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

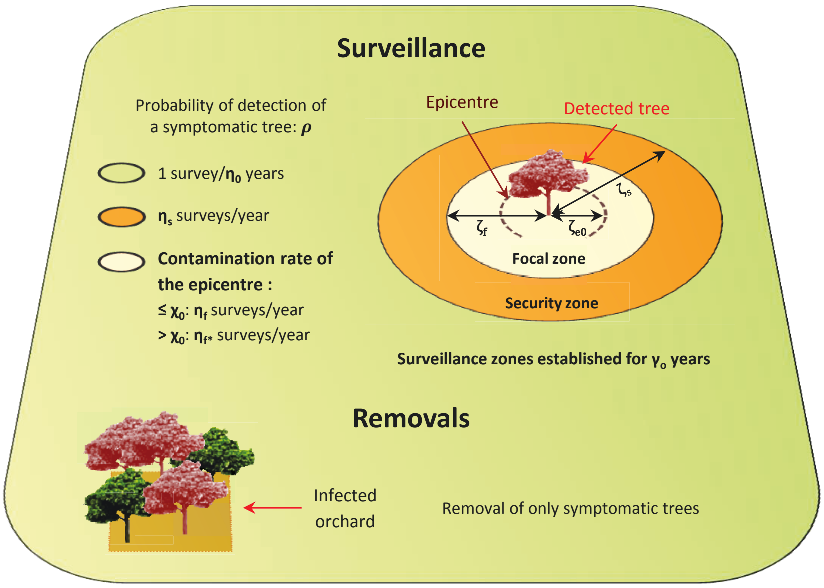


Fig. 1 Management actions implemented in the model.

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 γ_O = 0 or when ρ = 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 $\mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+N}$ 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 $f_{\min} = \min_{i=1,\dots,n} y_i$, 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 $\sigma^n(x)$):

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

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 $\mu(\mathbf{x}_1), \dots, \mu(\mathbf{x}_n)$ 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 $\mu(\mathbf{x}_1), \dots, \mu(\mathbf{x}_n)$ 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):

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

$$\mathbb{X} = \{\mathbb{X}_i, \mathbb{X}_J, \mathbb{X}_{-iJ}\} \tag{5}$$

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

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

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 $\{1,\dots,d\} \setminus I$) if there exists at least one $i \in I$ such that $x_i = c_i$ (I a subset of $\{1,\dots,d\} \setminus J$):

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

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 $\{1,\dots,d\} \setminus I$) if $\mathbf{A}\mathbf{x}_I = \mathbf{b}$, with I a subset of $\{1,\dots,d\} \setminus J$), $\mathbf{A}$ 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 $\widetilde{\mathbb{X}}$ and index the GP on $\widetilde{\mathbb{X}}$, 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:

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

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:

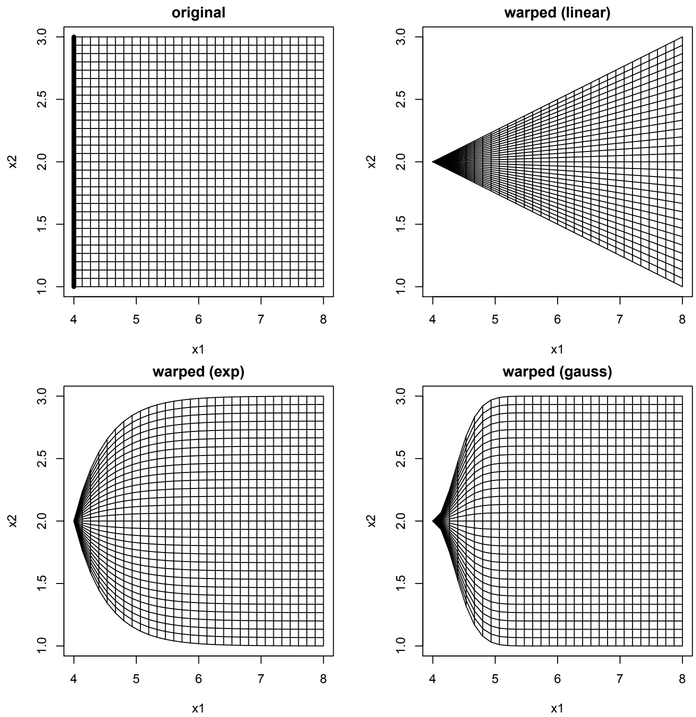


Fig. 2 Three deformations of a 2D space. The local invariance is at $x_1 = 0$, highlighted with larger lines.

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:

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

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 $J = \{1, 2\}$);

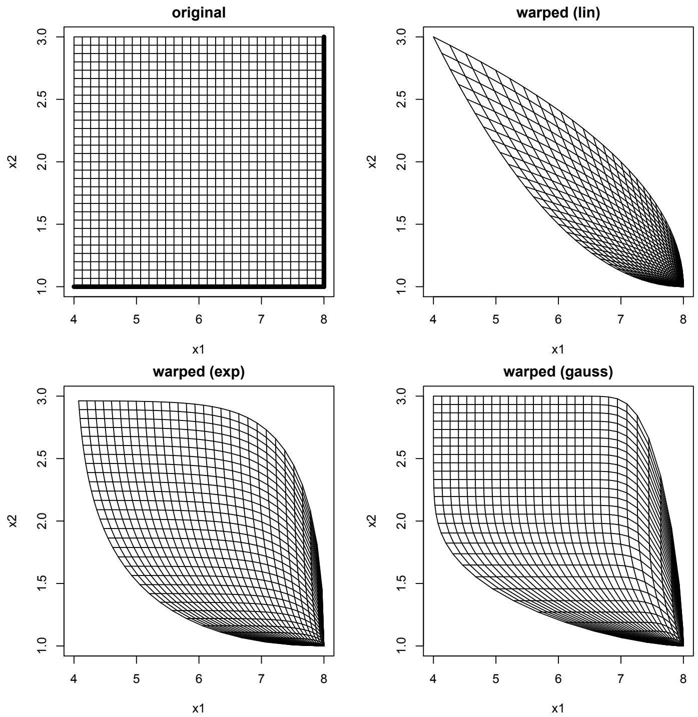


Fig. 3 Three deformations of a 2D space, with invariance at $x_1 = 8$ OR $x_2 = 1$, highlighted with larger lines.

- 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.

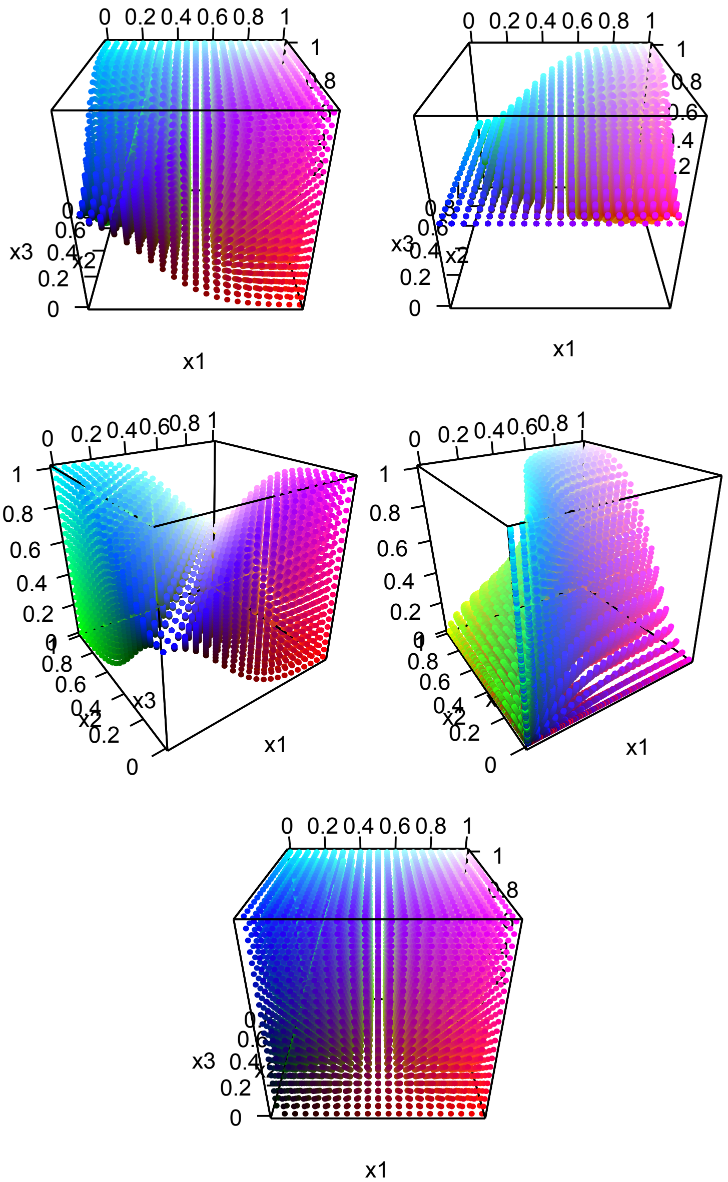


Fig. 4 Four deformations of the unit cube under different invariances: AND (top left), OR (top right), LINEAR (middle left), CIRCULAR (middle right). The bottom figure shows the orginal space.

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.

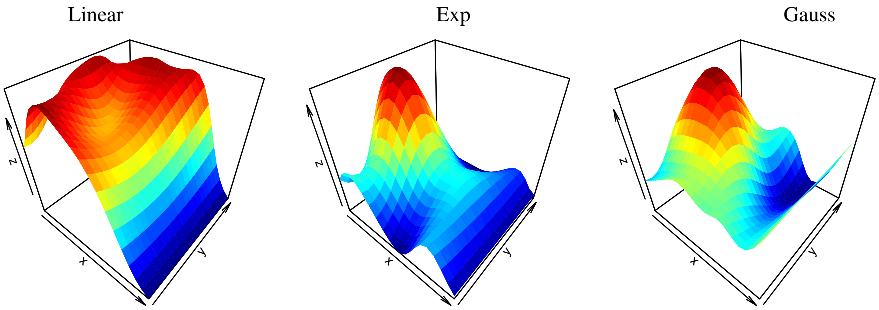


Fig. 5 Three GP realizations using warping functions as shown previously.

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]

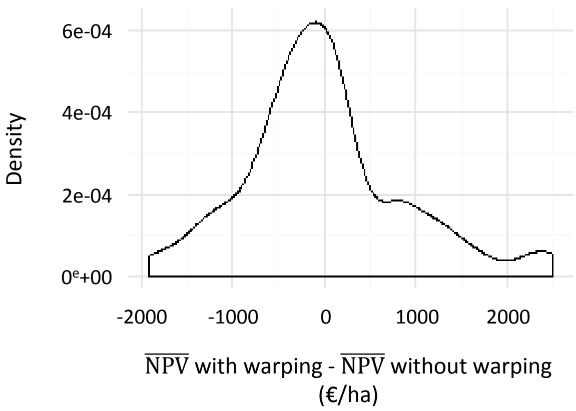


Fig. 6 Comparison of $\overline{NPV}$ obtained at the end of the optimization with and without warping.

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

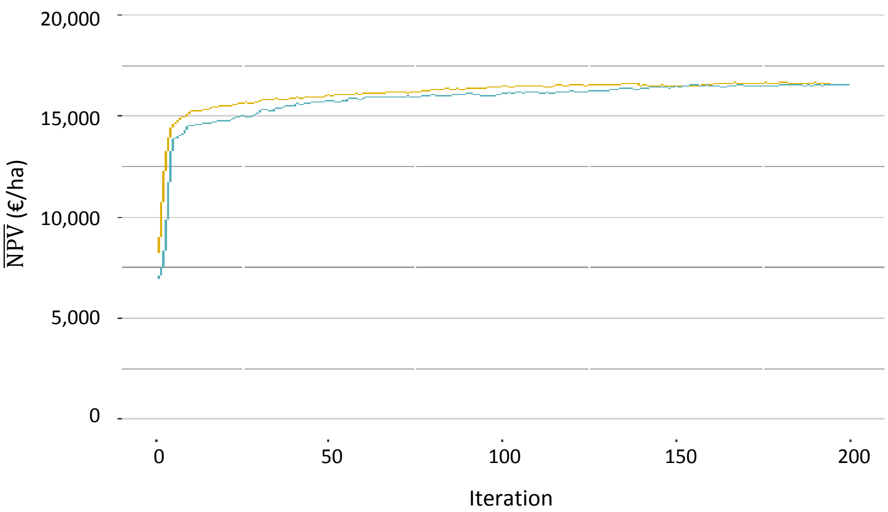


Fig. 7 Comparison of $\overline{NPV}$ obtained during optimizations with and without warping. Yellow and blue lines represent the mean of the $\overline{NPV}$ selected at each iteration for the 50 optimizations respectively perfomed with and without the warping step.

