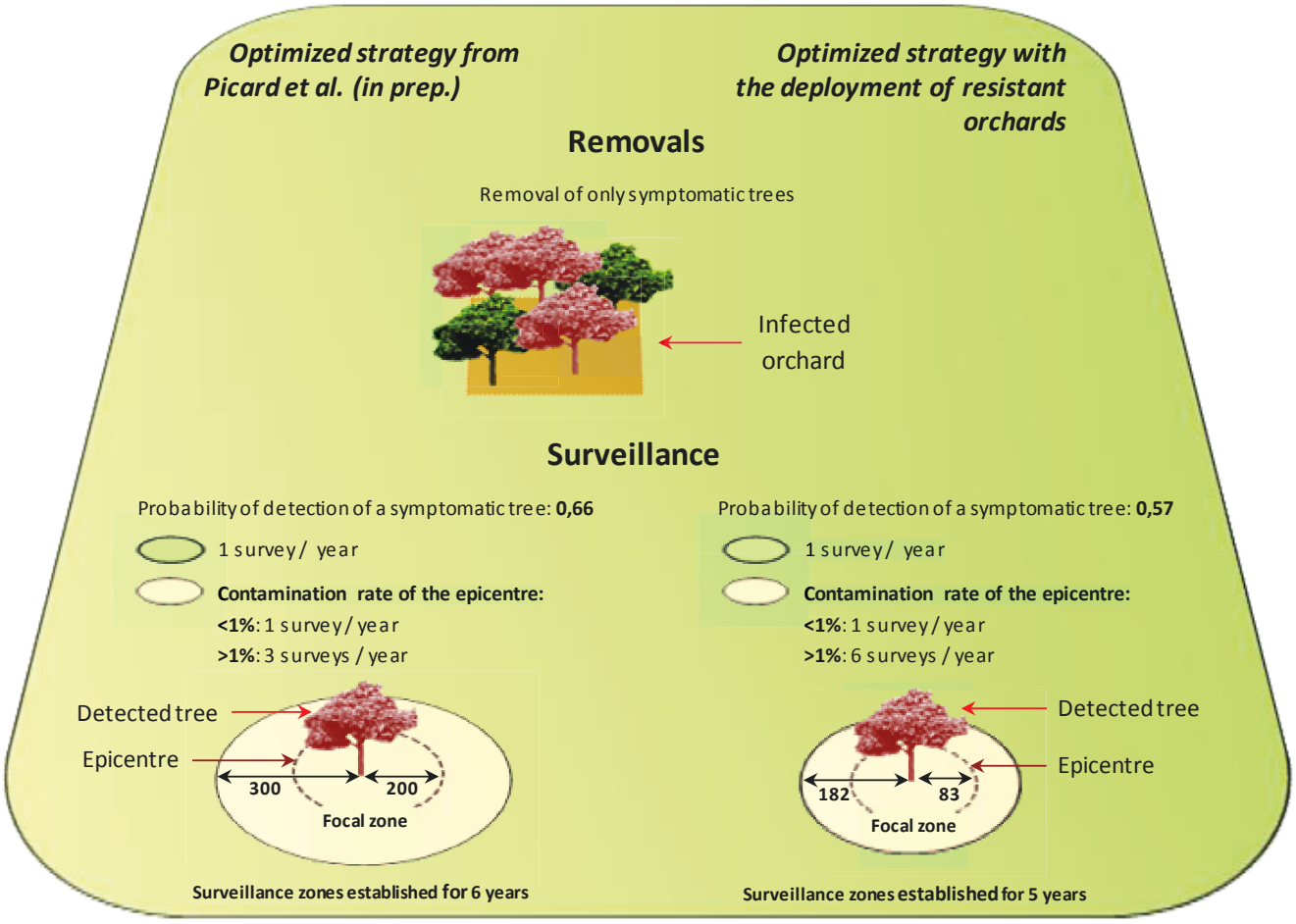


Figure 3: Management actions for optimized strategies of sharka. These strategies were obtained for established epidemics by Picard et al. (in prep) (left) and in this study (right) as a result of the optimization of the $\overline{\mathbf{NPV}}$ for landscape H with replacement of removed orchards in a uniform way (without replacement threshold).

4. Discussion

This work aimed to understand the influence of the allocation of resistant orchards on 3 landscapes varying by their level of patch aggregation. We performed simulations of sharka spread and management with 10 different scenarios of resistant orchards deployment for each landscape aggregation level, and for emerging and established epidemics. These scenarios also included different management strategies in order to assess the influence of the application of such management on the optimal deployment of resistant varieties, which has never been studied before.

In the absence of disease management, the uniform allocation is the most effective strategy whatever the level of landscape aggregation and the epidemic type. These results are consistent with those found by Papaïx et al. (2014a) and Papaïx et al. (2014b), which suggest to mix resistant and susceptible patches. In practice, this means that, when deployment of a resistant cultivar is the only disease control option, each grower should spread its resistant orchards regularly and, ideally, coordinate with its neighbors to maintain regularity across farm boundaries. In addition, others studies recommend to mix resistant and susceptible within a patch (Holt and Chancellor 1999; Mundt 2002; Skelsey et al. 2010) because « disease severity for the mixtures decreased with increasing number of genotype units » (Mundt and Brophy 1988). Thus, it may be interesting to test the effect of mixing resistant and susceptible cultivar in the same orchard with our simulation model. However, such mixture could be problematic for farmers who generally plant the same cultivar in one orchard to facilitate the cultural operations.

On the contrary, whatever the level of landscape aggregation and the epidemic type, when enough control is exerted on the disease (by the French management or the optimized strategy from Picard et al. in prep), the type of allocation of resistant cultivars in the landscape does not influence the NPV. This last point is important in practice because it implies that no collective decision has to be made on which orchards can or cannot be replaced by resistant cultivars, even if the disease management strategy changes in the future. In addition, when the management strategy is optimized in the context of established epidemics with a uniform replantation of resistant orchards (without threshold), we found a strategies which lead to very similar $\overline{\mathrm{NPV}}$ to Picard et al. (in prep) (although our strategy outperformed slightly the previous one). Both strategies are comparable even if the required surveillance is slightly different (in particular, regarding the strategy optimized for landscape H, the presence of resistant orchards enables to reduce even more the surveillance radius around detected trees). In practice, a strategy adapted to the allocations of resistant orchards thus seems possible to deploy, but such efforts may not be necessary since the optimized strategy proposed by Picard et al. (in prep) remains largely efficient.

Nevertheless, we have to keep in mind that the optimization was here performed for one scenario. We might optimize the management strategy of the pathogen in other contexts to confirm our results, as for example, with the use of a threshold controlling the number of orchards that can be replaced by resistant orchards each year.

To go even further in this work, it may be interesting to accept the removal and replacement of some orchards by resistant varieties in the model, even if they still productive. Indeed, in regions with a

high prevalence, farmers might anticipate and replace susceptible orchards by resistant ones before incurring production losses. In addition, our simulations imposed here to keep at least 50% of the patches with susceptible orchards in the landscape, which is not necessarily realistic although this is recommended to avoid resistance breakdowns. In situations where more than 50% of the orchards could be replaced by resistant cultivars, the disease spread may be widely slowed, and the optimal strategy may change. It could thus be another point for reflection. This work might also be improved by taking into account the possibility of resistance breakdown. Indeed, the varietal composition of the landscape can influence a population resistance level (Sapoukhina et al. 2009; Papaïx et al. 2011, 2017). For instance, studies show how such composition influences the resistance level of cereals varieties by altering the structure of the pathogen populations (Papaïx et al. 2011; Rimbaud et al. 2018c). Particularly it was shown that, simulating a resistance breakdown, the uniform deployment of resistant varieties would be optimal (Sapoukhina et al. 2009). Indeed, this study shows that random patterns can reduce both density and genetic diversity of the pathogen population and delay invasion. By contrast, aggregated allocations diversify pathogen population and, hence, reduce the efficacy of resistance genes. However, simulating resistance breakdown for sharka is complex for now because there is little knowledge about resistance mechanisms and their durability, but, although we did not account for the resistance breakdown in our study, our conclusions still the same: the uniform allocation of resistant cultivar is recommended.

ACKNOWLEDGMENTS

The authors thank François Bonnot for his help on the use of the algorithm used to define the uniform allocation of resistant cultivars. This work was supported by the CIRAD-UMR AGAP HPC Data Center of the South Green Bioinformatics platform (http://www.southgreen.fr), and we are grateful to Sébastien Ravel for the associated help.

LITERATURE CITED

Cambra, M., Capote, N., Myrta, A., and Llácer, G. 2006. *Plum pox virus* and the estimated costs associated with sharka disease. EPPO Bull. 36:202–204.

Cirilli, M., Rossini, L., Geuna, F., Palmisano, F., Minafra, A., Castrignanò, T., Gattolin, S., Ciacciulli, A., Babini A., and Liverani, A. 2017. Genetic dissection of Sharka disease tolerance in peach (P. persica L. Batsch). BMC Plant Biol. 17(1):192.

Dondini, L., Lain, O., Vendramin, V., Rizzo, M., Vivoli, D., Adami, M., Guidarelli, M., Gaiotti, F., Palmisano, F., and Bazzoni, A. 2011. Identification of QTL for resistance to plum pox virus strains M and D in Lito and Harcot apricot cultivars. Mol. Breed. 27(3):289–299.

García, J. A., Glasa, M., Cambra, M., and Candresse, T. 2014. *Plum pox virus* and sharka: a model potyvirus and a major disease. Mol. Plant Pathol. 15(3):226–241.

Hartmann, W., and Neumüller, M. 2006. Breeding for resistance: breeding for *Plum pox virus* resistant plums (*Prunus domestica* L.) in Germany. EPPO Bull. 36(2):332–336.

Holt, J., and Chancellor, T. C. B. 1999. Modelling the spatio-temporal deployment of resistant varieties to reduce the incidence of rice tungro disease in a dynamic cropping system. Plant Pathol. 48(4):453–461.

JORF 2011. Arrêté du 17 mars 2011 relatif à la lutte contre le *Plum Pox Virus*, agent causal de la maladie de la Sharka, sur les végétaux sensibles du genre Prunus. République Française. JORF n°0067 du 20 mars 2011. Paris, France. http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023728913&dateTexte=&categorieLien=id.

