



Rapport de stage

Développement d'un outil pour co-concevoir des itinéraires techniques économes en pesticides en vergers

Master Statistiques et sciences des données

Etudiant
GIRARD Julien

Lieu du stage
**Institut National de la Recherche Agronomique
Avignon**

Maîtres de stage
**MEMAH Mohamed-Mahmoud
BEVACQUA Daniele**

Encadrant universitaire
LECLERCQ SAMSON Adeline



Durée
Mars 2018 - Aout 2018

Résumé

Ce stage fait partie d'un projet innovant cherchant à concevoir des itinéraires techniques économes en pesticides en vergers et va mettre en œuvre différentes tâches. Dans ce rapport, on va s'intéresser à l'étude de modèle sur la monilia afin de mieux mieux connaître sa propagation avec différents paramètres mais on va aussi étudier son lien avec l'algorithme Qualitree qui traite des pêchers afin pouvoir étudier différents scénarios d'irrigations.

On va ainsi regarder le modèle existant et réaliser des analyses de sensibilités pour mieux connaître les paramètres. On va ensuite optimiser ce modèle afin d'obtenir des résultats le plus précis possible. On va s'intéresser surtout au crack pour travailler sur la monilia et on va donc modifier le modèle pour prendre en compte les surfaces de crack. On va ensuite pouvoir intégrer cela dans l'algorithme Qualitree qui va permettre d'obtenir des résultats de propagation lors de simulations d'arbres virtuelles. On va ainsi comparer nos modèles mais surtout des scénarios d'irrigations.

De ce fait, grâce à l'analyse de sensibilité, on va observer certains paramètres qui vont avoir beaucoup d'interactions et vont être importants. On va aussi définir un nouveau modèle prenant en compte le crack. On va optimiser nos deux modèles avec des techniques mono-critère et multi-critère. Après avoir intégré nos modèles, on a testé de créer de nouveaux arbres pour comparer les résultats. Enfin, on a pu rechercher quels scénarios de stress hydrique permettaient d'obtenir les meilleurs résultats et ainsi de réduire la propagation de la monilia dans les fruits.

Remerciements

Tout d'abord, je souhaiterai remercier Monsieur Memah Mohamed-mahmoud, mon maître de stage qui m'a permis de faire ce stage. Je souhaiterai aussi le remercier pour sa disponibilité, son aide et ses conseils tout au long du stage sur différents sujets comme par exemple sur la manière d'aborder certains problèmes.

Je remercie Monsieur Bevacqua Daniele, mon maître de stage, pour son aide à la compréhension du modèles et de ces paramètres.

Je remercie Madame Adeline Leclercq Samson, ma tutrice pédagogique pour son suivi et son tutorat lors du stage.

Je remercie Monsieur Valsesia Pierre pour ses conseils et son aide afin de maîtriser certains outils comme Qualitree et techniques très utile lors de ce stage.

Je voudrais aussi remercier toute l'équipe du laboratoire PSH qui m'on permis de m'intégrer au sein du service et de travailler dans une bonne ambiance.

Mon stage était financé par le projet ECOVERGER piloté par le ministère de l'Agriculture et de l'alimentation et le ministère de la Transition écologique et solidaire, avec l'appui financier de l'Agence française pour la biodiversité dans le cadre de l'APR "Résistance et pesticides" grâce aux crédits issus de la redevance pour pollutions diffuses attribués au financement du plan Ecophyto. Que les porteurs de ce projet soient remerciés pour ce soutien financier.

Table des matières

1	Introduction	4
2	Présentation du stage	5
2.1	L'INRA	5
2.1.1	Au niveau national	5
2.1.2	Lieu du stage	6
2.2	Objectifs	6
2.2.1	Contexte	6
2.2.2	Tâches	7
3	Modèle et données	8
3.1	Modèle original	8
3.2	Modèle adapté aux micro-fissures (cracks)	9
3.3	Jeux de données	9
4	Méthodologie et approches	10
4.1	Analyse de sensibilité	10
4.1.1	Méthode de Morris	10
4.1.2	Méthode de Sobol	12
4.1.3	Méthode FAST	13
4.2	Calibration des modèles	14
4.2.1	Formulation du problème	14
4.2.2	Choix algorithmique	15
4.2.3	Calibration des modèles sur les données des fruits infectés et sains simultanément	15
4.3	Intégration des modèles dans Qualitree	16
4.3.1	Présentation de Qualitree	16
4.3.2	Insertion dans Qualitree	16
4.3.3	Optimisation des scénarios	17
5	Résultats	18
5.1	Résultats de l'analyse de sensibilité	18
5.2	Résultats des estimations et optimisation	19
5.2.1	Test préliminaire	20
5.2.2	Minimisation de deux RMSE simultanément	21
5.2.3	Minimisation de deux NRMSE simultanément	24
5.3	Intégration dans Qualitree	26
5.3.1	Etude des simulations	26
5.3.2	Optimisation	28
6	Conclusion	29
7	Annexes	30