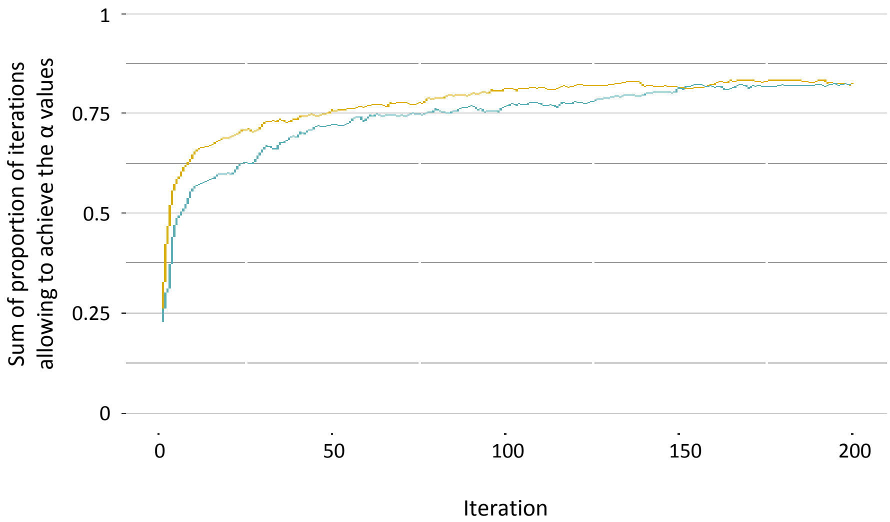


Fig. 8 Results of the algorithm using empirical distribution functions [10] with (yellow) and without (blue) warping ($\alpha \in [10{,}000;18{,}012.12]$).

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.

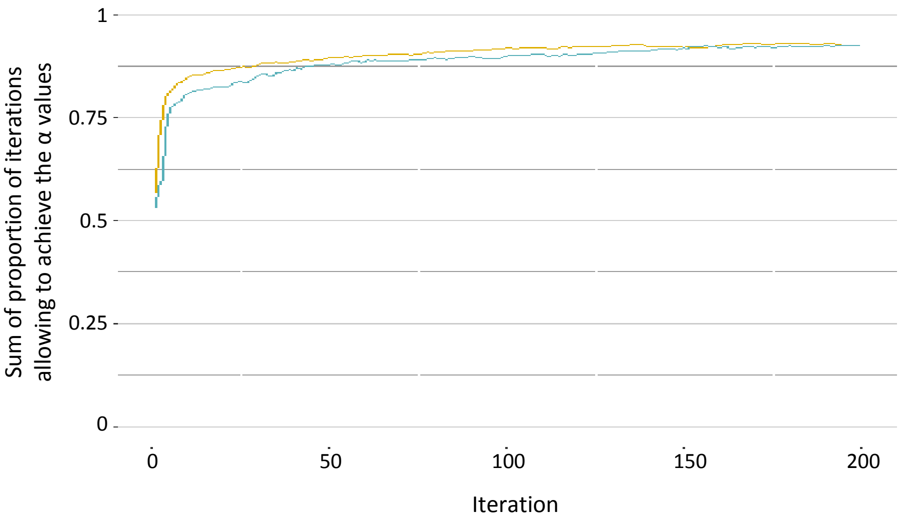
References

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.

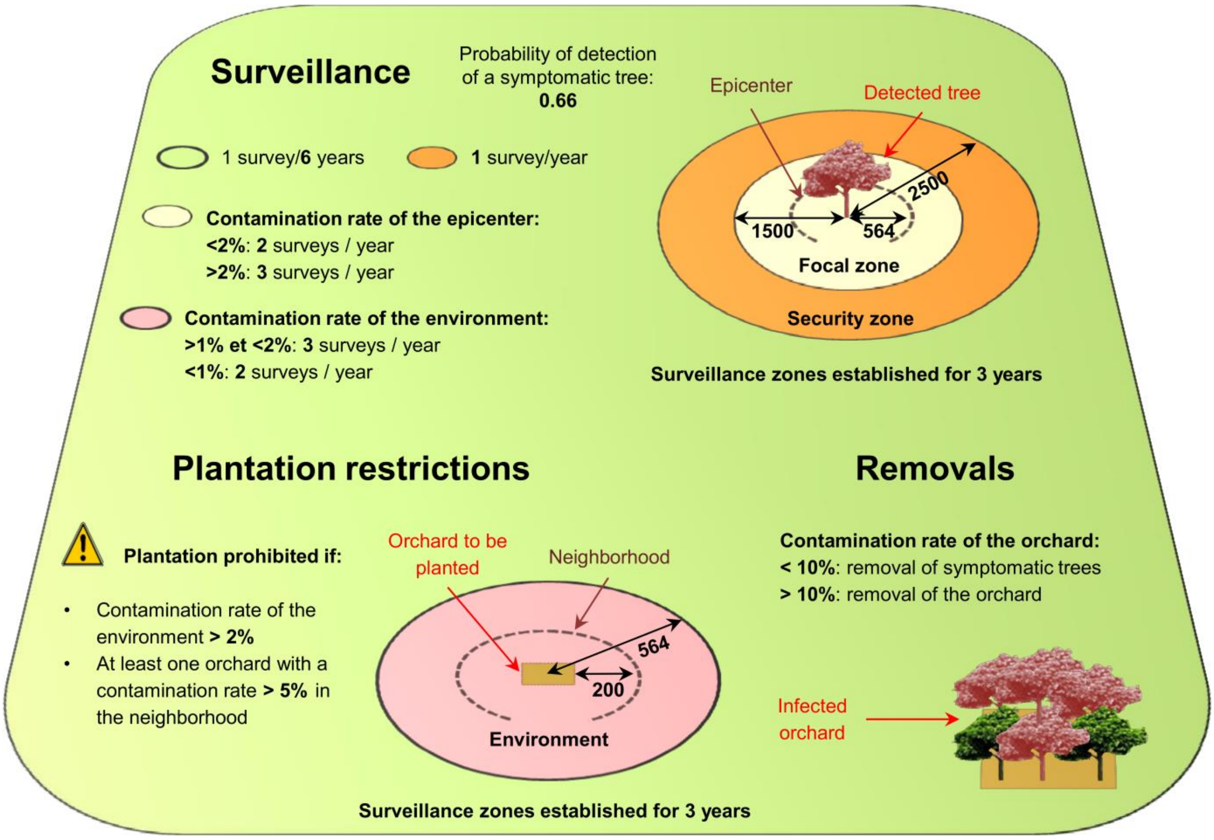


Fig 1. Management actions currently applied in France against sharka.

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_κ	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.

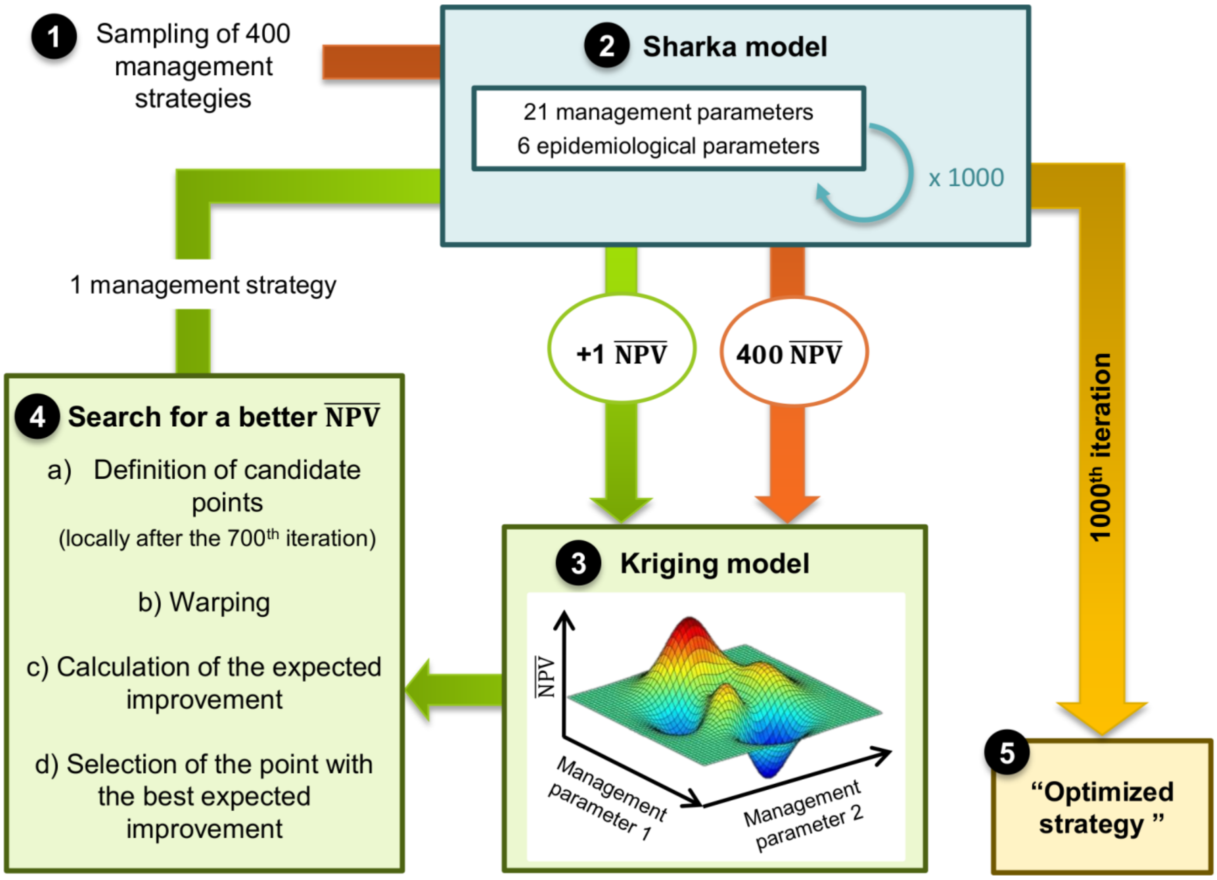


Fig 2. Optimization process of the NPV mean ($\overline{\mathbf{NPV}}$) for one level of patch aggregation. 1. Sampling of 400 management strategies (one strategy corresponds to 21 management parameters). 2. Simulations of each management strategy for 1,000 different epidemics (one epidemic corresponds to 6 epidemiological parameters), and calculation of the $\overline{\mathbf{NPV}}$. 3. Definition of a kriging model: example of a model with 2 management parameters. 4. Search for a better $\overline{\mathbf{NPV}}$: 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 $\overline{\mathbf{NPV}}$; after the 700th iteration, 110,000 candidate points of the parameter space are chosen locally around the best current $\overline{\mathbf{NPV}}$, 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 $\overline{\mathbf{NPV}}$. 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.