Mariette, S., Wong Jun Tai, F., Roch, G., Barre, A., Chague, A., Decroocq, S., Groppi, A., Laizet, Y, Lambert, P., and Tricon, D. 2016. Genome-wide association links candidate genes to resistance to *Plum Pox Virus* in apricot (Prunus armeniaca). New Phytol. 209(2):773–784.

Mikaberidze, A., Mundt, C. C., and Bonhoeffer, S. 2016. Invasiveness of plant pathogens depends on the spatial scale of host distribution. Ecol. Appl. 26(4):1238–1248.

Mundt, C.C. 2002. Use of multiline cultivars and cultivar mixtures for disease management. Annu. Rev. Phytopathol. 40(1):381–410.

Mundt, C.C., and Brophy, L. S. 1988. Influence of number of host genotype units on the effectiveness of host mixtures for disease control: a modeling approach. Phytopathology. 78(8):1087–1094.

Papaïx, J., Goyeau, H., Du Cheyron, P., Monod, H., and Lannou, C. 2011. Influence of cultivated landscape composition on variety resistance: an assessment based on wheat leaf rust epidemics. New Phytol. 191(4):1095–1107.

Papaïx, J., Touzeau, S., Monod, H., and Lannou, C. 2014a. Can epidemic control be achieved by altering landscape connectivity in agricultural systems? Ecol. Model. 284:35–47.

Papaïx, J., Adamczyk-Chauvat, K., Bouvier, A., Kiêu, K., Touzeau, S., Lannou, C., and Monod, H. 2014b. Pathogen population dynamics in agricultural landscapes: The Ddal modelling framework. Infect. Genet. Evol. 27:509–520.

Papaïx, J., Rimbaud, L., Burdon, J. J., Zhan, J., and Thrall, P. H. 2017. Differential impact of landscape-scale strategies for crop cultivar deployment on disease dynamics, resistance durability and long-term evolutionary control. Evol. Appl. 11(5), 705-717.

Pascal, T., Pfeiffer, F., and Kervella, J. 2002. Preliminary observations on the resistance to sharka in peach and related species. In *Acta Horticulturae*, International Society for Horticultural Science (ISHS), Leuven, Belgium, p. 699–704. Available at: https://doi.org/10.17660/ActaHortic.2002.592.98.

Picard, C., Soubeyrand, S., Jacquot, E., and Thébaud, G. in revision. Analyzing the influence of landscape aggregation on disease spread to improve management strategies.

Picard, C., Picheny, V., Bonnot, F., Soubeyrand, S., and Thébaud, G. in prep. In silico optimization of a strategy for landscape-wide plant disease management.

Picheny, V., and Ginsbourger, D. 2014. Noisy kriging-based optimization methods: a unified implementation within the DiceOptim package. Comput. Stat. Data Anal. 71:1035–1053.

Pilařová, P., Marandel, G., Decroocq, V., Salava, J., Krška, B., and Abbott, A. G. 2010. Quantitative trait analysis of resistance to plum pox virus in the apricot F1 progeny “Harlayne” × “Vestar.” Tree Genet. Genomes. 6(3):467–475.

Pleydell, D. R. J., Soubeyrand, S., Dallot, S., Labonne, G., Chadœuf, J., Jacquot, E., and Thébaud G. 2018. Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLoS Comput. Biol. 14(4): e1006085.

Polák, J., Kundu, J. K., Krška, B., Beoni, E., Komínek, P., Pívalova, J., and Jaro, J. 2017. Transgenic plum Prunus domestica L., clone C5 (cv. HoneySweet) for protection against sharka disease. J. Integr. Agric. 16(3):516–522.

Rimbaud, L., Dallot, S., Gottwald, T., Decroocq, V., Jacquot, E., Soubeyrand, S., and Thébaud G. 2015. Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annu. Rev. Phytopathol. 53:357–378.

Rimbaud, L., Bruchou, C., Dallot, S., Pleydell, D. R. J., Jacquot, E., Soubeyrand, S., and Thébaud, G. 2018a. Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. R. Soc. Open Sci. 5:171435.

Rimbaud, L., Dallot, S., Bruchou, C., Thoyer, S., Jacquot, E., Soubeyrand, S., and Thébaud G. 2018b. Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. bioRxiv 315747. (doi: https://doi.org/10.1101/315747)

Rimbaud, L., Papaïx, J., Rey, J.-F., Barrett, L. G., and Thrall, P. H. 2018. Assessing the durability and efficiency of landscape-based strategies to deploy plant resistance to pathogens. PLOS Comput. Biol. 14(4):e1006067.

Sapoukhina, N., Durel, C.E., and Cam, B. L. 2009. Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations. Theor. Ecol. 2(4):229.

Sapoukhina, N., Tyutyunov, Y., Sache, I., and Arditi, R. 2010. Spatially mixed crops to control the stratified dispersal of airborne fungal diseases. Ecol. Model. 221(23):2793–2800.

Skelsey, P., Rossing, W. A. H., Kessel, G. J. T., and van der Werf, W. 2010. Invasion of Phytophthora infestans at the landscape level: how do spatial scale and weather modulate the consequences of spatial heterogeneity in host resistance? Phytopathology. 100(11):1146–1161.

Vera Ruiz, E. M., Soriano, J. M., Romero, C., Zhebentyayeva, T., Terol, J., Zuriaga, E., Llacer, G., Abbott, A., and Badenes, M. 2011. Narrowing down the apricot *Plum pox virus* resistance locus and comparative analysis with the peach genome syntenic region. Mol. Plant Pathol. 12(6):535–547.

Zuriaga, E., Romero, C., Blanca, J. M., and Badenes, M. L. 2018. Resistance to Plum Pox Virus (PPV) in apricot (Prunus armeniaca L.) is associated with down-regulation of two MATHd genes. BMC Plant Biol. 18(1):25.

SUPPORTING INFORMATION

Supplementary Table S1: Statistical comparison of $\overline{\mathbf{NPV}}$ obtained for emerging epidemics under 90 scenarios. These scenarios correspond to the complete factorial design for: 3 management strategies x 3 landscape aggregation levels x 2 resistance availability thresholds x 5 resistance allocations. Letters correspond to the result of a Tukey HSD test performed for each management strategy scenario: for instance, scenarios without management strategy presenting the same letter are not significantly different. To facilitate easy reading, for each management strategy, the boxes representing the scenario leading to the best $\mathbf{NPV_{10\%}}$ are colored, as well as the scenarios which are not significantly different.

		Aggregation level	Replacement threshold of resistant varieties	Allocation scenarios of resistant orchards				
				22	42	102	R	U
Emerging epidemics	Without management	High	With	o	o	n	n	n
			Without	mn	m	l	k	j
		Medium	With	i	cdefgh	bcdefg	ghi	fghi
			Without	efgh	defgh	abcde	abcde	abcd
		Low	With	defh	cdefgh	abcd	bcdef	bcdef
			Without	abcde	ab	abc	a	a
	French management	High	With	f	f	f	f	f
			Without	e	defgh	de	de	de
		Medium	With	c	c	c	c	c
			Without	b	b	b	b	b
		Low	With	b	b	b	b	b
			Without	a	a	a	a	a
	Optimized strategy from Picard et al. (in prep)	High	With	i	i	i	i	i
			Without	h	h	h	h	h
		Medium	With	g	fg	fg	fg	efg
			Without	cde	abc	bcd	bcd	bcd
		Low	With	bcd	bcd	bcd	def	bcd
			Without	abc	ab	abc	a	a

Supplementary Table S2: Statistical comparison of $\overline{\mathbf{NPV}}$ obtained for established epidemics under 90 scenarios. These scenarios correspond to the complete factorial design for: 3 management strategies x 3 landscape aggregation levels x 2 resistance availability thresholds x 5 resistance allocations. Letters correspond to the result of a Tukey HSD test performed for each management strategy scenario: for instance, scenarios without management strategy presenting the same letter are not significantly different. To facilitate easy reading, for each management strategy, the boxes representing the scenario leading to the best $\mathbf{NPV}_{\overline{10\%}}$ are colored, as well as the scenarios which are not significantly different.

		Aggregation level	Replacement threshold of resistant varieties	Allocation scenarios of resistant orchards				
				22	42	102	R	U
Established epidemics	Without management	High	With	o	o	n	mn	m
			Without	l	l	k	j	i
		Medium	With	h	gh	gh	gh	g
			Without	f	ef	de	de	d
		Low	With	c	c	c	c	bc
			Without	ab	a	a	a	a
	French management	High	With	g	g	g	g	g
			Without	ef	e	f	ef	f
		Medium	With	d	d	d	d	d
			Without	c	c	c	c	c
		Low	With	b	b	b	b	b
			Without	a	a	a	a	a
	Optimized strategy from Picard et al. (in prep)	High	With	f	f	f	f	f
			Without	e	e	e	e	e
		Medium	With	d	d	d	d	d
			Without	c	c	c	c	c
		Low	With	b	b	b	b	b
			Without	a	a	a	a	a

Résultats clés de l'Article 7

OPTIMISATION DANS UN PAYSAGE DE LA REPARTITION DE VARIETES RESISTANTES ET OPTIONS DE LUTTE

- **Influence des stratégies de gestion sur la VAN lorsque des vergers sensibles sont remplacés par des résistants**
  - 5 scénarios de répartition (aléatoire, uniforme, et 3 niveaux d'agrégation différents) des variétés résistantes ont été testés pour chacun des 3 paysages variant par le niveau d'agrégation de leurs parcelles.
  - En l'absence de gestion, les résultats montrent que la répartition uniforme des variétés résistantes est la plus efficace, que ce soit dans le cas d'épidémies émergentes ou installées.
  - Lorsqu'une stratégie de gestion est appliquée (stratégie de gestion française ou stratégie optimisée obtenue dans l'article 6), la façon de répartir les variétés résistantes dans le paysage n'a pas d'influence sur la VAN. Ce point est important dans la pratique car, même si la stratégie de gestion de la maladie est remplacée à l'avenir par une stratégie optimisée, aucune décision collective ne doit être prise sur les vergers, qu'il y ait ou non des remplacements de vergers sensibles par des résistants.