Chapitre 1

Introduction

De nos jours, une des grandes missions de l'agriculture est de satisfaire des besoins alimentaires croissants de la société, tout en réduisant les nuisances causées à l'environnement (Foley et al. 2011). On sait que les fruits sont des produits de haute qualité nutritionnelle et ils vont ainsi être une composante majeure de l'alimentation et de la santé humaine (INPES 2002). Cependant, on voit que leurs productions ont de fortes pressions de bioagresseurs (ravageurs) qui peuvent conduire à des pertes de récolte importantes. Une perte sera une baisse du rendement et/ou un déficit de qualité des fruits ce qui pourrait entraîner une insécurité alimentaire et causer des pertes économiques pour les producteurs.

On sait aussi que la plupart des dommages en arboriculture fruitière sont dûs à des bioagresseurs attaquant les fleurs ou les fruits (Aubertot et al. 2005). Malgré que les producteurs utilisent de plus en plus des méthodes alternatives, la protection des vergers repose encore beaucoup sur la lutte chimique (en préventif, en curatif ou en complément des méthodes alternatives). De plus, l'arboriculture fruitière représentait en valeur 4% du marché national des fongicides et 21% du marché des insecticides en 1998 (Aubertot et al. 2005).

C'est pour cela que le plan national Ecophyto a été lancé en France en 2009. Ce plan national a pour but de réduire de 50% l'utilisation des pesticides en France à l'horizon de 2018 tout en maintenant une agriculture économiquement performante et assurant un niveau de production élevé tant en quantité qu'en qualité.

Le projet ECOVERGER dans le cadre duquel rentre mon stage s'inscrit dans ce plan national. Ce projet vise à concevoir des profils d'itinéraires techniques en verger qui soient économes en pesticides et en accord avec les contraintes et les objectifs des producteurs pour en faciliter leur adoption.

Dans ce qui suit, je présente le contexte de mon stage, ses objectifs, et les missions qui m'étaient confiées.

Chapitre 2

Présentation du stage

2.1 L'INRA



2.1.1 Au niveau national

Mon stage s'est déroulé à l'institut national de la recherche agronomique (INRA) . L'INRA a été créé en 1946 dans le but de « nourrir la France » après la guerre. Cet institut compte aujourd'hui 8165 agents titulaires, dont 50,8% de femmes, répartis dans 250 unités de recherche et 48 unités expérimentales. L'INRA accueille 2616 stagiaires ainsi que 522 doctorants, l'institut compte aussi 4081 publications scientifiques en 2015.

L'INRA travaille dans différents domaines dont la recherche pour comprendre le monde vivant, faire de la recherche sur les communautés microbiennes, sur les épidémies animales et sur le changements de climat comme la sécheresse ou les variations de températures.

L'INRA a ainsi plusieurs objectifs :

- (i) La recherche de nouveaux moyens d'agir
- (ii) L'écoconception de produit alimentaires, de biomatériaux ou de bioénergies
- (iii) Adapter des cultures aux changements du climat
- (iv) Concevoir des élevages durables
- (v) La conception d'innovation qui va permettre le développement agricole, industrielle et de service
- (vi) Permettre une meilleur compréhension de certaines décision pour des acteurs privés ou publics
- (vii) Former à la recherche grâce à des doctorants et des partenariats avec des établissements de l'enseignement supérieur

L'INRA va ainsi chercher à mieux adapter les ressources aux besoins nutritionnels de l'Homme, en proposant une agriculture productive, compétitive, et respectueuse de l'environnement.

2.1.2 Lieu du stage

Mon stage s'est déroulé au centre Inra Provence-Alpes-Côte d'Azur qui est un centre de recherche tourné vers l'agroécologie des systèmes de culture sous serres et en vergers, ainsi que la modélisation de l'impact régionalisé du changement climatique à l'échelle du paysage.

Ce centre se place au 4ème rang national parmi les 17 centres Inra avec un budget consolidé de 54 millions d'euro. Il est composé de plus de 1000 agents avec 700 agents permanents. On compte aussi 26 unités répartis sur 10 sites différents.

J'étais plus précisément dans l'unité PSH (Plantes et systèmes de culture horticoles) qui utilise des approches d'écophysiologie et d'agroécologie pour mettre au point des systèmes de culture des fruits et légumes et de scénarios paysagers en zone méditerranéenne afin d'améliorer la qualité des produits récoltés et le respect de l'environnement.

Cette unité est dirigée par Hélène Gautier et compte une quarantaine de permanents. L'unité a ainsi des vergers, des serres, des chambres de culture, des enceintes climatiques et un insectarium. Elle étudie en particulier les plantes horticoles comme les tomates ou la laitue mais aussi les pêchers et les pommiers.