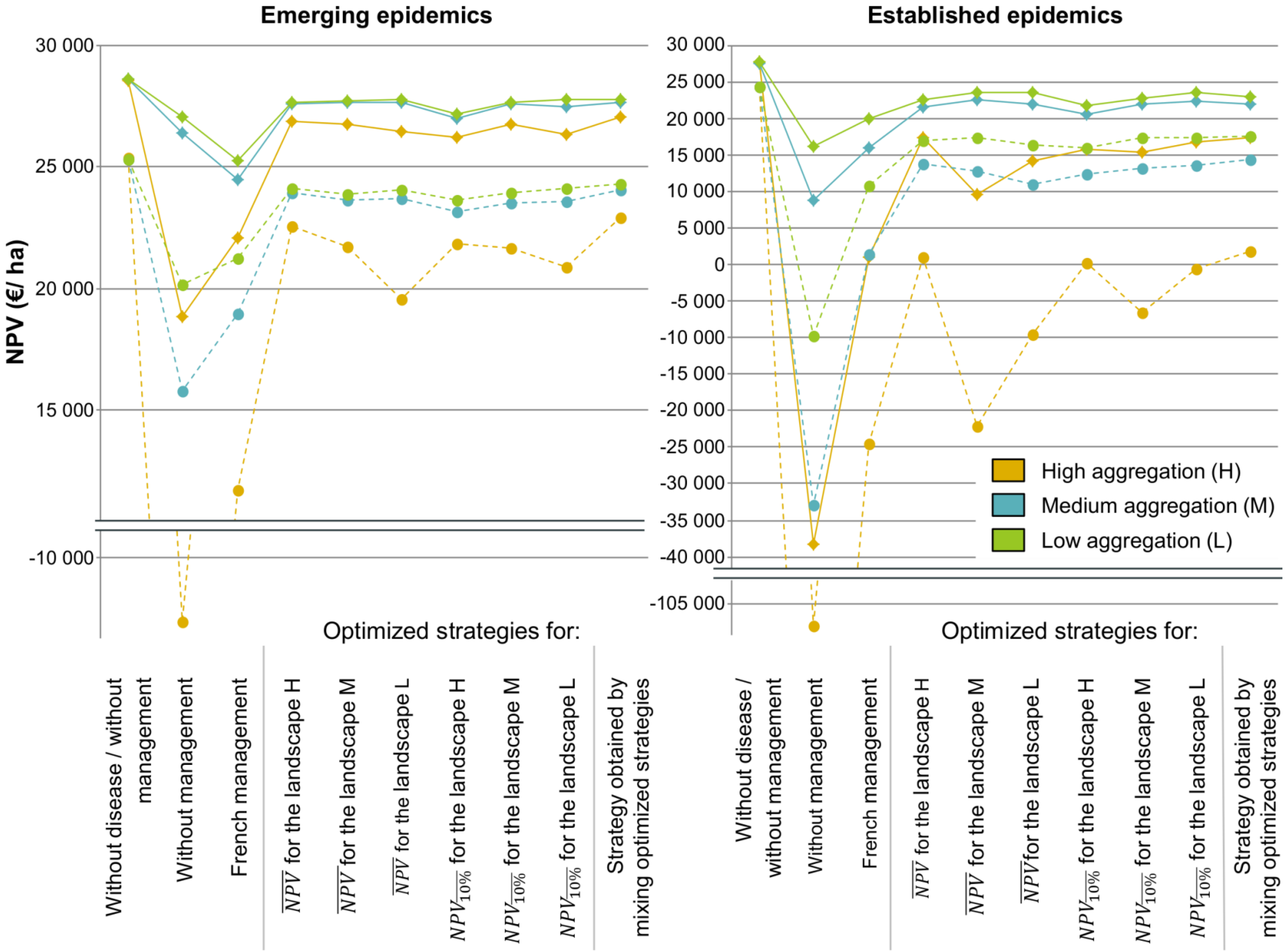


Fig 3. $\overline{\mathbf{NPV}}$ and $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ obtained after 10,000 simulations of PPV dispersal and management. The $\overline{\mathbf{NPV}}$ is represented by solid lines and the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ by dotted lines. For each epidemic type, the 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).

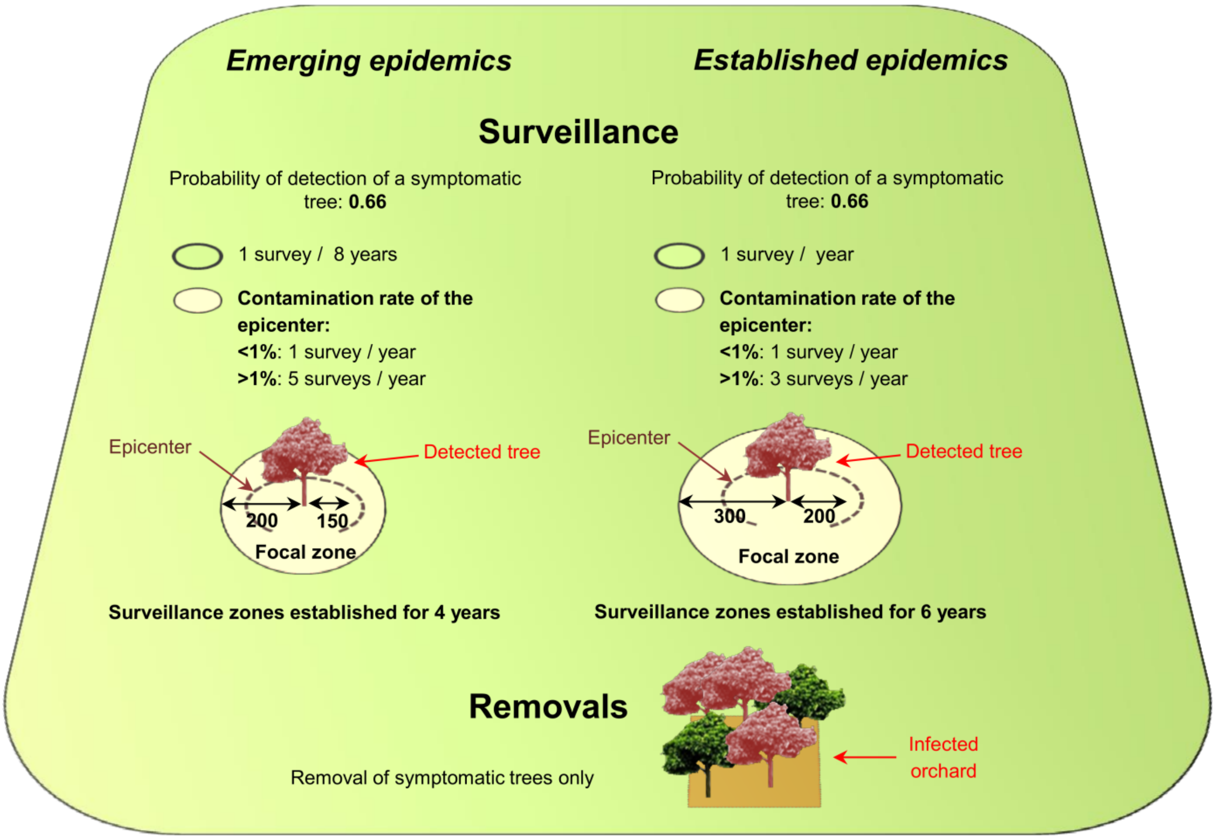


Fig 4. Management actions for the optimized strategies for emerging and established epidemics.

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.

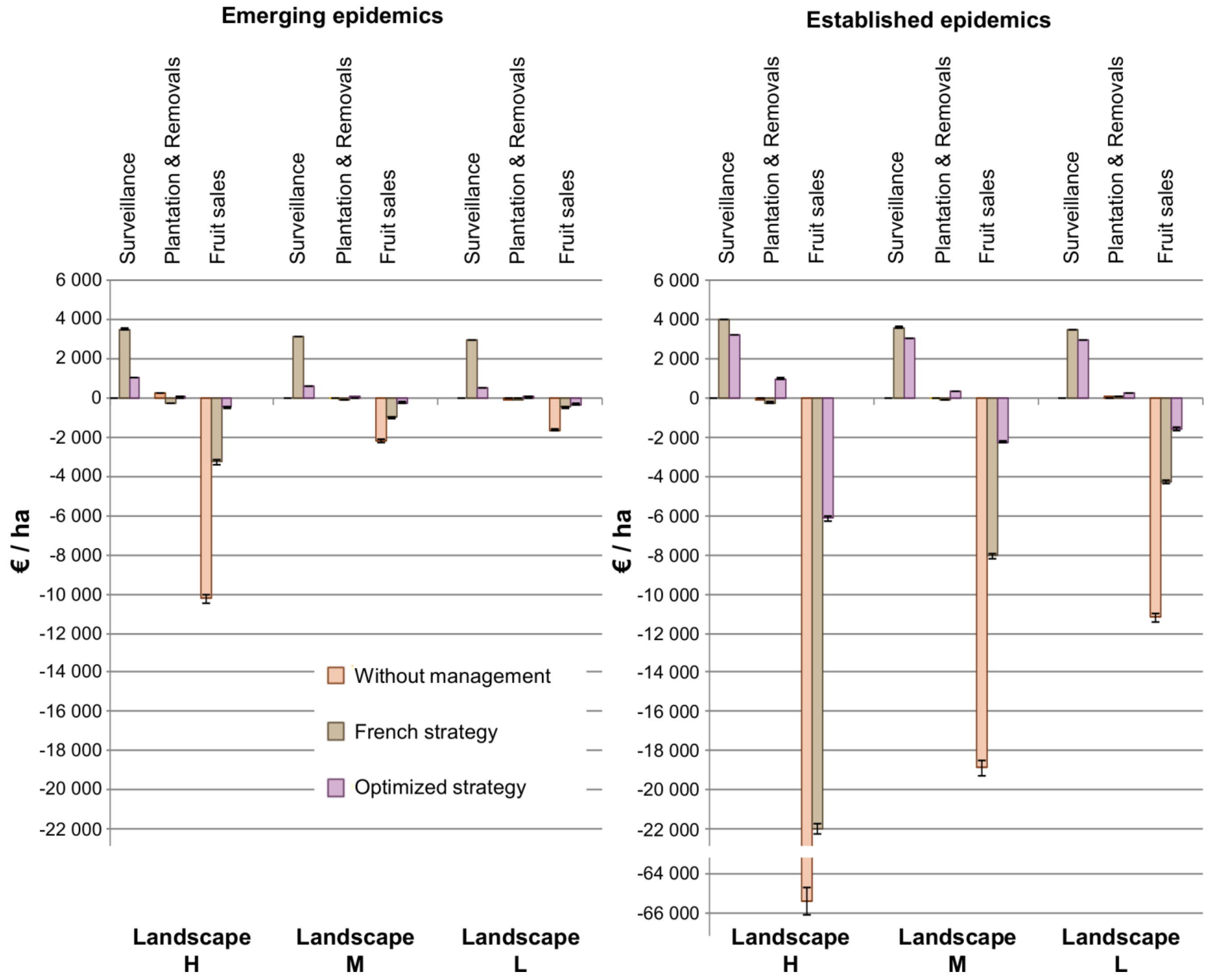


Fig 5. Comparison of the details of NPV components. Barplots represent the difference between simulations without disease (and without management) and (i) with disease and without management (red), (ii) with the French management strategy (brown), and (iii) with the optimized strategy (purple). When 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.

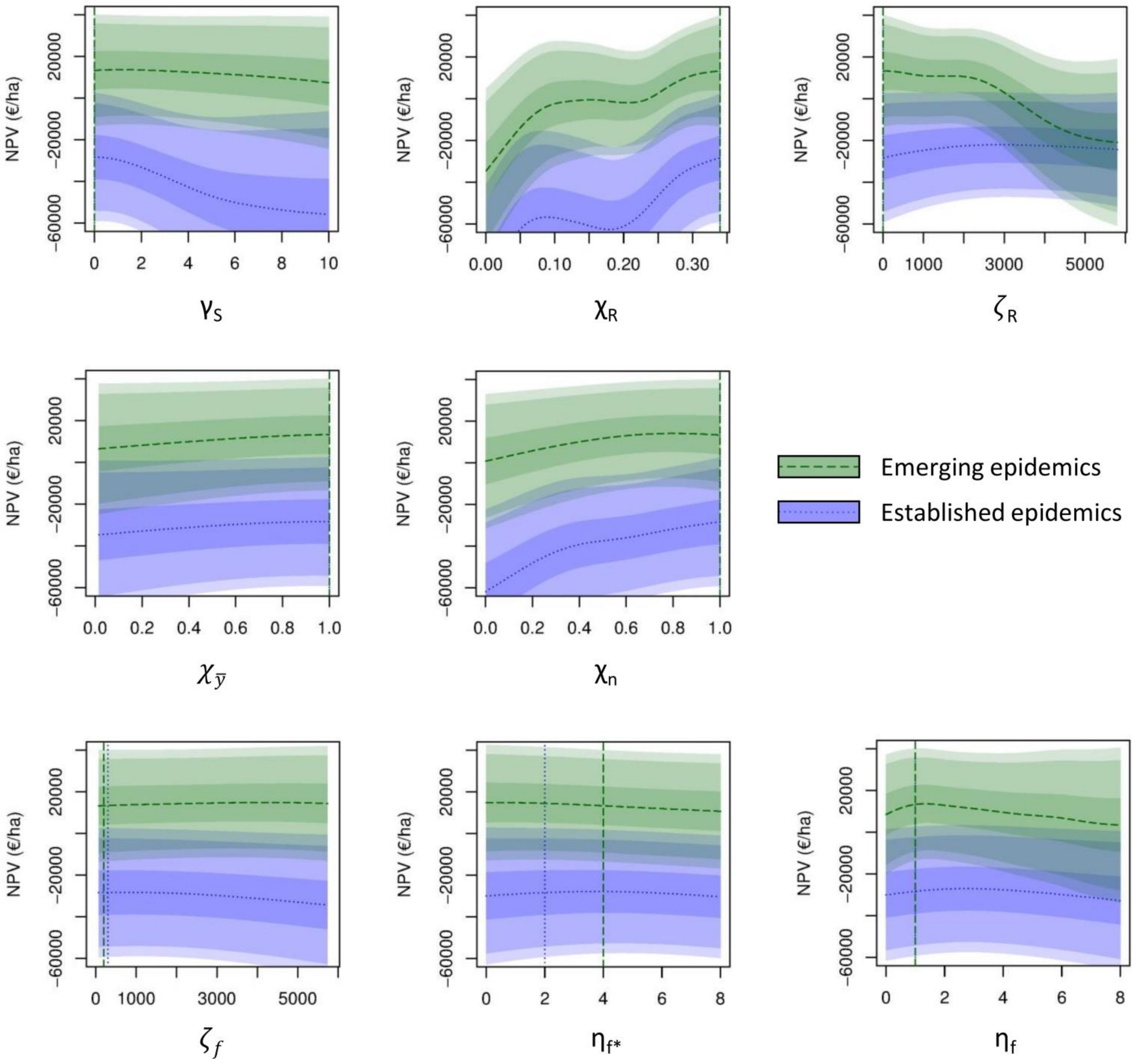


Fig 6. 2D view of kriging metamodels. These metamodels were obtained at the end of the optimization of the $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ for the most aggregated landscape (H) for emerging (green) and established (blue) epidemics. Each plot represents the influence of a single management parameter 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/