- **Optimisation de la gestion avec une répartition uniforme des variétés résistantes**
  - En présence de variétés résistantes déployées de façon uniformes dans le paysage, une nouvelle stratégie optimale peut être identifiée, et sa VAN est sensiblement identique à celle obtenue sur ce nouveau paysage avec la stratégie optimisée dans l'article 6.
  - Cette stratégie optimisée est proche de celle obtenue dans l'article 6 ; seules les fréquences de surveillance sont sensiblement différentes (en particulier, la présence de vergers résistants permet de réduire encore plus le rayon de surveillance autour des arbres détectés).

Conclusion générale
&
Perspectives

1. Conclusion générale

Les objectifs de ce travail de thèse étaient de caractériser la dynamique d'une épidémie et de simuler sa dispersion à travers des paysages variés pour analyser l'impact des caractéristiques du paysage et rechercher des stratégies de gestion adaptées. Cette démarche, nommée **PESO** (Parameter Estimation – Simulation – Optimization), est présentée dans l'article 1 (Picard et al. 2017). Ce travail s'appuie sur l'exemple de la sharka, causée par le ***Plum pox virus*** (PPV). Cette maladie, qui entraine des pertes économiques dans la majorité des régions productrices de ***Prunus***, est aujourd'hui règlementée en France par un arrêté national (JORF, 2011). Ce pathogène fait également l'objet de contraintes règlementaires du fait de son statut de pathogène de quarantaine et les expérimentations en champ sont interdites. Aussi, pour répondre aux objectifs de cette thèse, j'ai utilisé un modèle stochastique spatio-temporel permettant de simuler la dynamique épidémique ainsi que des stratégies de gestion variées (Pleydell et al. 2018; Rimbaud et al. 2018a, 2018b).

Afin d'estimer indépendamment la valeur du **paramètre de dispersion** implémenté dans ce modèle (Pleydell et al. 2018), nous avons utilisé un modèle permettant d'**estimer la dynamique de l'épidémie** grâce à des données spatio-temporelles et génétiques. Plus particulièrement, ce modèle nous a permis d'inférer « qui a infecté qui » dans un paysage et d'évaluer les paramètres épidémiologiques sous-jacents. Dans un premier temps, en appliquant ce modèle à des données simulées, nous avons montré que la prise en compte des arbres sains dans l'inférence permettait d'obtenir des résultats plus fiables. Par la suite, nous avons appliqué ce modèle d'inférence à des données réelles : nous disposions de données épidémiologiques et génétiques d'une épidémie de sharka, échantillonnées dans plusieurs vergers de *Prunus*. Cependant, les résultats de l'inférence ne nous ont pas paru vraisemblables et l'estimation des chaines de transmission pourrait être largement améliorée. Néanmoins, les estimations des autres paramètres épidémiologiques se rapprochent des résultats de Pleydell et al. (2018) qui ont été utilisés dans la suite de ma thèse.

Des **simulations d'épidémies** de sharka ont ensuite été réalisées à l'aide du modèle spatio-temporel intégrant des paysages différant par le niveau d'agrégation de leurs parcelles. Nous avons montré que le paysage influence l'efficacité des stratégies de gestion. De plus, des analyses de sensibilité nous ont permis de mettre en évidence 2 paramètres de gestion influents : ils concernent les interdictions de planter des vergers ainsi que les arrachages de vergers entiers.

La dernière partie de notre démarche consistait à **optimiser les paramètres de gestion de l'épidémie** grâce aux résultats d'analyses de sensibilité puis à l'aide d'un algorithme d'optimisation. Les résultats

ont montré *in silico* qu'il devrait être possible d'améliorer économiquement la stratégie française actuellement en vigueur. Nous avons identifié des stratégies optimisées spécifiques à chaque niveau d'agrégation du paysage, ainsi que des stratégies efficaces pour l'ensemble des paysages simulés (dans le cas d'épidémies émergentes ou installées). Ces stratégies requièrent moins de moyens que la stratégie française actuelle. En effet, elles ne prévoient aucune interdiction de plantation et n'imposent pas l'arrachage de vergers entiers (ce qui est en accord avec les résultats des analyses de sensibilité), et les zones de surveillance sont plus restreintes. De plus, le nombre de prospections à l'intérieur de chaque zone surveillée est inférieur à celui de la stratégie française pour les stratégies optimisées dans le cas d'épidémies émergentes, et plus élevé dans le cas d'épidémies installées. Enfin, nous avons étudié l'influence de la répartition de variétés résistantes dans un paysage. Nous avons montré qu'en l'absence de gestion, une répartition uniforme devrait permettre de limiter les dégâts provoqués par la sharka. Néanmoins, lorsqu'une stratégie de gestion est appliquée, la répartition des variétés résistantes n'influe pas sur les sorties du modèle : les variétés résistantes peuvent être remplacées de manière uniforme ou agrégée dans le paysage, la VAN ne sera pas modifiée.

Les résultats de mes travaux devraient pouvoir guider les décisions des acteurs en charge de la gestion des épidémies.

2. Perspectives

2.1. Amélioration du processus PESO

Le processus PESO, proposé et décliné au cours de ma thèse, est basé sur un ensemble d'hypothèses qui nous ont paru pertinentes par rapport à nos objectifs. Toutefois, certaines de ces hypothèses peuvent avoir des conséquences sur les résultats obtenus. Cette partie présente des pistes pour améliorer les 3 étapes du processus PESO.

2.1.1. Estimation des paramètres épidémiologiques

- **Estimer les paramètres gouvernant l'épidémie à partir de nouveaux jeux de données**

Dans le chapitre 1, nous avons tenté d'estimer les paramètres épidémiologiques de la sharka à partir d'un jeu de données différent de celui utilisé jusqu'à présent (Pleydell et al. 2018). Néanmoins, les résultats n'ont pas été satisfaisants car la méthode utilisée (inférence via un modèle généticospatio-temporel) est nettement plus efficace quand l'échantillonnage est étalé dans le temps. Or, nos

données correspondaient à seulement 3 années consécutives d'échantillonnage, ce qui est très peu. En effet, étant donnée la période de latence de la maladie qui peut aller de plusieurs mois à plusieurs années et la détection imparfaite des arbres symptomatiques, il est probable que peu d'arbres détectés la dernière année aient été infectés pendant la période d'échantillonnage. Nous avons donc utilisé les paramètres épidémiologiques estimés par Pleydell et al. (2018) dans tous nos travaux. Cependant, ces estimations ont été réalisées à partir de données d'incidence à l'échelle de la parcelle. Pour être plus précis, il pourrait être intéressant de réaliser de nouveau les estimations des paramètres épidémiologiques avec le modèle généticospatio-temporel en échantillonnant des données à l'échelle de l'arbre sur plusieurs années.

2.1.2. Simulation de l'épidémie

- **Améliorer la simulation des dates et de la probabilité de détection**

Les stratégies de gestion implémentées dans notre modèle requièrent parfois de prospecter plusieurs fois un même verger au cours d'une année. Dans nos simulations, les dates de prospection de ce verger sont tirées aléatoirement dans la période d'observation, ce qui n'est pas le cas en réalité. En effet, les dates de prospections sont établies judicieusement afin de maximiser les chances de détecter des arbres infectés qui étaient asymptomatiques (ou qui avaient des symptômes difficilement visibles) lors de prospections précédentes. Ainsi, les prospections visuelles sont réparties pour surveiller les arbres dans différentes conditions (par exemple à divers moments de la période de prospection sur fleurs et feuilles, ou bien dans diverses conditions météorologiques). L'intérêt de prospecter plusieurs fois par an un même verger est donc surement sous-estimé dans nos simulations. Pour y remédier, les dates des prospections dans le modèle pourraient être réparties de manière plus homogène dans le temps.

Par ailleurs, la probabilité de détecter un arbre infecté dans notre modèle est la même pour tous les arbres et cette probabilité est constante dans le temps. Or en réalité, cette probabilité de détection peut être différente en fonction de l'organe prospecté (fleur, feuille), de l'âge du verger et de l'âge de l'infection car ces facteurs peuvent influer sur la visibilité des symptômes. Le processus de détection des hôtes symptomatiques intégré dans le modèle pourrait donc être adapté pour permettre de prendre en compte l'évolution des symptômes au cours du temps (en faisant par exemple varier la probabilité de détection en fonction du mois de l'année et donc du stade végétatif des arbres). De plus, les observations des arbres symptomatiques sont simulées de manière indépendante dans notre modèle, ce qui a surement entrainé une surestimation du taux de

détection (dans les vergers, lorsqu'un arbre infecté n'est pas détecté lors d'une campagne de surveillance, il n'est pas rare qu'il ne soit également pas détecté lors de la prospection suivante).

- **Simuler plus précisément la structure du paysage**

Dans cette étude, nous avons réalisé des simulations sur des paysages comportant 400 parcelles plus ou moins agrégées. Bien que cet aspect permette d'obtenir des résultats plus génériques qu'avec l'étude de Rimbaud et al. (2018a) qui proposait une optimisation sur un paysage unique, des caractéristiques du paysage qui influencent la dispersion du virus pourraient encore être intégrées dans le modèle (Brunker et al. 2018; Ostfeld et al. 2005). Ainsi, des structures paysagères telles que les montagnes, les lacs, les rivières, les forêts ou encore les routes pourraient être représentées dans le paysage. Elles pourraient agir comme des barrières ou des corridors qui empêcheraient ou favoriseraient la dispersion du virus. De plus, des zones comprenant des espèces hôtes du puceron pourraient être simulées : la dispersion du virus sur les parcelles cultivées voisines serait alors accentuée. Il peut être difficile de prendre en compte ces caractéristiques du paysage dans le cas d'une étude générique, mais la possibilité de les prendre en compte dans le modèle peut devenir rapidement intéressante lorsque l'on souhaite réaliser des simulations à partir d'un paysage réel (pour optimiser les stratégies de gestion d'une épidémie dans une région particulière par exemple).