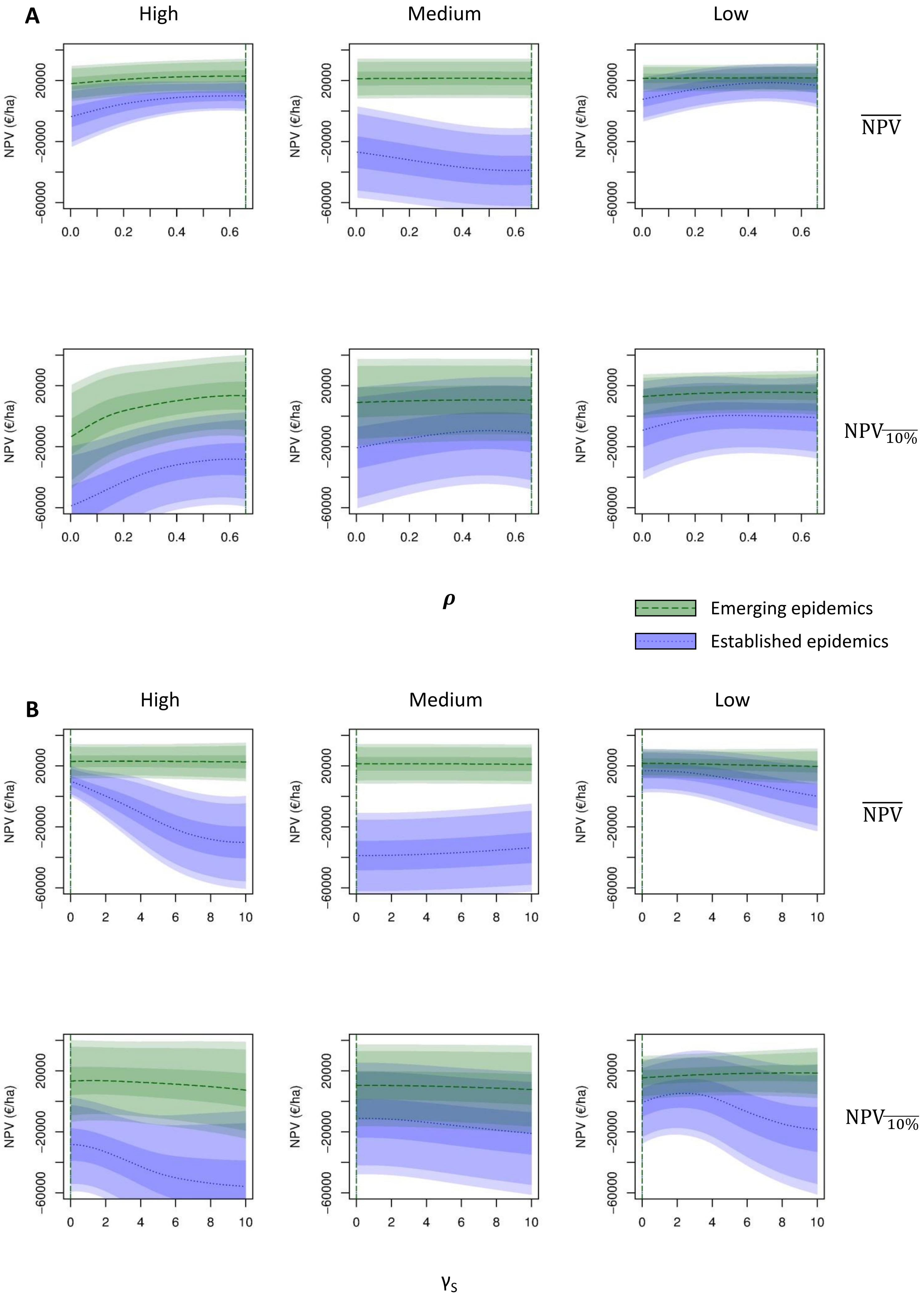
53. Filipe JAN, Cobb RC, Meentemeyer RK, Lee CA, Valachovic YS, Cook AR, et al. Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden Oak Death in Northern Californian forests. PLOS Comput Biol. 2012; 8(1): e1002328. doi:10.1371/journal.pcbi.1002328.

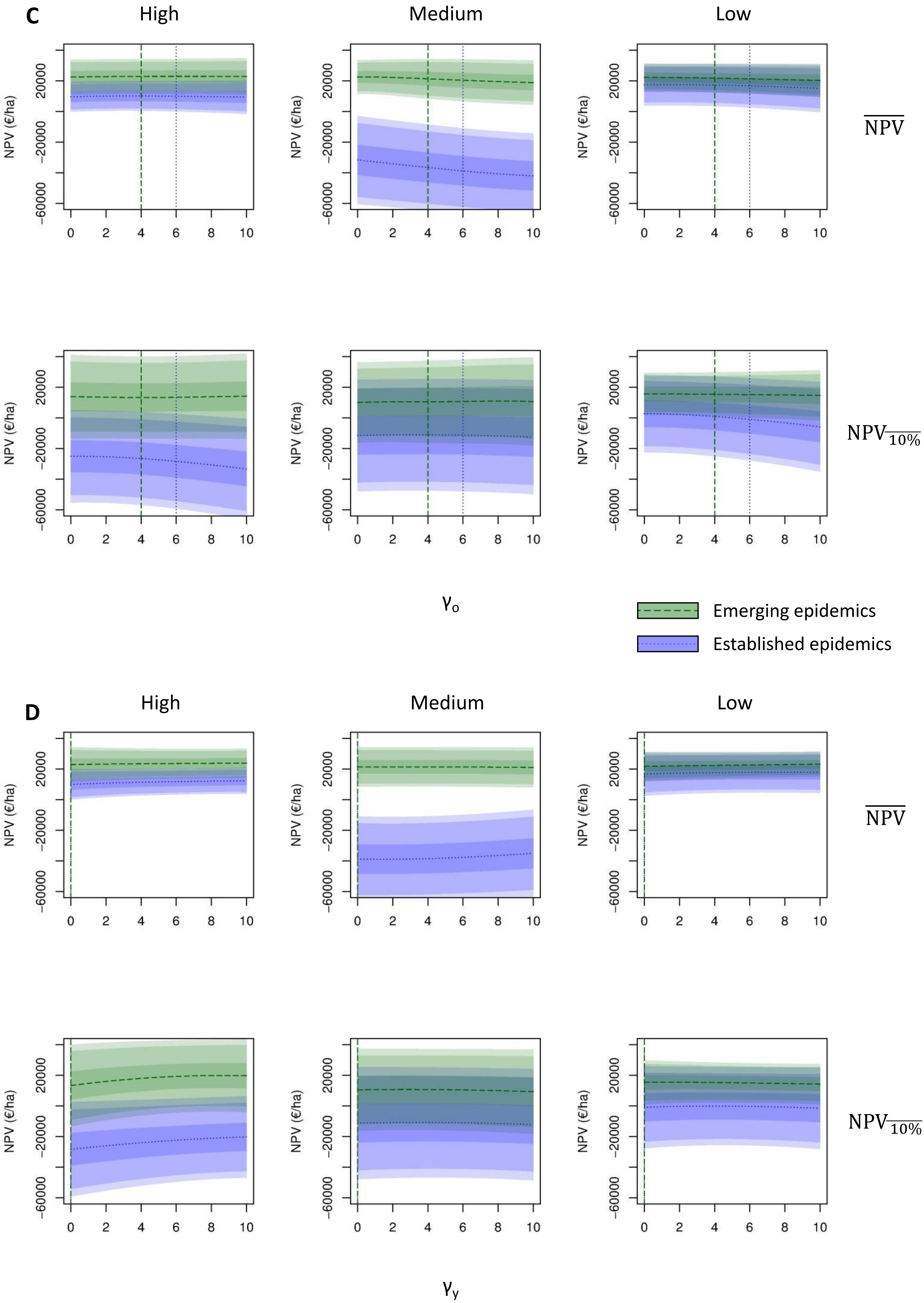
54. Parnell S, Gottwald TR, Gilks WR, van den Bosch F. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. J Theor Biol. 2012; 305: 30–36. doi:10.1016/j.jtbi.2012.03.009.

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.

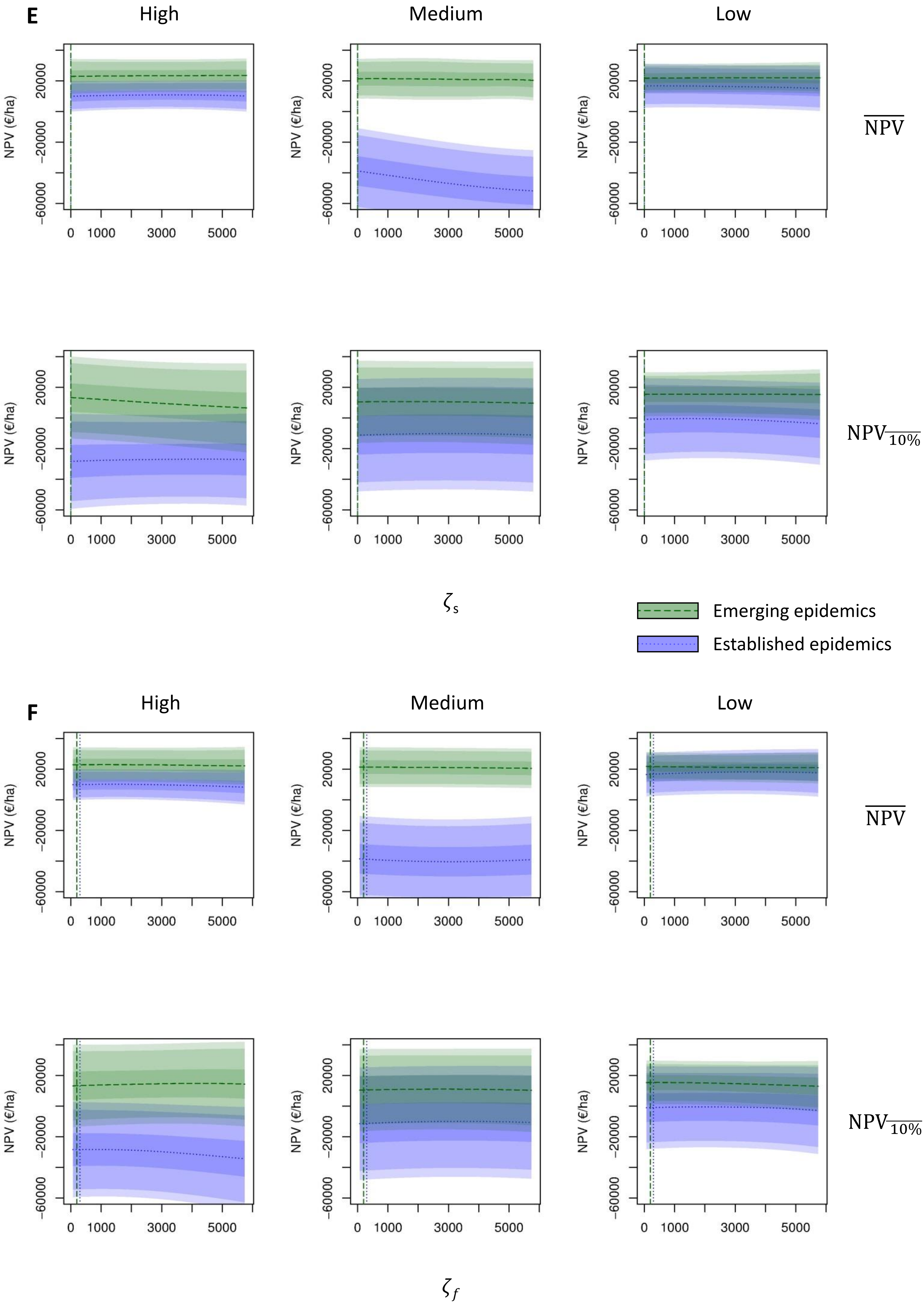
Supporting information

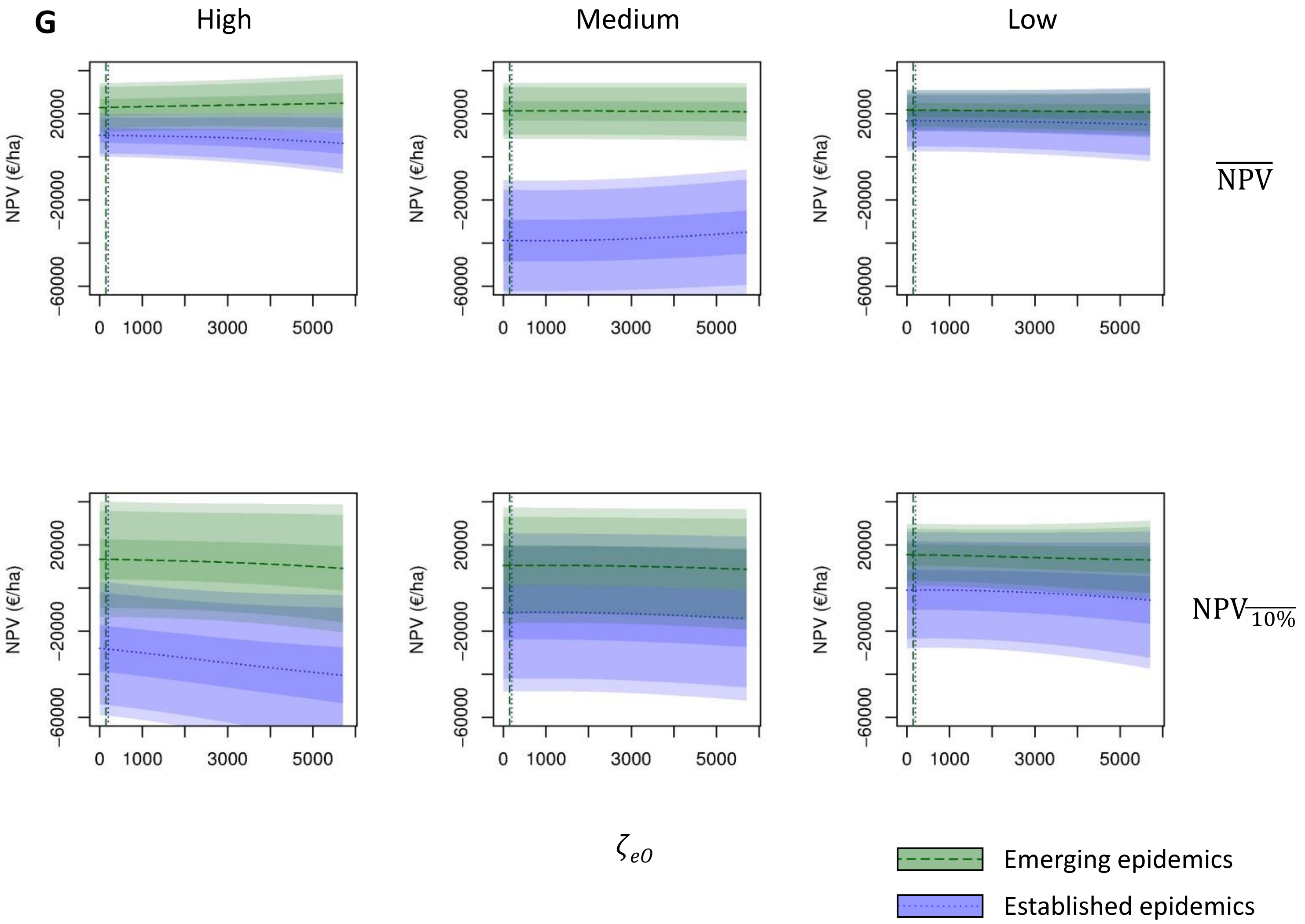
S1 Fig. 2D view of the kriging metamodels. These metamodels were obtained at the end of the optimization of $\overline{\mathbf{NPV}}$ (top row) and $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ (bottom row) for the 3 levels of landscape aggregation (from left to right) in case of emerging (green) and established (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/η0, 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)
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ρ
Emerging epidemics
Established epidemics
B
High
Medium
Low
NPV (€/ha)
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
γ_S

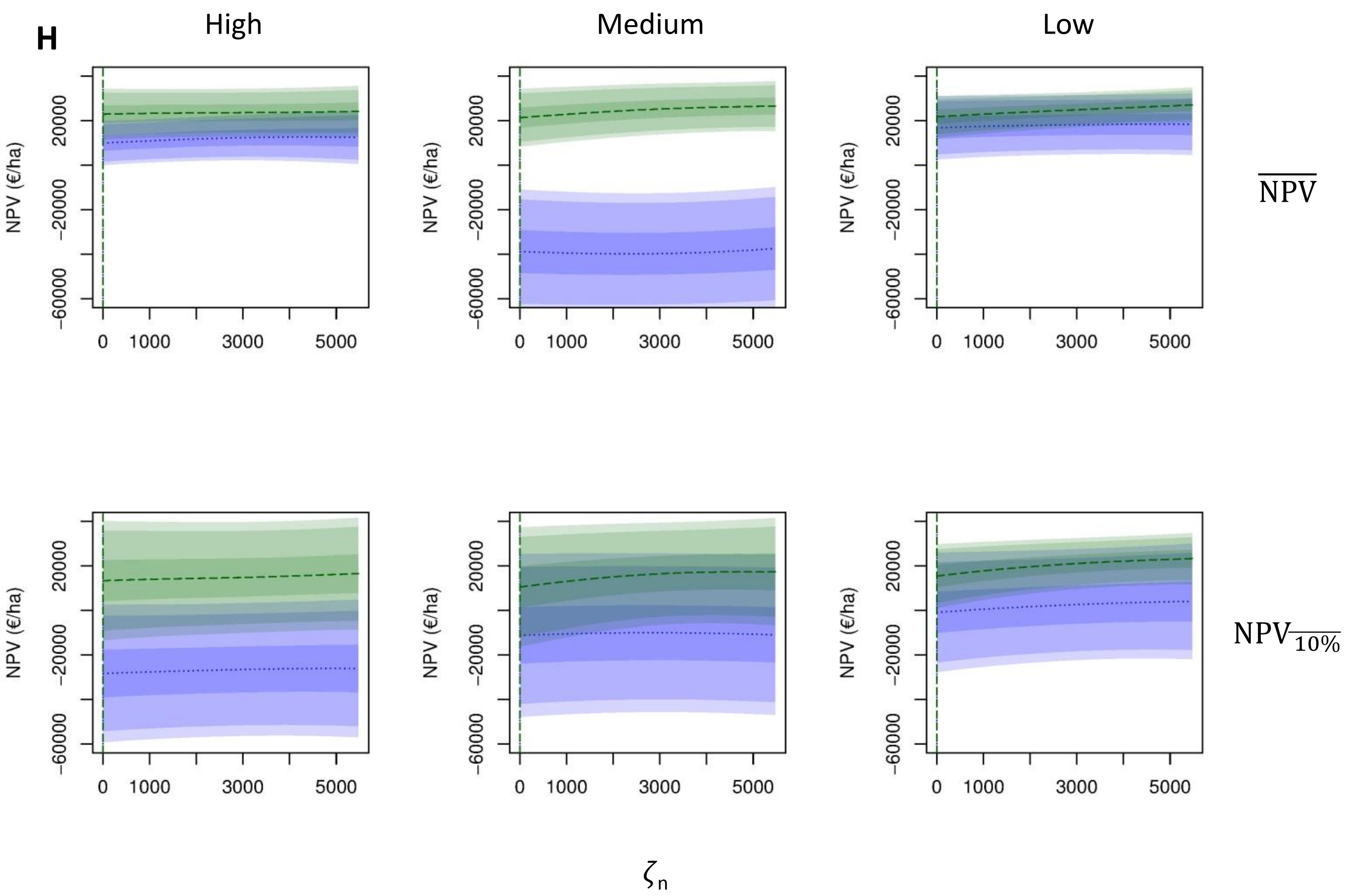


C
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
2
4
6
8
10
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
γ_o
Emerging epidemics
Established epidemics
D
High
Medium
Low
NPV (€/ha)
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
γ_y

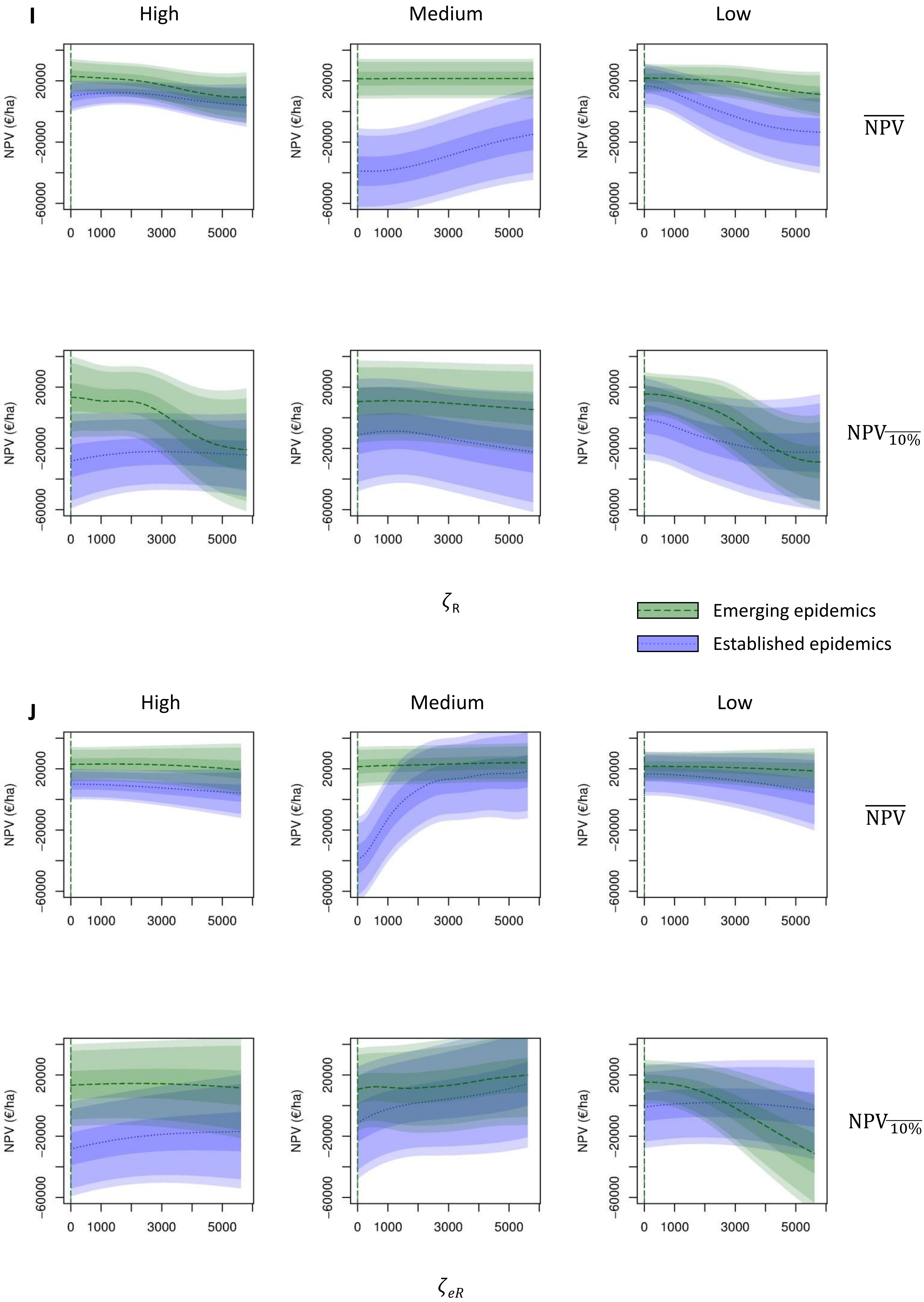

E
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
1000
3000
5000
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_s
Emerging epidemics
Established epidemics
F
High
Medium
Low
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
ζ_f