Dans l'article 4, nous avons analysé l'influence de la taille du paysage sur différentes stratégies de gestion en répliquant un paysage réel. Du fait du temps de calcul qui augmente rapidement avec le nombre de parcelles prises en compte, nous avons par la suite optimisé la stratégie de gestion seulement pour des paysages simulés composés de 400 parcelles (article 6). Cependant, la taille du paysage étudié pourrait avoir des conséquences sur les résultats des optimisations. En effet, une stratégie tendant vers l'éradication n'aura pas nécessairement la même influence sur la VAN suivant la taille du paysage. Ainsi, sur un petit paysage, une surveillance renforcée ou des arrachages massifs couteront cher et ne seront pas compensés par les bénéfices obtenus dans une zone non infectée. A l'inverse, sur un grand paysage infecté de façon hétérogène, les mesures de gestion pourraient se traduire par des coûts inférieurs aux bénéfices obtenus dans une zone non infectée du paysage. Le processus d'introduction simulé dans notre modèle limite cet effet car le virus est régulièrement introduit dans des zones indemnes par la plantation de vergers infectés. En revanche, la taille du paysage pourrait influencer la VAN dans des situations où la rareté des introductions garantit qu'une proportion suffisante du paysage reste indemne du virus.

Pour finir, pour réaliser des simulations en limitant le temps de calcul, le modèle utilisé dans cette thèse fonctionne à l'échelle du verger. Au cours d'une simulation, nous ne connaissons que le pourcentage d'arbres infectés de chaque parcelle, mais nous ignorons la position des arbres infectés dans les vergers. Ceux-ci sont considérés répartis uniformément sur la surface du verger, alors que les arbres infectés par le PPV au sein d'un verger sont généralement agrégés (Dallot et al. 2003; Gottwald et al. 1995, 2013). Dans notre modèle, si les arbres infectés sont regroupés sur le côté d'une parcelle, nous supposons que toutes les parcelles voisines peuvent être infectées avec la même probabilité. Or, la probabilité que la parcelle la plus proche du groupe d'arbres infectés soit infectée devrait être plus élevée que pour les autres parcelles. Pour être plus précis, il pourrait être intéressant de reprogrammer le modèle à l'échelle de l'arbre, notamment si l'objectif est de simuler la dispersion d'un pathogène à travers des parcelles de formes et de tailles variées.

- **Adapter le modèle de simulation à différentes maladies**

L'approche globale PESO présentée dans l'article 1 est générique, néanmoins, nous l'avons ici appliquée à la sharka qui présente des caractéristiques spécifiques. Pour transposer cette démarche à une autre maladie épidémique, il est nécessaire d'estimer les paramètres épidémiologiques de cette maladie (afin de définir leurs domaines de définition) en étudiant ses caractéristiques (vecteurs, mode de transmission, dynamique épidémique, etc.) ainsi que la réponse de ses hôtes à l'infection (périodes d'incubation et de latence, visibilité des symptômes, etc.). Des paysages doivent également être simulés pour que les caractéristiques des parcelles composées de plantes hôtes soient réalistes (taille, forme et niveau d'agrégation des parcelles). Enfin, l'étape la plus complexe consiste à modifier la simulation des modalités de gestion si des méthodes de lutte autres que la surveillance, l'arrachage, et l'interdiction de plantation existent pour cette maladie et à adapter les mécanismes du modèle aux caractéristiques de la maladie. Dans notre modèle les hôtes sont par exemple caractérisés par 5 statuts différents au cours d'une simulation (« sains », « infecté », « infectieux cachés », « infectieux détecté » et « arraché ») car il a été montré pour la sharka que la période d'incubation et de latence était de même durée (Rimbaud et al. 2015a). Cependant, pour adapter la démarche PESO à une maladie qui cause des symptômes après que l'hôte soit infectieux (ou l'inverse), un statut doit être ajouté au modèle.

Le processus PESO présenté dans cette thèse pourrait être par exemple adapté à la tristeza (qui cause des dommages sur les agrumes notamment en Espagne, au Portugal et en Floride) car elle présente des caractéristiques similaires à la sharka. En effet, cette maladie, causée par un pathogène

de quarantaine, fait l'objet de contraintes réglementaires obligatoires : la méthode de lutte est également l'arrachage des arbres contaminés. De plus, sa transmission se fait par le biais de pucerons ou par la plantation de matériel infecté (produit en pépinière). Néanmoins, des différences notables avec la sharka entraineraient une modification du modèle. Par exemple, à la différence de la sharka, le virus de la tristeza des agrumes est transmis par pucerons sur le mode semi-persistant (le puceron acquiert rapidement le virus et peut le transmettre directement à un autre hôte, et ce durant une trentaine de minutes). La fonction de dispersion du modèle devra alors être changée. De plus, la simulation de la gestion doit être adaptée car la surveillance de cette maladie se fait moyennant des techniques biologiques, sérologiques et moléculaires.

2.1.3. Optimisation de la gestion d'une épidémie

- **Optimiser la gestion sur la base d'un risque épidémique**

Dans l'article 4, les résultats des analyses de sensibilité ont montré que le paramètre qui correspond à la connectivité de la parcelle de première introduction a une forte influence sur la VAN (pour le paysage le plus agrégé, ce paramètre est d'ailleurs le 3ème plus influent parmi l'ensemble des paramètres épidémiologiques et de gestion). Ce résultat s'explique facilement : si la maladie est introduite dans une parcelle très connectée, l'épidémie risque de se répandre vite et les dommages seront importants. La stratégie de gestion de l'épidémie pourrait donc être réfléchie en prenant en compte un indice de risque épidémique des parcelles. Pour cela, la fréquence des prospections ou le seuil d'arrachage des vergers pourraient être définis en fonction du risque encouru pour chaque parcelle, et du risque qu'elle fait peser sur les autres si elle est contaminée. Par exemple, on pourrait accorder plus d'attention aux parcelles connectées les unes aux autres (l'indice de risque serait ici défini par la connectivité des parcelles). Nous pourrions nous inspirer de travaux qui proposent un échantillonnage basé sur le risque, développés pour des maladies spécifiques (te Beest et al. 2011; Gottwald et al. 2013; Hyatt-Twynam et al. 2017; Meentemeyer et al. 2011; Robinet et al. 2012) ou dans un cadre plus générique (Ostfeld et al. 2005; Parnell et al. 2014).

- **Développer des stratégies de gestion adaptatives dans le temps**

Dans les différents travaux effectués au cours de ma thèse, la simulation de la dispersion du pathogène et de sa gestion se déroule sur une période de 30 ans. La VAN est calculée sur cette même période et correspond à la somme actualisée (taux d'actualisation de 4%) des marges brutes de

chaque année, ce qui implique qu'une plus grande importance est accordée à la marge brute des premières années de l'épidémie qu'aux dernières. Le critère économique d'optimisation que nous avons utilisé favorise donc des marges brutes élevées les premières années de l'épidémie (Figure 1A). Cependant, les stratégies optimisées pourraient ne pas être adaptées aux dernières années de l'épidémie alors qu'on peut penser que les *Prunus*, et sans doute la sharka, seront encore présents en France au-delà de la 30ème année. Par ailleurs, le scénario de gestion optimal d'une épidémie pourrait varier dans le temps. Ainsi, pour une épidémie installée, une stratégie requérant de nombreuses prospections pourrait être utile quelques années pour limiter la dispersion de la maladie et faire diminuer la prévalence ; cependant, il est possible que cette stratégie ne soit plus optimale par la suite : une gestion moins stricte pourrait alors être moins couteuse mais tout aussi efficace épidémiologiquement.

Pour pallier ce problème, la stratégie de gestion d'une épidémie pourrait varier au cours du temps en fonction de l'état épidémique des parcelles dans le paysage (i.e. en permettant aux paramètres de gestion de prendre des valeurs différentes au cours du temps). Pour cela, il serait intéressant de développer une méthode qui permette d'optimiser les paramètres de gestion au cours du temps.

Dans un premier temps, nous pourrions optimiser un enchainement de plusieurs scénarios de gestion pour lesquels les paramètres seraient estimés au début de l'épidémie (Figure 1B). Cette approche permettrait de proposer une gestion optimisée pour un ensemble d'épidémies (avec la $\overline{\mathrm{VAN}}$ ou la $\mathrm{VAN}_{\overline{10\%}}$). Cependant, elle impliquerait de multiplier le nombre de paramètres de gestion par le nombre de scénarios de gestion appliqués. Le temps de calcul peut alors rapidement devenir important.

Alternativement, les paramètres de gestion pourraient être optimisés au cours du temps en fonction de l'état épidémique des parcelles, c'est-à-dire en intégrant en temps réel les données relatives à l'état sanitaire des cultures (Figure 1C). Avec cette approche, l'optimisation des paramètres de gestion serait réalisée sur un paysage réel (l'état sanitaire des parcelles serait donc connu grâce aux données de surveillance). Les paramètres de gestion seraient optimisés comme dans l'article 6, c'est-à-dire en réalisant plusieurs itérations composées chacune de plusieurs réplicats d'épidémies, mais en partant toujours du même état épidémique des parcelles (dans l'article 6, l'état initial des parcelles était défini par la simulation de 5 ou 15 ans d'épidémie avant l'optimisation de la gestion). Ce processus pourrait être utile en pratique pour les gestionnaires du risque : à chaque pas de temps, ils mettraient à jour leur base de données sur l'état épidémique des parcelles (à partir des données de surveillance), mais également sur l'apparition, la disparition ou la modification de forme d'un

verger. Une nouvelle stratégie de gestion pourrait alors être identifiée pour un paysage donné. Pour faciliter la mise en place d'un tel processus à l'échelle nationale, on pourrait imaginer que l'optimisation de la gestion se fasse chaque année sur des bassins de production différents (i.e. si le pas de temps est de 5 ans, la gestion serait optimisée pour 20% des régions chaque année). De plus, pour améliorer la précision des simulations, les données de surveillance pourraient être utilisées à chaque pas de temps pour inférer les valeurs des paramètres épidémiologiques (voire les erreurs de prédiction du modèle) et les réajuster. Néanmoins, l'estimation de ces paramètres épidémiologiques demande une quantité de travail importante au niveau méthodologique.