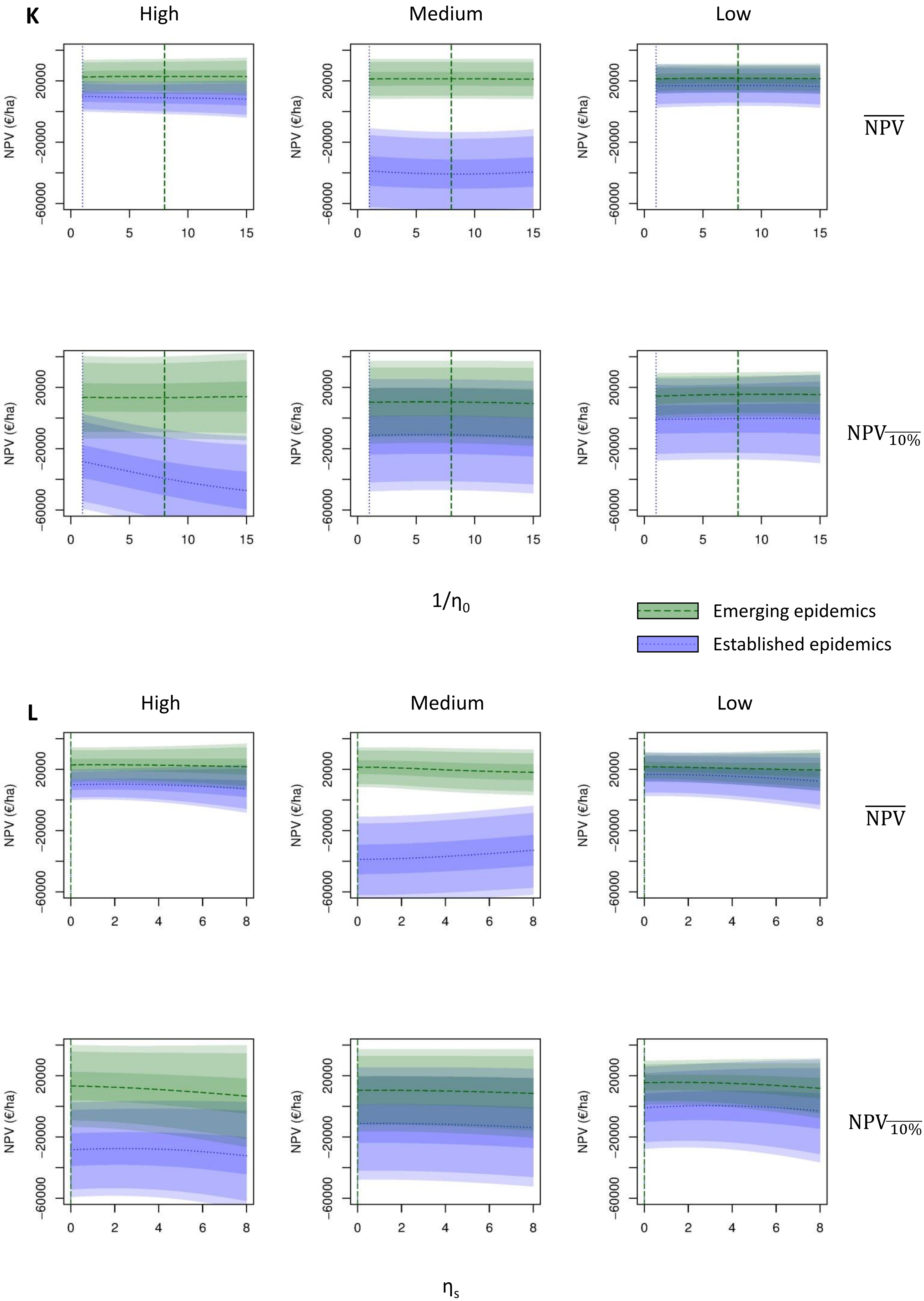


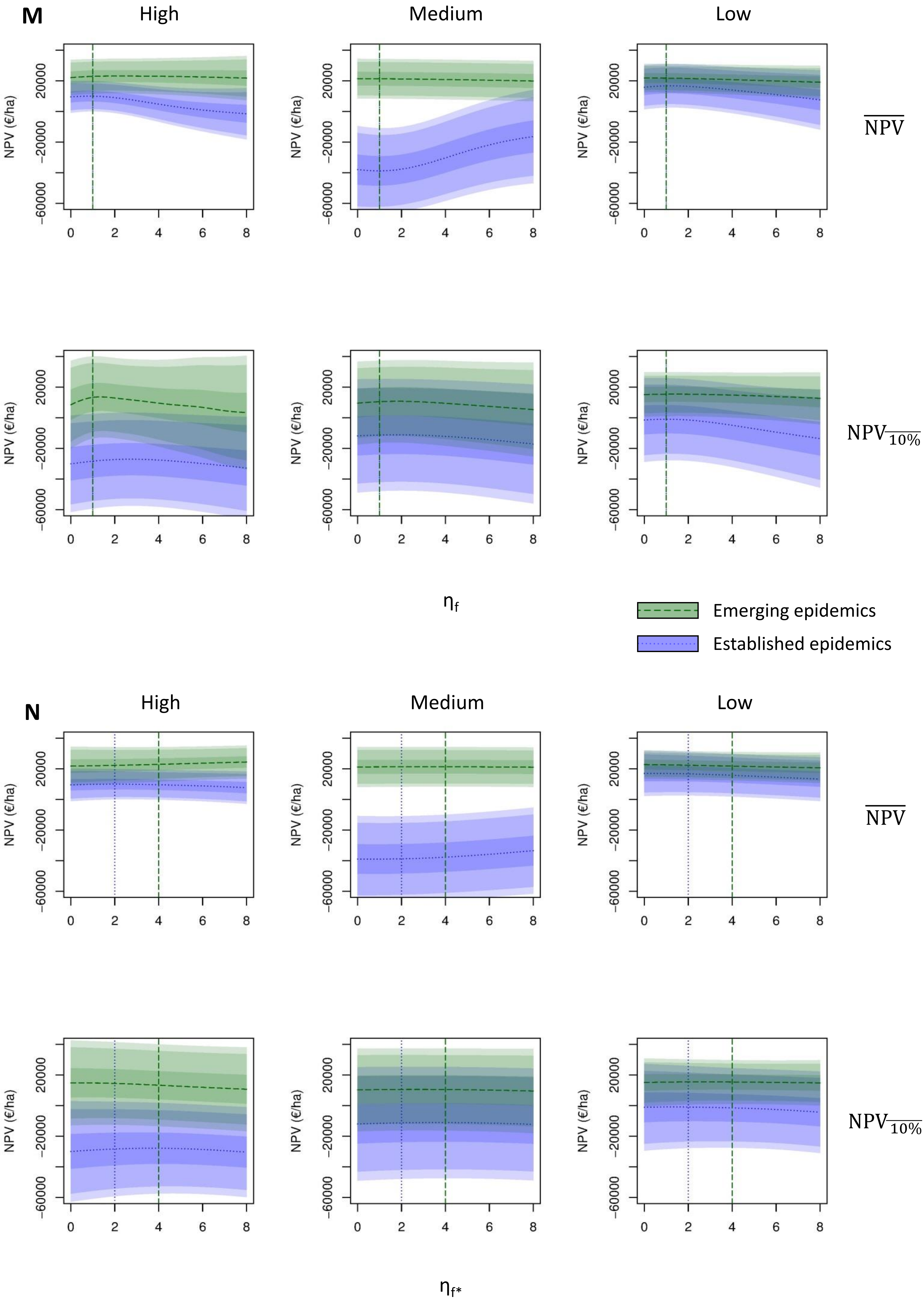
G
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
1000
3000
5000
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
ζ_{e0}
Emerging epidemics
Established epidemics

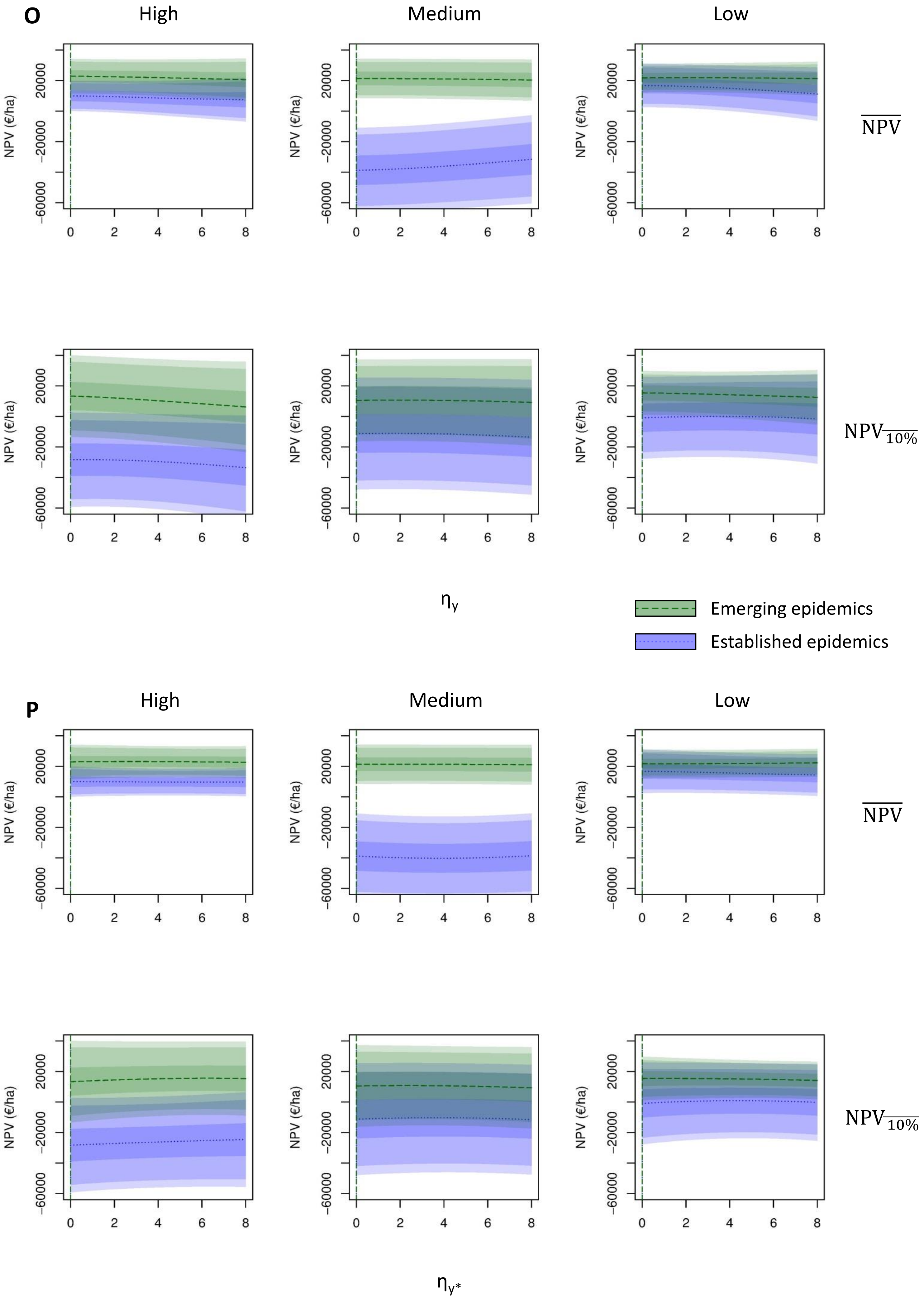


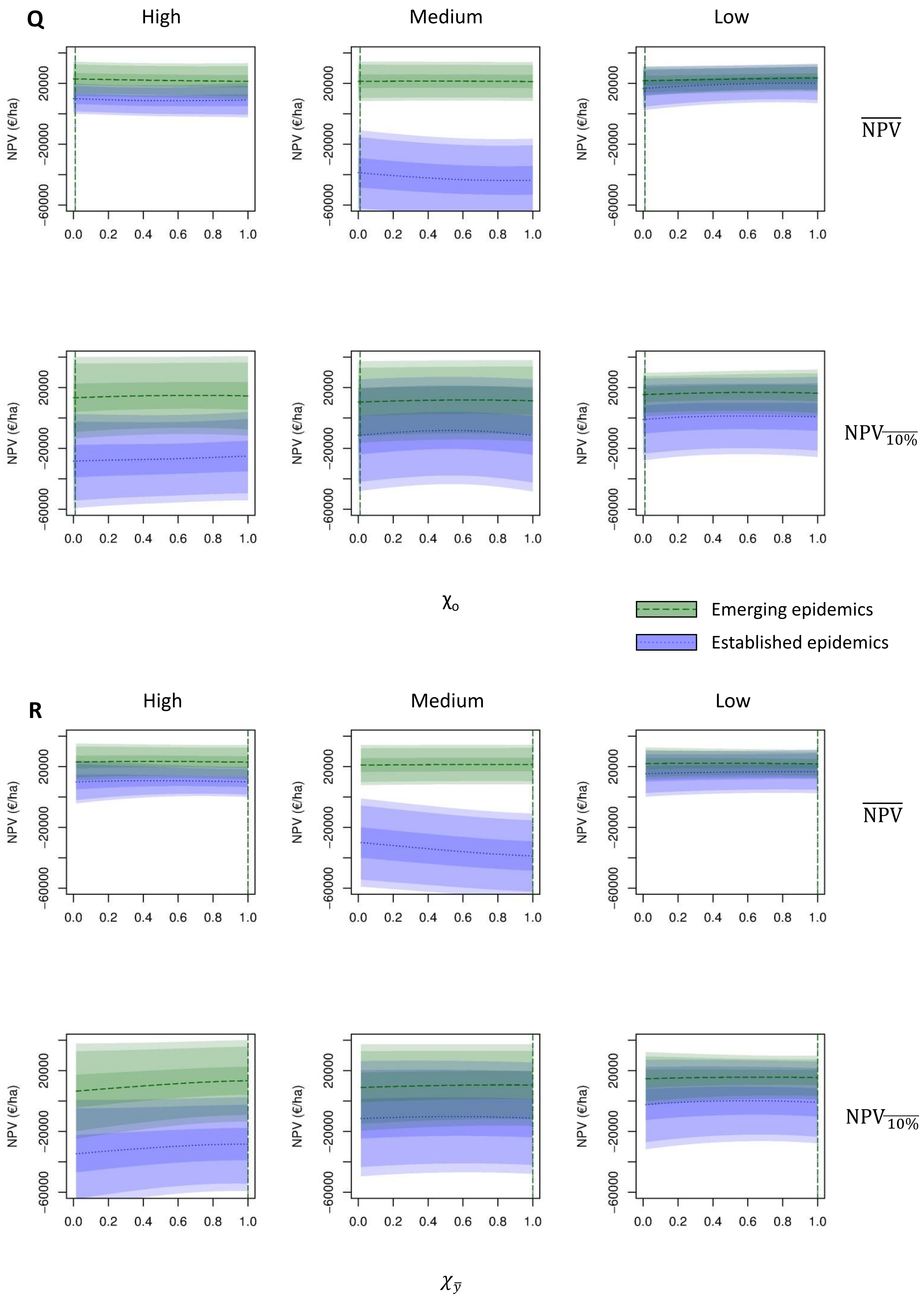
H
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
1000
3000
5000
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
ζ_{n}


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

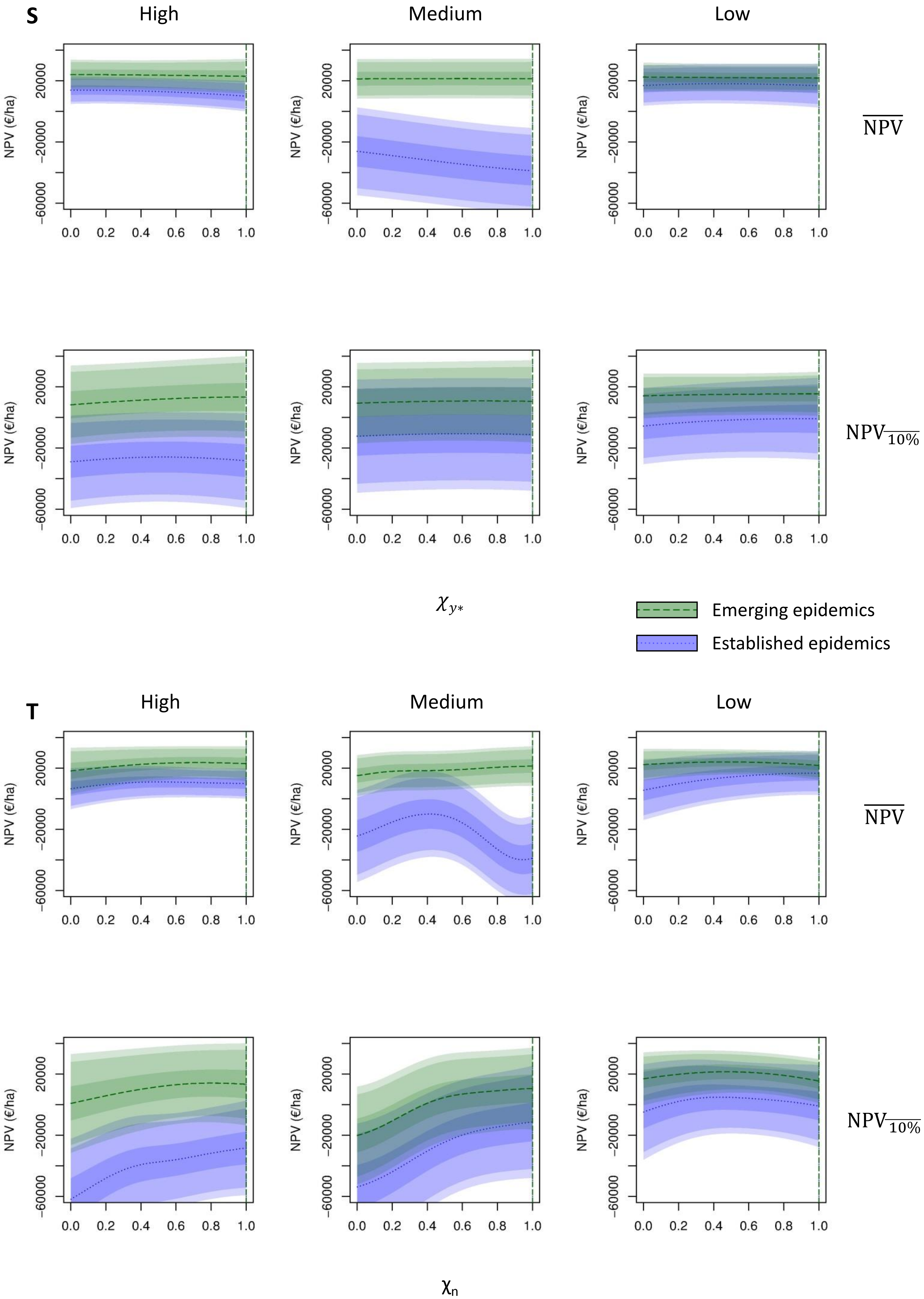

K
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
5
10
15
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
$1/\eta_0$
Emerging epidemics
Established epidemics
L
High
Medium
Low
NPV (€/ha)
0
2
4
6
8
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
η_s


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
η_{f*}

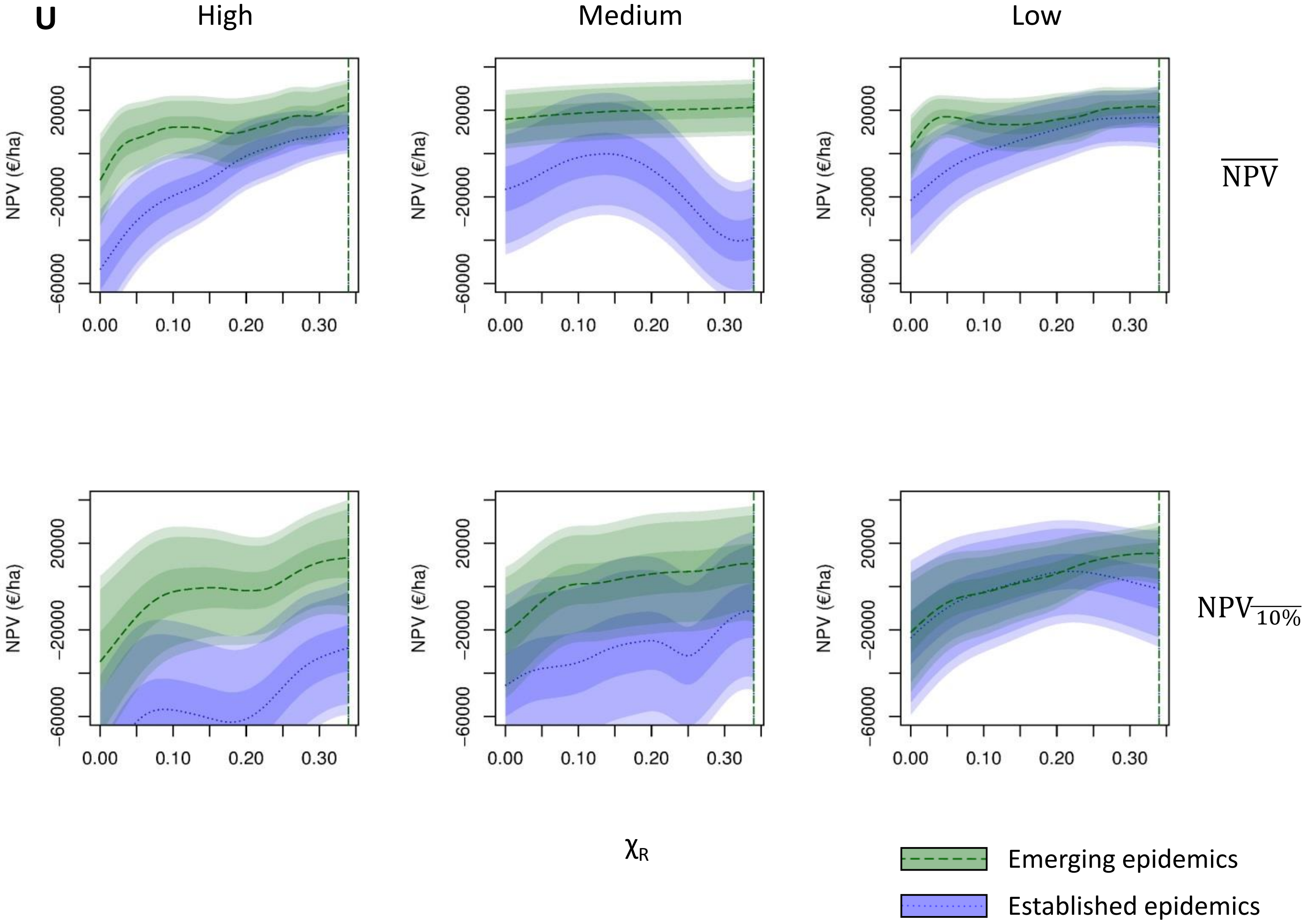

O
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0
2
4
6
8
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
η_y
Emerging epidemics
Established epidemics
P
High
Medium
Low
NPV (€/ha)
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
η_{y*}



Q
High
Medium
Low
NPV (€/ha)
20000
-20000
-60000
0.0
0.2
0.4
0.6
0.8
1.0
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
χ_o
Emerging epidemics
Established epidemics
R
High
Medium
Low
$\chi_{\bar{y}}$



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)
$\overline{\mathrm{NPV}}$
$\mathrm{NPV}_{\overline{10\%}}$
χ_n



U
High
Medium
Low
NPV (€/ha)
$\overline{\text{NPV}}$
$\text{NPV}_{\overline{10\%}}$
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).

			Emerging epidemics							Established epidemics						
			Optimization criterions						Strategy obtained by mixing some optimized strategies	Optimization criterions						Strategy obtained by mixing some optimized strategies
		French management	$\overline{\mathrm{NPV}}$			$\overline{\mathrm{NPV}_{10\%}}$				$\overline{\mathrm{NPV}}$			$\overline{\mathrm{NPV}_{10\%}}$			
Management parameters			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**
γS	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. $\overline{\mathrm{NPV}}$ and $\mathrm{NPV}_{\overline{10\%}}$ (€/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	
		$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$	$\overline{\mathrm{NPV}}$	$\mathrm{NPV}_{\overline{10\%}}$
[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: $\boldsymbol{c_O = 182 \times \rho}$, 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.