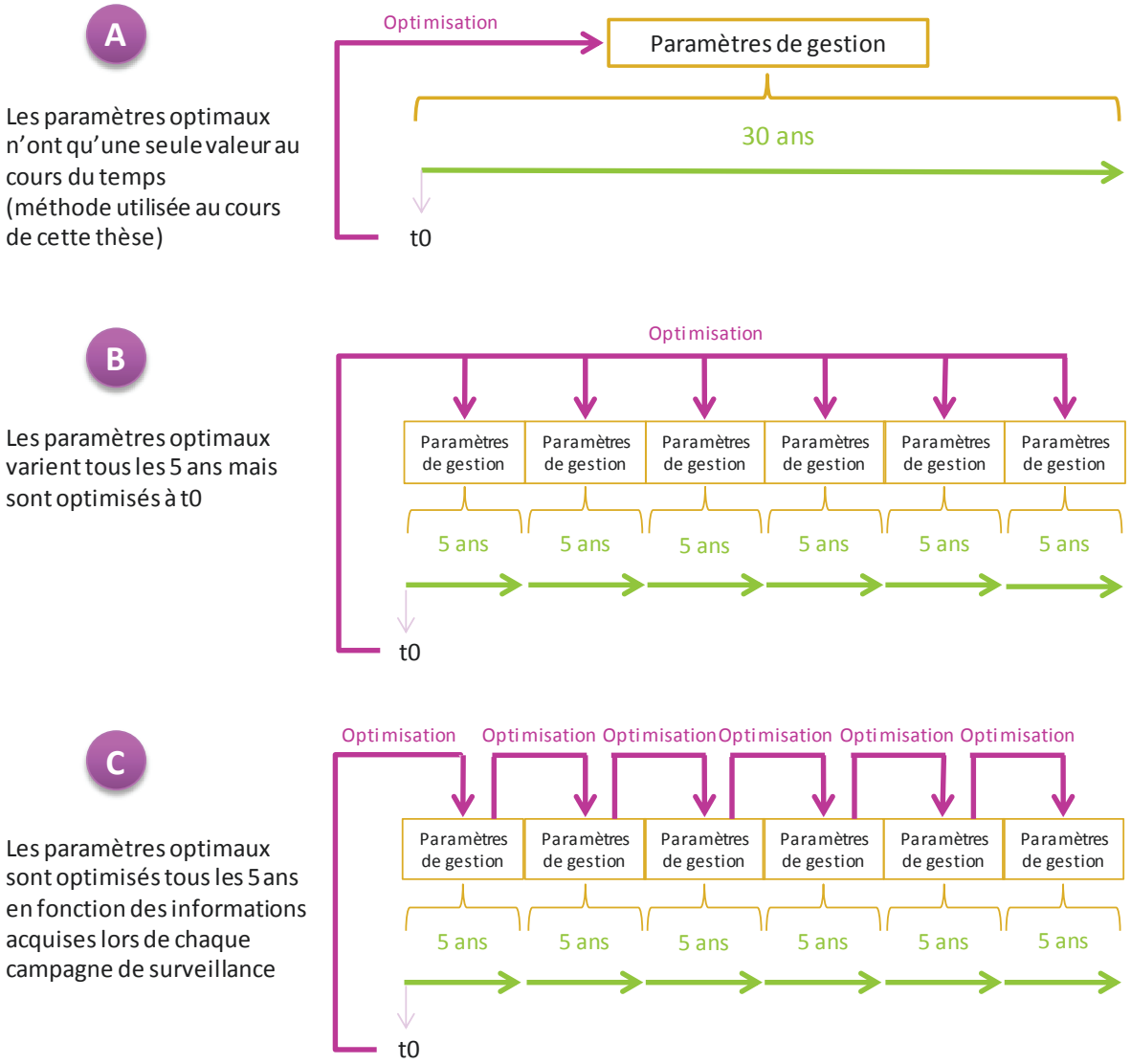


Figure 3 : Méthodes d'optimisation des paramètres de gestion d'une épidémie au cours du temps (exemple avec un pas de temps de 5 ans).

2.2. Quelle place pour les modèles épidémiologiques dans la gestion des maladies ?

Nous avons vu dans la partie précédente que les résultats d'un modèle peuvent être incertains car ils font l'objet de nombreuses hypothèses. Les modèles sont néanmoins de plus en plus utilisés car ils permettent d'apporter des réponses qu'il est difficile (voire impossible) d'obtenir avec des essais sur le terrain, notamment lorsque l'on souhaite analyser l'efficacité de stratégies de gestion des maladies à grande échelle (cf. introduction).

En France, des modèles statistiques sont par exemple utilisés par les membres du réseau en charge de la surveillance biologique du territoire (chambres d'agriculture, instituts techniques, FREDON, coopératives, etc.) pour donner des prévisions de maladies directement aux agriculteurs et à leurs conseillers. Pour chaque filière, ces membres se chargent de rédiger le Bulletin de Santé du Végétal avec des conseils de culture basés sur ces prévisions (Alim'agri 2015, Michel et al. 2016, 2017). Les stratégies de surveillance et de gestion des maladies à l'échelle nationale peuvent également être élaborées à l'aide de modèles avec le soutien et l'expertise de l'Anses (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail) ou des instituts de recherche comme l'INRA. Ainsi, l'Anses a centralisé les données de surveillance de la bactérie *Xylella fastidiosa* à l'échelle nationale (ANSES 2018 ; de Jerphanion et al. 2016 ; Legendre et al. 2017). Ces données ont été utilisées dans des travaux de modélisation menés par l'INRA pour comprendre la dynamique de la bactérie et envisager les conséquences de différents scénarios de surveillance et de lutte (ANSES 2018 ; Soubeyrand et al. 2018). Les résultats obtenus pourraient permettre à la DGAL d'élaborer et de réviser le plan de surveillance (B.O. agri 2017) et le plan d'urgence (B.O. agri 2018) pour contrôler cette bactérie qui affecte de nombreuses plantes cultivées. Un autre exemple est celui de la saisine de l'Anses sur l'évaluation des mesures d'urgence destinées à prévenir la propagation du nématode du pin dans l'Union européenne (Anses 2015). Dans ce cadre, un modèle a été développé pour simuler la dispersion de l'insecte vecteur, la transmission du nématode du pin et enfin la coupe des arbres sur une surface de rayon croissant. Le modèle a notamment montré l'inefficacité de l'arrachage des arbres dans un rayon de 500m autour d'un arbre infecté (qui est pour l'instant appliqué ; EPPO 2009). Ces résultats vont être pris en compte par la DGAL et une nouvelle norme devrait être publiée rapidement. (X. Tassus, communication personnelle). Pour finir, ma thèse (financée par l'INRA et l'Anses), apporte des éléments qui pourraient être pris en compte par la DGAL pour améliorer la gestion de la sharka.

Néanmoins, bien que le système mis en place en France (annexe 1) semble efficace pour gérer de nombreuses maladies, les modèles n'ont pas été largement utilisés pour la gestion des organismes nuisibles de quarantaine pour soutenir la prise de décision en matière de santé des plantes (Suffert et al. 2016). Cela peut s'expliquer par le fait que les résultats des recherches scientifiques basées sur des modèles sont parfois difficiles à interpréter par les gestionnaires. En effet, les gestionnaires peuvent se demander à quel point les hypothèses du modèle peuvent influencer les résultats obtenus. Le choix d'une stratégie de gestion n'est donc pas toujours évident et le processus de réflexion peut être très long car les intérêts des différents acteurs de la filière (arboriculteurs, pépiniéristes, CTIFL, DGAL pour le cas de la sharka) peuvent diverger. De plus, les acteurs impliqués dans ce processus ont chacun leurs compétences ; le dialogue peut donc être difficile.

Pour pallier la difficulté de communication entre scientifiques et gestionnaires du risque, les modèles doivent tenir compte des incertitudes, et celles-ci doivent pouvoir être facilement interprétées. Dans cette thèse, nous avons par exemple choisi de présenter la moyenne des résultats de nos simulations, ainsi que la moyenne du plus faible décile des résultats. Ce dernier critère prend en compte l'aversion au risque des acteurs. Il a été choisi en considérant que les arboriculteurs n'accepteraient pas une stratégie de gestion qui conduise trop souvent à une faible VAN. De rares études ont également proposé des résultats tenant compte du niveau d'aversion au risque des gestionnaires (Cunniffe et al. 2015b, 2016), mais la plupart présentent des résultats basés sur la moyenne. Pour certains modèles, d'autres critères d'incertitude pourraient être développés à l'instar de ceux présentés par les modèles météorologiques : par exemple, Météo-France communique les prévisions estimées les plus probables parmi différents scénarios, et fournit en complément un indice de confiance pour les prévisions à partir du quatrième jour (Météo France, 2018). De plus, pour rendre les modèles accessibles, certaines études ont développé des plateformes interactives sur internet (Cunniffe et al. 2015b). Néanmoins, les scientifiques peuvent être réticents à développer ce genre d'interfaces, qui peuvent être mal utilisées, mal interprétées et ensuite blâmées pour les mauvaises prédictions dans leur application. L'implication des gestionnaires du risque dans la construction des modèles pourrait alors être une option intéressante, en les impliquant par exemple aux moments clés des programmes de recherche.

Pour résumer, les modèles pourraient être mieux valorisés en favorisant les échanges entre gestionnaires du risque et scientifiques et en consolidant le lien entre tous les acteurs de la gestion des maladies, comme cela se fait déjà lors de conférences, de réunions de réseaux ou via des partenariats (DigitAg, 2018; EFSA, 2018).

Annexe 1

Les organismes en charge de l'élaboration et de l'application des stratégies de gestion contre les épidémies

En France, la gestion des épidémies se fait tout d'abord au niveau national par le biais du ministère de l'agriculture et de l'alimentation. Le ministère est divisé en 3 directions dont la *Direction générale de l'alimentation* (DGAL 2014). « La DGAL veille à la sécurité et à la qualité des aliments à tous les niveaux de la chaîne alimentaire, ainsi qu'à la santé et à la protection des animaux et des végétaux ». La DGAL élabore le dispositif juridique correspondant à ses missions et en contrôle l'application. Au niveau international, elle assure la promotion des modèles alimentaire, sanitaire et phytosanitaire français (Alim'agri 2018b). Pour atteindre ses objectifs, la DGAL travaille en coordination avec les services de l'État dans les régions et départements avec les différents acteurs concernés : professionnels du monde agricole, associations, consommateurs, etc. » (Alim'agri 2018b).

Pour mettre en œuvre ses actions sur le terrain, la DGAL s'appuie sur les services déconcentrés de l'État (DGAL 2014) comme les directions régionales de l'alimentation, de l'agriculture et de la forêt (DRAAF). Au sein de la DRAAF, les services régionaux de l'alimentation (SRAL) mettent en œuvre la politique de la DGAL (DRAAF PACA 2015). En particulier, les SRAL participent à la coordination des plans d'intervention sanitaire d'urgence départementaux, font appliquer la règlementation relative à la surveillance biologique du territoire et au maintien du bon état sanitaire des végétaux, et veillent à la mise en place de l'ensemble des dispositifs régionaux de surveillance. La surveillance est coordonnée et/ou réalisée par la FREDON (Fédération régionale de défense contre les organismes nuisibles) en verger et en pépinières (FREDON 2018).

Afin d'améliorer le dispositif de gestion des maladies, l'ensemble des organismes présentés ci-dessus travaillent en collaboration avec des organismes indépendants tels que l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail). L'ANSES fournit un appui scientifique et technique pour l'élaboration des plans de surveillance et contribue ainsi à la qualité de la surveillance réalisée par les services de l'État. Elle précise également les modalités les plus adaptées pour assurer la pertinence et la fiabilité des données collectées. L'ANSES collabore également avec des organismes comme l'INRA qui réalise des recherches finalisées en épidémiologie.

Références bibliographiques

Alim'agri (2015) Le BSV, pour quoi ? Pour qui ? http://agriculture.gouv.fr/le-bsv-pour-quoi-pour-qui

Alim'agri (2018a) Le plan Écophyto, pour réduire l'utilisation des produits phytosanitaires en France. http://agriculture.gouv.fr/le-plan-ecophyto-pour-reduire-lutilisation-des-produits-phytosanitaires-en-france

Alim'agri. (2018b) DGAL | Alim'agri. http://agriculture.gouv.fr/mots-cl%C3%A9s/dgal

Allen R. (1978) Epidemiological factors influencing the success of roguing for the control of bunchy top disease of bananas in New South Wales. Aust. J. Agric. Res. 29(3):535–544.

Anses (2015) Évaluation des mesures d'urgence destinées à prévenir la propagation du nématode du pin dans l'Union européenne. Avis de l'Anses. Rapport d'expertise collective. https://www.anses.fr/fr/system/files/SVEG2014sa0103Ra.pdf

ANSES (2018) Qui est Xylella fastidiosa ? | Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail. https://www.anses.fr/fr/content/qui-est-xylella-fastidiosa.

Barnier J.B., Grafton-Cardwell E. and Polek M. (2010) Citrus tristeza virus (CTV): then and now. Citrograph. 6:16-23.

te Beest D.E., Hagenaars T.J., Stegeman J.A., Koopmans M.P. and van Boven, M. (2011) Risk based culling for highly infectious diseases of livestock. Vet. Res. 42(1):81.

Bertschinger L. (1997) *Modeling plant virus disease epidemics: development and use of simulation models*. International Potato Center.

B.O. agri (2017) Plan de surveillance pluriannuel national de Xylella fastidiosa. https://info.agriculture.gouv.fr/gedei/site/bo-agri/instruction-2017-653

B.O. agri (2018) Plan national d'intervention sanitaire d'urgence Xylella fastidiosa. https://info.agriculture.gouv.fr/gedei/site/bo-agri/instruction-2017-653

Brunker K., Lemey P., Marston D.A., Fooks A.R., Lugelo A., Ngeleja C., Hampson K. and Biek, R. (2018) Landscape attributes governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs. Mol. Ecol. 27(3):773-788.

Cambra M., Capote N., Myrta A. and Llácer G. (2006) *Plum pox virus* and the estimated costs associated with sharka disease. EPPO Bull. 36(2)202–204.

Campbell C.L. and Maden L.V. (1990) *Introduction to plant disease epidemiology*. John Wiley Sons.

Chan M.-S. and Jeger M.J. (1994) An analytical model of plant virus disease dynamics with roguing and replanting. J. Appl. Ecol. 31:413–427.

Courcoul A., Monod H., Nielen M., Klinkenberg D., Hogerwerf L., Beaudeau F. and Vergu, E. (2011) Modelling the effect of heterogeneity of shedding on the within herd Coxiella burnetii spread and identification of key parameters by sensitivity analysis. J. Theor. Biol. 284(1):130–141.

Coutts S.R., van Klinken R.D., Yokomizo H. and Buckley Y.M. (2010) What are the key drivers of spread in invasive plants: dispersal, demography or landscape: and how can we use this knowledge to aid management? Biol. Invasions 13(7):1649–1661.

Cunniffe N.J., Laranjeira F.F., Neri F.M., DeSimone R.E. and Gilligan C.A. (2014) Cost-effective control of plant disease when epidemiological knowledge is incomplete: modelling Bahia Bark Scaling of Citrus. PLOS Comput Biol 10(8):e1003753.

Cunniffe N.J., Koskella B., Metcalf C.J., Parnell S., Gottwald T.R. and Gilligan C.A. (2015a) Thirteen challenges in modelling plant diseases. Epidemics 10, 6–10.

Cunniffe N.J., Stutt R.O.J.H., DeSimone R.E., Gottwald T.R. and Gilligan C.A. (2015b) Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. PLOS Comput Biol 11(4):e1004211.

Cunniffe N.J., Cobb R.C., Meentemeyer R.K., Rizzo D.M. and Gilligan C.A. (2016) Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc. Natl. Acad. Sci. 113(20):5640–5645.

Dallot S., Gottwald T., Labonne G. and Quiot J.-B. (2003) Spatial pattern analysis of sharka disease (Plum pox virus strain M) in peach orchards of southern France. Phytopathology 93(12):1543–1552.

de Jerphanion P., Joudar S., Poliakoff F., Olivier V., Legendre B., Raynaud P., Hendrikx P., Bronner A., Grosman J., Jullien J., Chauvel G., Bourgouin B., Flot J.L., Poirier A., Fajardi V. and Manceau C. (2016) *Xylella fastidiosa* bilan de la surveillance en 2015. Phytoma 694:43–48.

DGAL (2014) La DGAL en bref. http://agriculture.gouv.fr/sites/minagri/files/dgal-rapdac-2014-eb-fr-web_cle4bb1ecenbref.pdf.

DigitAg (2018) #DigitAg Institut Convergences Agriculture Numérique. http://www.hdigitag.fr/fr/

Doll R. and Hill A.B. (1964) Mortality in relation to smoking: ten years' observations of british doctors. Br. Med. J. 1(5395):1399–1410.

DRAAF PACA (2015) SRAL - Service Régional de l'alimentation . http://draaf.paca.agriculture.gouv.fr/SRAL-Service-Regional-de-l

Dzahini-Obiatey H., Ameyaw G.A. and Ollennu L.A. (2006) Control of cocoa swollen shoot disease by eradicating infected trees in Ghana: A survey of treated and replanted areas. Crop Prot. 25(7):647–652.

EFSA (2018) Joint EFSA-EPPO Workshop: Modelling in Plant Health – how can models support risk assessment of plant pests and decision-making? https://www.efsa.europa.eu/fr/events/event/161212

Epanchin-Niell R.S., Haight R.G., Berec L., Kean J.M. and Liebhold A.M. (2012) Optimal surveillance and eradication of invasive species in heterogeneous landscapes. Ecol. Lett. 15(8):803–812.

Epanchin-Niell R.S., Brockerhoff E.G., Kean J.M. and Turner J.A. (2014) Designing cost-efficient surveillance for early detection and control of multiple biological invaders. Ecol. Appl. 24(6):1258–1274.

EPPO. (2009) *Bursaphelenchus xylophilus* and its vectors: procedures for official control. EPPO Bulletin, 39(3) :454-459.

FAO (2017) Glossaire des termes phytosanitaires. http://www.fao.org/fileadmin/user_upload/faoterm/PDF/ISPM_05_2016_Fr_2017-04-26_PostCPM12_InkAm.pdf

Fauquet C. and Fargette D. (1990) African cassava mosaic virus: etiology, epidemiology and control. Plant Dis. 74(6):404-411.

Filipe J.A.N., Cobb R.C., Meentemeyer R.K., Lee C.A., Valachovic Y.S., Cook A.R., Rizzo D.M. and Gilligan C.A. (2012) Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden Oak Death in Northern Californian forests. PLOS Comput Biol 8(1):e1002328.

Fossati D., Kellerhals M., Spring J.L. and Mascher F. (2006) Sélection de variétés pour une agriculture durable : trois exemples suisses. Le courrier de l'environnement de l'INRA. (30):91-100.

FREDON (2018). FREDON Occitanie. http://www.fredonoccitanie.com/

Galus C. (2008) Une épidémie de rouille noire, redoutable parasite du blé, menace les grands pays céréaliers d'Asie. Le Monde.fr. http://www.lemonde.fr/planete/article/2008/04/21/une-epidemie-de-rouille-noire-redoutable-parasite-du-ble-menace-les-grands-pays-cerealiers-d-asie_1036464_3244.html

Gibson G.J. (1997) Investigating mechanisms of spatiotemporal epidemic spread using stochastic models. Phytopathology 87(2):139–146.

Gibson R.W. and Aritua V. (2002) The perspective of sweetpotato chlorotic stunt virus in sweetpotato production in Africa: a review. Afr. Crop Sci. J. 10:281–310.

Gibson G.J., Gilligan C.A. and Kleczkowski A. (1999) Predicting variability in biological control of a plant—pathogen system using stochastic models. Proc. R. Soc. Lond. B Biol. Sci. 266(1460):1743–1753.

Gibson R.W., Aritua V., Byamukama E., Mpembe I. and Kayongo J. (2004) Control strategies for sweet potato virus disease in Africa. Virus Res. 100(1):115–122.

Gilligan C.A. (1985) *Mathematical modelling of crop disease*. Academic Press Inc.

Gilligan C.A. (1990) Comparison of disease progress curves. New Phytol. *115*(2):223–242.

Gilligan C.A. (2002) An epidemiological framework for disease management. A. in B. Research. *38*(1):64.

Gilligan C.A. and Kleczkowski A. (1997) Population dynamics of botanical epidemics involving primary and secondary infection. Philos. Trans. R. Soc. B Biol. Sci. *352*(1353):591–608.

Gilligan C.A. and Van Den Bosch F. (2008) Epidemiological models for invasion and persistence of pathogens. Annu. Rev. Phytopathol. 46:385–418.