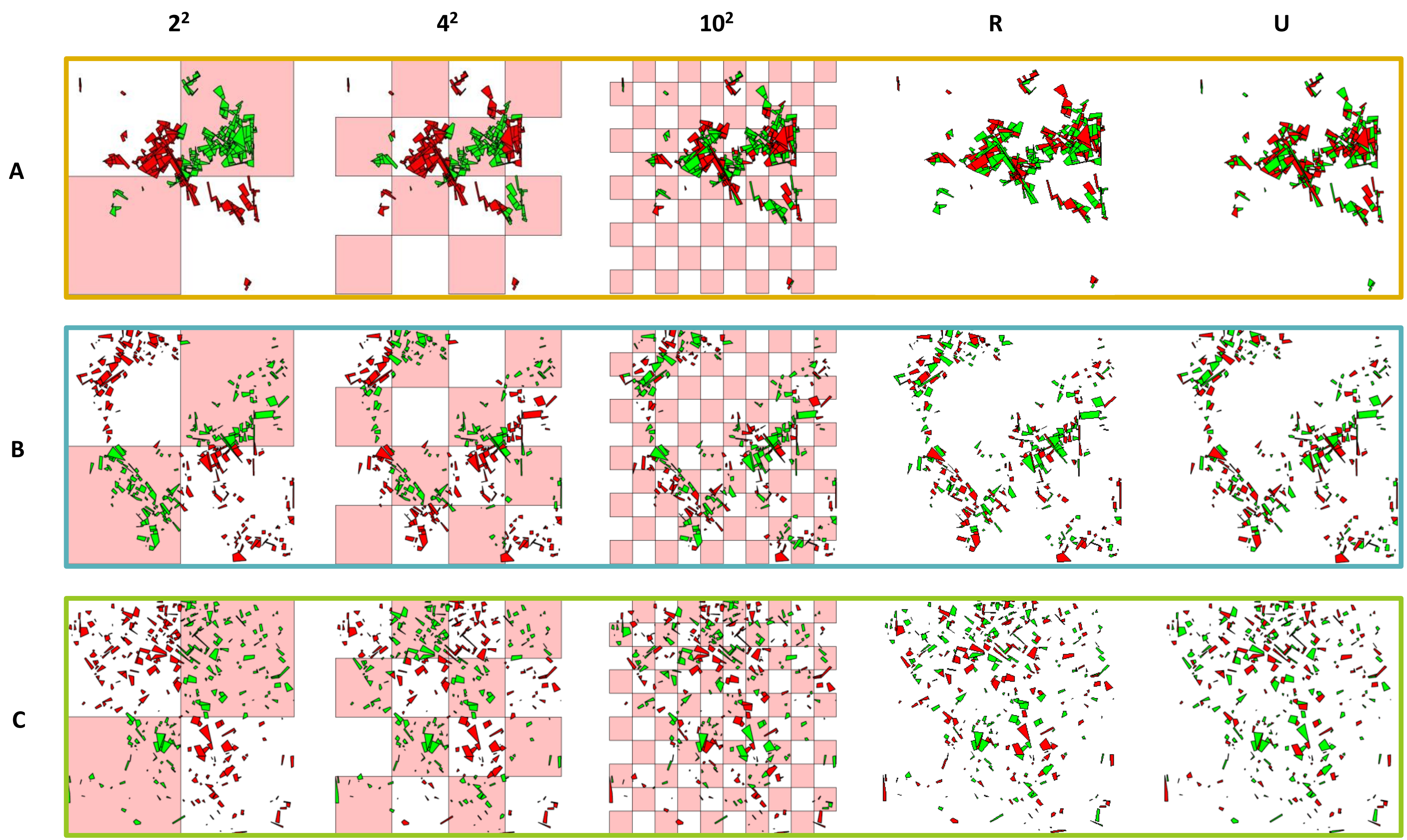


Figure 1: Spatial allocations of resistant cultivars on landscapes with high (A), medium (B), low (C) levels of patch aggregation. Green patches are located in the resistance zone, where resistant or susceptible orchards can be planted. Red patches are located in the susceptible zone, where only susceptible 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.

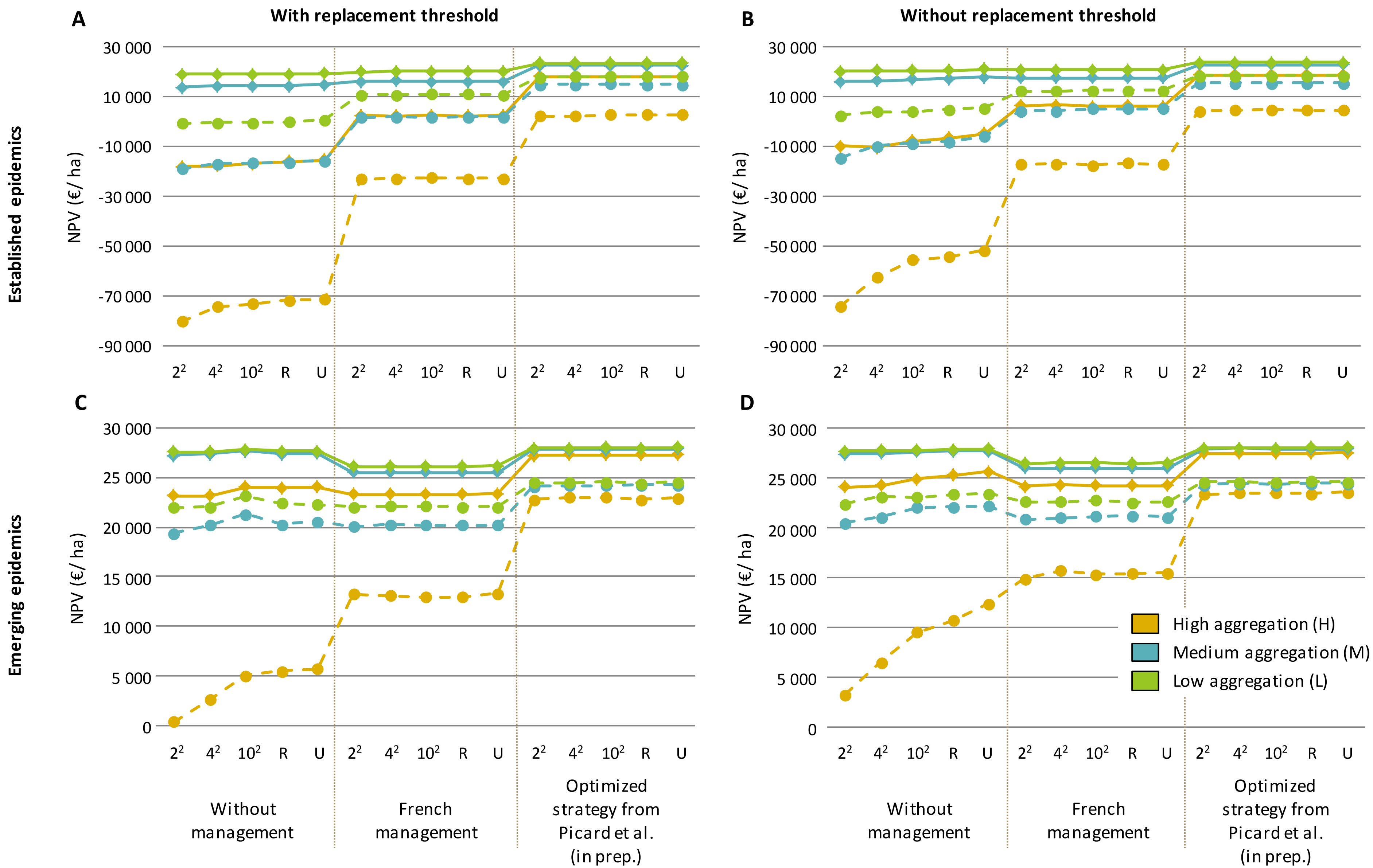


Figure 2: $\overline{\mathrm{NPV}}$ (solid lines) and $\mathrm{NPV}_{\overline{10\%}}$ (dotted lines) obtained after 10,000 simulations of PPV dispersal and management. Simulations were performed for 5 scenarios of allocation of the resistant cultivars in the landscape: 2^4. distribution in 4 squares 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 $\overline{\text{NPV}}$ 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 $\overline{\text{NPV}}$ 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: $\overline{\text{NPV}}$ (€/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 $\overline{\text{NPV}}$ (or the $\text{NPV}_{\overline{10\%}}$) corresponding to simulations performed on a landscape with the management optimized for the $\overline{\text{NPV}}$ (or the $\text{NPV}_{\overline{10\%}}$) for the same landscape (e.g. the $\overline{\text{NPV}}$ of simulations performed on landscape H with management parameters optimized for the $\overline{\text{NPV}}$ on the landscape H).

	Landscape H		Landscape M		Landscape L	
Optimized strategies optimized for :	$\overline{\text{NPV}}$	$\text{NPV}_{\overline{10\%}}$	$\overline{\text{NPV}}$	$\text{NPV}_{\overline{10\%}}$	$\overline{\text{NPV}}$	$\text{NPV}_{\overline{10\%}}$
$\overline{\text{NPV}}$ of landscape H	**18974**	4977	22983	15557	23917	18544
$\overline{\text{NPV}}$ of landscape M	17166	-1698	**23236**	14231	24375	18323
$\overline{\text{NPV}}$ 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

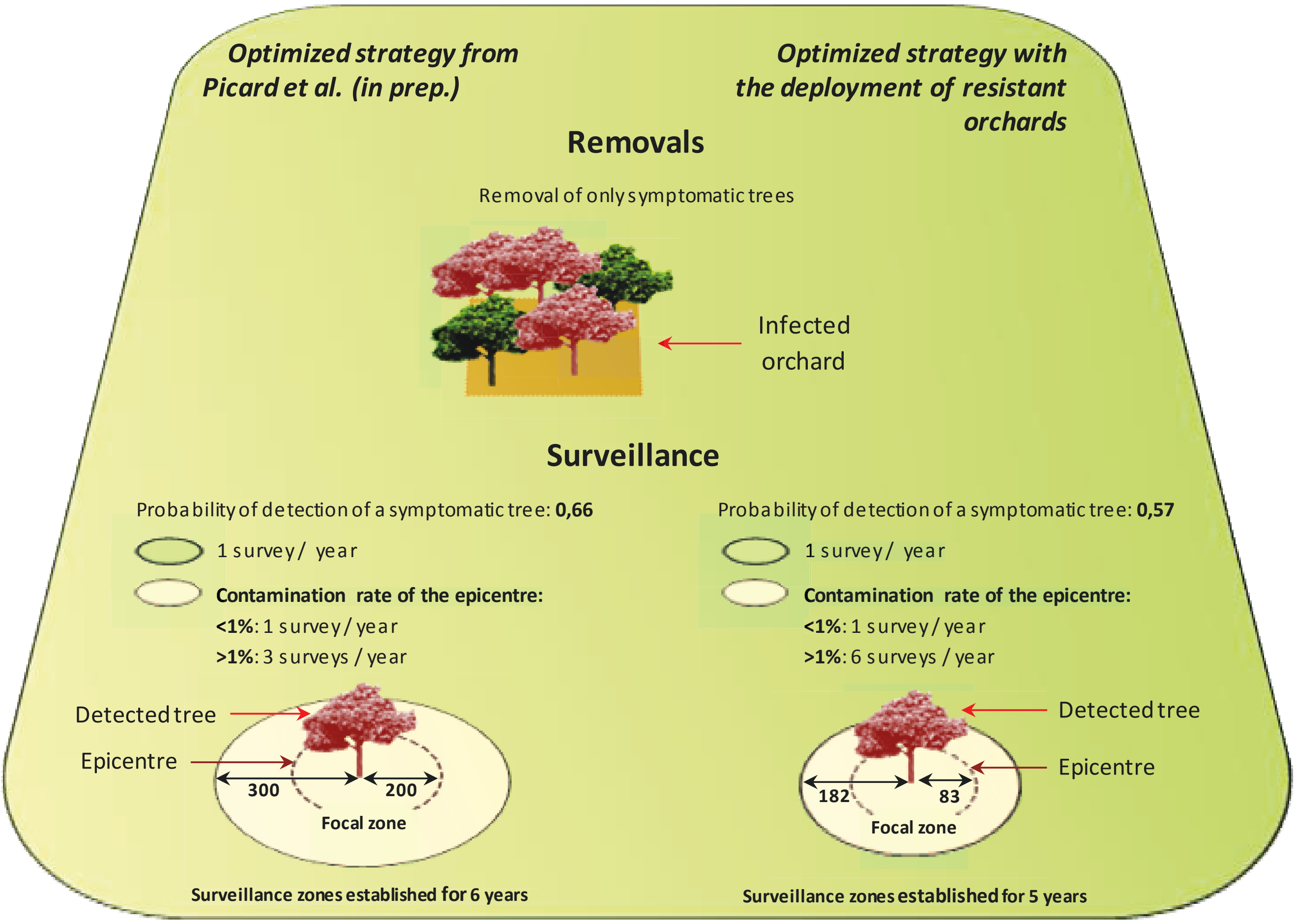


Figure 3: Management actions for optimized strategies of sharka. These strategies were obtained for established epidemics by Picard et al. (in prep) (left) and in this study (right) as a result of the optimization of the $\overline{\mathbf{NPV}}$ for landscape H with replacement of removed orchards in 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 $\mathbf{NPV}_{\overline{\mathbf{10\%}}}$ 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