Gottwald T.R., Avinent L., Llácer G., de Mendoza A.H. and Cambra M. (1995) Analysis of the spatial spread of sharka (plum pox virus) in apricot and peach orchards in eastern Spain. Plant Dis. 79(3):266–278.

Gottwald T.R., Gibson G.J., Garnsey S.M. and Irey M. (1999) Examination of the Effect of Aphid Vector Population Composition on the Spatial Dynamics of Citrus Tristeza Virus Spread by Stochastic Modeling. Phytopathology. 89(7):603–608.

Gottwald T.R., Wierenga E., Luo W. and Parnell S. (2013) Epidemiology of Plum pox 'D' strain in Canada and the USA. Can. J. Plant Pathol. 35(4):442–457.

Grafton-Cardwell E.E., Stelinski L.L. and Stansly P.A. (2013) Biology and Management of Asian Citrus Psyllid, Vector of the Huanglongbing Pathogens. Annu. Rev. Entomol. 58:413–432.

Gubbins S. and Gilligan C.A. (1999) Invasion thresholds for fungicide resistance: deterministic and stochastic analyses. Proc. R. Soc. Lond. B Biol. Sci. 266(1437):2539–2549.

Hall R.J., Gubbins S. and Gilligan C.A. (2004) Invasion of drug and pesticide resistance is determined by a trade-off between treatment efficacy and relative fitness. Bull. Math. Biol. *66*(4):825–840.

Holt J. and Chancellor T.C.B. (1996) Simulation Modelling of the Spread of Rice Tungro Virus Disease: The Potential for Management by Roguing. J. Appl. Ecol. 33:927–936.

Holt J., Colvin J. and Muniyappa V. (1999) Identifying control strategies for tomato leaf curl virus disease using an epidemiological model. J. Appl. Ecol. 36(5):625–633.

Hyatt-Twynam S.R., Parnell S., Stutt R.O.J.H., Gottwald T.R., Gilligan C.A. and Cunniffe N.J. (2017) Risk-based management of invading plant disease. New Phytol. 214(3):1317–1329.

Irwin M.E., Kampmeier G.E. and Jeger M.J. (1989) Vector behaviour, environmental stimuli, and the dynamics of plant virus epidemics. Spat. Compon. Plant Dis. Epidemics 14–39.

Jeger M.J. (1989) *Spatial components of plant disease epidemics*. Englewood Cliffs, USA: Prentice Hall.

Jeger M.J. and Chan M.S. (1995) Theoretical aspects of epidemics: uses of analytical models to make strategic management decisions. Can. J. Plant Pathol. 17(2):109–114.

Jenicek M. and Clèroux R. (1982) *Épidémiologie: principes: techniques: applications.* EDISEM / MALOINE.

JORF (2011) Arrêté du 17 mars 2011 relatif à la lutte contre le Pum Pox Virus, agent causal de la maladie de la Sharka, sur les végétaux sensibles du genre Prunus. JORF n°0067 du 20 mars 2011. République Française. Paris, France. NOR: AGRG1105295A.
http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000023728913&dateTexte=&categorieLien=id.

Kellerhals M. (2004) En collaboration avec Agroscope FAW Wadenswil et l'Institut de recherches en agriculture biologique Portrait des varietes de pommes resistantes a la tavelure. Rev. Suisse Vitic. Arboric. Hortic. 36(1):29–36.

Kermack W. and Mckendrick A. (1927) A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Math. Phys. Eng. Sci. 115:700–721.

Kleczkowski A., Bailey D.J. and Gilligan C.A. (1996) Dynamically generated variability in plant-pathogen systems with biological control. Proc. R. Soc. Lond. B Biol. Sci. 263(1371):777–783.

Kranz J. (2012). *Epidemics of Plant Diseases: Mathematical Analysis and Modeling*. Springer Science & Business Media.

Last J. (2001) *A dictionary of epidemiology*. New York: Oxford University Press.

Legendre B., Denancé N. and Olivier V. (2017) Xylella fastidiosa: the status of the infection and control measures in France. Xylella Fastidiosa Olive Quick Decline Syndr. OQDS P 33.

Llácer G. and Cambra M. (2006) Hosts and symptoms of Plum pox virus: fruiting Prunus species. EPPO Bull. 36(2):219–221.

Lurette A., Touzeau S., Lamboni M. and Monod H. (2009) Sensitivity analysis to identify key parameters influencing Salmonella infection dynamics in a pig batch. J. Theor. Biol. 258(1):43–52.

MacMahon B., Pugh T.F. and Ipsen J. (1960) *Epidemiologie Methods*. Epidemiologie Methods.

Madden L.V., Reynolds K.M., Pirone T.P. and Raccah B. (1988) Modeling of tobacco virus epidemics as spatio-temporal autoregressive integrated moving-average processes. Phytopathology 78(10):1361–1366.

Mariette S., Wong Jun Tai F., Roch G., Barre A., Chague A., Decroocq S., Groppi A., Laizet, Y., Lambert P., Tricon D., et al. (2016) Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (Prunus armeniaca). New Phytol. 209(2):773–784.

Martins D.S., Ventura J.A., Lima R., Culik M.P., Costa H. and Ferreira P.S.F. (2012) Interaction between Papaya meleira virus (PMeV) infection of papaya plants and Mediterranean fruit fly infestation of fruits. Crop Prot. 36:7–10.

May R.M. and Anderson R.M. (1979) Population biology of infectious diseases: Part II. Nature 280(5722) :455–461.

McLean G.D., Garret R.G. and Ruesink W.G. (1986) *Plant virus epidemics: monitoring, modelling and predicting outbreaks*. Academic Press.

Meentemeyer R.K., Cunniffe N.J., Cook A.R., Filipe J.A.N., Hunter R.D., Rizzo D.M. and Gilligan C.A. (2011) Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). Ecosphere 2(2):1–24.

Merrill, R. M. (2012) *Introduction to epidemiology*. Jones & Bartlett Publishers.

Météo France (2018) Prévision probabiliste - Météo France. http://www.meteofrance.com/prevision-probabiliste

Michel L., Brun F., Piraux F. and Makowski D. (2016) Estimating the incidence of Septoria leaf blotch in wheat crops from in-season field measurements. European journal of plant pathology, 146(1) :17-35.

Michel L., Brun F. and Makowski D. (2017) A framework based on generalised linear mixed models for analysing pest and disease surveys. Crop Protection. 94:1-12.

Mollentze N., Nel L.H., Townsend S., Roux K., Hampson K., Haydon D.T. and Soubeyrand, S. (2014). A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proc. R. Soc. Lond. B Biol. Sci. 281(1782):20133251.

Morelli M.J., Thébaud G., Chadœuf J., King D.P., Haydon D.T. and Soubeyrand S. (2012) A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. PLoS Comput Biol 8(11):e1002768.

Németh M. (1986) *Virus, mycoplasma and Rickettsia diseases of fruit trees.* M. Nijhoff/Dr. W. Junk.

Ostfeld R.S., Glass G.E. and Keesing F. (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol. Evol. 20(6):328–336.

Parnell S., Gottwald T.R., van den Bosch F. and Gilligan C.A. (2009) Optimal strategies for the eradication of Asiatic Citrus Canker in heterogeneous host landscapes. Phytopathology 99(12):1370–1376.

Parnell S., Gottwald T.R., Gilligan C.A., Cunniffe N.J. and van den Bosch F. (2010) The effect of landscape pattern on the optimal eradication zone of an invading epidemic. Phytopathology 100(7):638–644.

Parnell S., Gottwald T.R., Gilks W.R. and van den Bosch F. (2012) Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. J. Theor. Biol. 305:30–36.

Parnell S., Gottwald T.R., Riley T. and van den Bosch F. (2014) A generic risk-based surveying method for invading plant pathogens. Ecol. Appl. 24(4):779–790.

Pavé A. (1994). *Modélisation en biologie et en écologie*. Aléas.

Perry R.D. and Fetherston J.D. (1997) Yersinia pestis--etiologic agent of plague. Clin. Microbiol. Rev. 10(1):35–66.

Pethybridge S.J. and Madden L.V. (2003) Analysis of spatiotemporal dynamics of virus spread in an Australian hop garden by stochastic modeling. Plant Dis. 87(1):56–62.

Picard C., Rimbaud L., Hendrikx P., Soubeyrand S., Jacquot E. and Thébaud G. (2017) PESO: a modelling framework to help improve management strategies for epidemics – application to sharka. EPPO Bull. 47(2):231-236.

Pleydell D.R.J., Soubeyrand S., Dallot S., Labonne G., Chadœuf J., Jacquot E. and Thébaud G. (2018) Estimation of the dispersal distances of an aphid-borne virus in a patchy landscape. PLOS Comput. Biol. 14(4):e1006085.

Rimbaud, L., Dallot, S., Delaunay, A., Borron, S., Soubeyrand, S., Thébaud, G., & Jacquot, E. (2015a). Assessing the mismatch between incubation and latent periods for vector-borne diseases: The case of sharka. *Phytopathology*, *105*(11), 1408-1416.

Rimbaud L., Dallot S., Gottwald T., Decroocq V., Jacquot E., Soubeyrand S. and Thébaud G. (2015b) Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. Annu. Rev. Phytopathol. 53:357–378.

Rimbaud L., Dallot S., Bruchou C., Thoyer S., Jacquot E., Soubeyrand S. and Thébaud G. (2018a) Heuristic optimisation of the management strategy of a plant epidemic using sequential sensitivity analyses. BioRxiv 315747.

Rimbaud L., Bruchou C., Dallot S., Pleydell D.R.J., Jacquot E., Soubeyrand S. and Thébaud G. (2018b) Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. R. Soc. Open Sci. *5*(1):171435.

Robinet C., Kehlenbeck H., Kriticos D.J., Baker R.H.A., Battisti A., Brunel S., Dupin M., Eyre D., Faccoli M., Ilieva Z. et al. (2012) A suite of models to support the quantitative assessment of spread in pest risk analysis. PLOS ONE. *7*(10):e43366.

Rosegrant M.W. and Cline S.A. (2003) Global Food Security: Challenges and Policies. Science 302(5652):1917–1919.

Ruesink, W. G. (1986) Soybean mosaic virus epidemiology: a model and some implications. Plant virus epidemics: Monitoring, modelling, and predicting outbreaks, 295-313.

Scorza R., Ravelonandro M., Callahan A.M., Cordts J.M., Fuchs M., Dunez J. and Gonsalves D. (1994) Transgenic plums (Prunus domestica L.) express the plum pox virus coat protein gene. Plant Cell Rep. 14(1):18–22.

Shaw M.W. (1994) Modeling stochastic processes in plant pathology. Annu. Rev. Phytopathol. 32(1):523–544.

Sisterson M.S. and Stenger D.C. (2012) Roguing with replacement in perennial crops: conditions for successful disease management. Phytopathology 103(2):117–128.

Snow, J. (1855) *On the mode of communication of cholera*. John Churchill.

Soubeyrand S. (2016) Construction of semi-Markov genetic-space-time SEIR models and inference. J. Société Fr. Stat. 157(1):129–152.

Soubeyrand S., de Jerphanion P., Martin O., Saussac M., Manceau C., Hendrikx P. and Lannou, C. (2018) Inferring pathogen dynamics from temporal count data: the emergence of Xylella fastidiosa in France is probably not recent. New Phytologist. 219(2):824-836.

Strange R.N. and Scott P.R. (2005) Plant disease: a threat to global food security. Annu. Rev. Phytopathol. 43:83–116.

Suffert M., Petter F. and Ward M. (2016) Summary and conclusions of the Joint EFSA–EPPO Workshop: Modelling in Plant Health – how can models support risk assessment of plant pests and decision-making? Parma, Italy, 12–14 December 2016 EFSA–EPPO Modelling Workshop. EPPO Bulletin, 47(2):211-212.

Taylor R.A., Mordecai E.A., Gilligan C.A., Rohr J.R. and Johnson L.R. (2016) Mathematical models are a powerful method to understand and control the spread of Huanglongbing. PeerJ. 4:e2642.

Thresh J.M. and Owusu G.K. (1986) The control of cocoa swollen shoot disease in Ghana: an evaluation of eradication procedures. Crop Prot. 5(1):41–52.

UNAIDS (2015) UNAIDS. Global Fact Sheet.
http://www.unaids.org/sites/default/files/media_asset/AIDS_by_the_numbers_2015_fr.pdf.

Van Der Plank J.E.V. (1963) *Plant diseases: epidemics and control*. Academic Press.

Vanhaute E., Paping R. and Ó Gràda C. (2006) The European Subsistence Crisis of 1845–1850: A Comparative Perspective. Working Paper Series, UCD Centre for Economic Research.

Zadoks J.C. (1971) Systems analysis and the dynamics of epidemics. Phytopathology 600-610.

Liste des publications, présentations et posters

Publications

Publiées

Picard C.*, Dallot S.*, Brunker K., Berthier K., Roumagnac P., Soubeyrand S., Jacquot E. and Thébaud G. (2017) Exploiting genetic information to trace plant virus dispersal in landscapes. *Annual Review of Phytopathology* 55

Picard C., Rimbaud L., Hendrikx P., Soubeyrand S., Jacquot E. and Thébaud G. (2017) PESO: a modelling framework to help improve management strategies for epidemics – application to sharka. *EPPO Bulletin. 47*(2), 231-236.

Acceptées

Picard C., Soubeyrand S., Jacquot E. and Thébaud G. Analyzing the influence of landscape aggregation on disease spread to improve management strategies. *In revision (Phytopathology).*

En préparation

Picheny V., Picard C. and Thébaud G. Impact of input warping on the Bayesian optimisation of the management of a plant disease using a complex epidemiological model.

Picard C., Picheny V., Bonnot F., Soubeyrand S. and Thébaud G. *In silico* optimization of a strategy for landscape-wide plant disease management.

Picard C., Picheny V., Soubeyrand S. and Thébaud G. Optimization of the spatio-temporal deployment of resistant cultivars and disease control options.

Picard C., Dallot S., Jacquot E., Thébaud G. and Soubeyrand S. Accounting for uninfected hosts in transmission tree reconstruction.

Communications orales

03/2018 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Optimisation in silico de la gestion d'une épidémie chez les plantes à l'échelle du paysage. *Réunion annuelle du réseau ModStatSAP*. Paris, France, 19 mars 2018.

01/2018 Picard C., Soubeyrand S., Jacquot E., Thébaud G. In silico optimisation of sharka management. *Réunion annuelle du projet E-SPACE.* Montpellier, France, 25 janvier 2018.

11/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Optimisation in silico de la gestion d'une maladie des plantes à l'échelle du paysage. *Réunion annuelle du réseau MEXICO*. Montpellier, France, 17 novembre 2017.

10/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Optimisation in silico de la gestion d'une épidémie chez les plantes à l'échelle du paysage. *Journées scientifiques et doctorales de l'ANSES.* Maisons-Alfort, France, 25 octobre 2017.

09/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. In silico optimisation of plant disease management at the landscape scale. *Séminaire ITK*. Clapiers, France, 21 septembre 2017.

09/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. In silico optimisation of plant disease management at the landscape scale. *Séminaire BGPI*. Montpellier, France, 14 septembre 2017.

05/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Optimisation in silico de la gestion d'une maladie des plantes à l'échelle du paysage. *Réunion annuelle du réseau EcoStat*. Nantes, France, 22 mai 2017.

01/2017 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Influence of landscape characteristics on the optimal control of a plant virus. *16èmes rencontres de virologie végétale (RVV2017).* Aussois, France, 15-19 janvier 2017. **[Prix de la meilleure présentation de doctorant]**

07/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Influence des caractéristiques paysagères sur la dispersion et la gestion d'une épidémie chez les plantes*. Réunion annuelle du projet E-SPACE.* Montpellier, France, 8 juillet 2016.

06/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Influence of landscape characteristics on plant disease control*. 8ème journée des doctorants SPE.* Toulouse, France, 29 juin - 1er juillet 2016.

06/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Impact du paysage sur la gestion d'une épidémie chez les plantes. *Séminaire BGPI*. Montpellier, France, 23 juin 2016.

06/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Influence des caractéristiques paysagères sur la dispersion et la gestion d'une épidémie chez les plantes. *Printemps de Baillarguet*. Montpellier, France, 2 juin 2016.

01/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Optimisation in silico de la gestion spatialisée d'une épidémie chez les plantes. *Réunion annuelle du réseau ModStatSAP*. Nantes, France, 27 janvier 2016.

Posters

12/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Analyzing the influence of landscape characteristics on disease spread and management strategies. *Joint EFSA-EPPO Workshop: Modelling in Plant Health – how can models support risk assessment of plant pests and decision-making?* Parme, Italie, 12-13 décembre 2016.

05/2016 Picard C., Soubeyrand S., Jacquot E., Thébaud G. Analyzing the influence of landscape characteristics on disease spread and management strategies. *13th International Plant Virus Epidemiology Symposium (IPVE).* Avignon, France, 6-10 juin 2016. **[2ème prix du meilleur poster de doctorant]**

Abstract

How to identify effective strategies to manage the epidemics impacting crops?

The answer to this question is not obvious because of the complexity of epidemics (they simultaneously depend on biological processes, patch organization in the landscape, and human interventions). To answer that question, models can be used because of their ability to simulate many scenarios. Indeed, they can enable to estimate epidemiological parameters, to assess the efficiency of different management strategies and to optimize them.

In this context, we developed a generic framework for *in silico* optimization of plant disease management strategies at the landscape scale. This framework is based on (i) the characterization of the epidemic dynamics to estimate its key parameters, (ii) the use of a spatially explicit model to simulate the dynamics and management of this disease and (iii) the use of numerical optimization methods to identify efficient management strategies. Here, we apply this approach to sharka, caused by *Plum pox virus*. This disease has a worldwide impact on the *Prunus* industry and is associated with huge management costs in many countries. In France, the management strategy for this disease principally consists of orchard surveillance, removal of trees that might contribute to the epidemic propagation, and plantation bans.

My PhD indicates that some strategies should be more profitable than the one currently applied in France. These strategies require less orchard surveillance and do not impose any plantation ban. It turns out that some of these strategies are efficient for all tested landscape types, which may be interesting in practice for risk managers. This approach was applied to the sharka, but could be transposed to optimize the management of other epidemics.

Keywords: epidemiology, management strategies, modeling, optimization, landscape.

Résumé

Comment identifier des stratégies efficaces pour gérer les épidémies affectant les cultures ?

La réponse à cette question n'est pas évidente du fait de la complexité des épidémies (elles dépendent à la fois de processus biologiques, de l'organisation des parcelles dans le paysage, et des interventions humaines). Pour y répondre, des modèles peuvent être utilisés en raison de leur capacité à simuler de nombreux scénarios. En effet, ils peuvent permettre d'estimer des paramètres épidémiologiques, d'évaluer l'efficacité de différentes stratégies de gestion et de les optimiser.

Dans ce contexte, nous avons développé une approche générique pour optimiser *in silico* la gestion d'une maladie des plantes à l'échelle du paysage. Cette approche est basée sur (i) la caractérisation de la dynamique épidémique pour estimer ses paramètres clés, (ii) l'utilisation d'un modèle spatialement explicite pour simuler la dynamique et la gestion de cette maladie,et (iii) l'utilisation de méthodes d'optimisation numérique pour identifier des stratégies de gestion efficaces. Ici, nous appliquons cette approche à la sharka, causée par le *Plum pox virus*. Cette maladie a un impact mondial sur la filière *Prunus* et est associée à d'importants coûts de gestion dans de nombreux pays. En France, la stratégie de gestion de cette maladie repose notamment sur la surveillance des vergers, l'arrachage des arbres pouvant contribuer à la propagation de l'épidémie, et des restrictions de plantation.

Ma thèse indique que certaines stratégies devraient être plus rentables que celle actuellement appliquée en France. Ces stratégies nécessitent moins de surveillance des vergers et n'imposent aucune restriction de plantation. Il s'avère que certaines de ces stratégies sont efficaces pour tous les types de paysage testés, ce qui peut être intéressant en pratique pour les gestionnaires du risque. Cette approche a été appliquée à la sharka, mais elle pourrait être transposée pour optimiser la gestion d'autres épidémies.

Mots clés : épidémiologie, stratégies de gestion, modélisation, optimisation, paysage.