L'unité a divers enjeux avec la production des connaissances et des outils d'intégration de ces connaissances pour permettre la transition vers une horticulture écologiquement intensive favorisant la réduction d'intrants et la qualité des produits. Un autre enjeu sera d'accompagner cette transition avec des collaborations et les partenaires régionaux, institutionnels et privés.

2.2 Objectifs

2.2.1 Contexte

Mon stage s'inscrit dans le cadre du projet ECOVERGER (APR PESTICIDES-Ecophyto 2015-2018) qui cherche à concevoir des itinéraires techniques économes en pesticides en vergers guidée par les contraintes et les objectifs des agriculteurs. Pour ce faire, les partenaires de ce projet proposent de combiner des approches expérimentales et de modélisation. Un objectif spécifique de ce projet est de développer un cadre de modélisation générique pour des cultures fruitières qui représente les processus impliqués dans la régulation des bioagresseurs et l'élaboration du rendement et de la qualité des fruits ainsi que leur pilotage par des pratiques, et qui intègre des indicateurs de performance co-définis avec les acteurs. Ce cadre de modélisation sera éprouvé sur trois pathosystèmes d'importance économique : manguier-mouches des fruits et manguier-cécidomyies des fleurs à La Réunion ; pêcher-pourriture brune (moniliose) dans le sud-est de la France.

Ce projet va avoir différents objectifs avec un premier objectif de créer un outil permettant entre acteurs et chercheurs mobilisable de communiquer et de concevoir des itinéraires techniques économes en pesticides en vergers. Le second objectif de ce projet est de définir des profils candidats d'itinéraires techniques permettant un compromis entre les indicateurs de performance définis selon les conditions, les objectifs et les contraintes spécifiques aux producteurs.

On dispose d'un modèle épidémiologique développé par Daniele Bevaqua (Bevacqua et al. 2017). Ce modèle simule la diffusion de la moniliose dans un verger de pêcher en lien avec des pratiques culturales. Ce modèle va se servir de courbe de croissance pour prédire l'épidémiologie de la moniliose dans une parcelle de pêcher.

Plus exactement, le modèle de Daniele Bevaqua va traiter de la propagation de la pourriture brune dans les vergers. Cette infection est causée par la *Monilinia* qui va surtout toucher les fruits à noyaux comme les pêches ou les abricots. La monilia va créer sur le fruit des taches noires ornées de cercles concentriques et/ou de pustules blanches. Avec de bonnes conditions environnementales ainsi que l'ab-

sence de traitement, va causer de nombreuse perte dans le verger. Aujourd'hui, cette infection est présente dans les régions tempérées et est traité avec des produits chimiques qui ont un fort coût sur l'environnement. Ce modèle va nous permettre d'étudier la propagation de cette infection au cours du temps et ainsi appliquer de meilleures stratégies de contrôle.

2.2.2 Tâches

Les tâches qui m'étaient confiées au début de mon stage sont :

1) La première tâche a été dans un premier temps d'analyser la sensibilité du modèle original afin de déterminer quels sont les paramètres les plus influents. Puis on réalise une re-calibration du modèle sur 2014 et 2015.

2) La deuxième tâche consistait à remplacer la courbe de croissance du fruit par la courbe de surface de crack en entrée du modèle épidémiologique. Puis de re-calibrer ce modèle modifié avec les données des années 2014 et 2015. Enfin, de comparer les sorties obtenues par les deux modèles (original et modifié).

3) Dans la troisième tâche, on a intégré le modèle modifié dans Qualitree. Après cela, on a simulé des scénarios avec Qualitree et étudié les sorties obtenues grâce aux modèles. Et enfin, on a réaliser une optimisation du scénario de stresse hydrique grâce à la création de nouveau scénarios.

Chapitre 3

Modèle et données

3.1 Modèle original

On s'intéresse au modèle sur la propagation de la pourriture brune (moniliose) dans un verger de pêche développé par Daniele Bevacqua. Ce modèle contient 3 équations et 8 paramètres. Il permet de calculer les densités par mètre carré des fruits infectés, exposés et susceptibles. Ce modèle a été initialement paramétré grâce aux données (comptages de fruits infectés et sains) obtenues en 2014 et à la littérature. Il est de la forme :

$$\begin{cases} \frac{dS}{dt} &= \eta * E - \lambda * I * S \\ \frac{dE}{dt} &= \lambda * I * S - \eta * E - \sigma(t) * E \\ \frac{dI}{dt} &= \sigma(t) * E - \rho * I \end{cases} \quad (3.1)$$

Avec :

$$\sigma(t) = \begin{cases} 0 & \text{if } W(t) < fw_crit \\ \gamma_m * W(t) & \text{if } W(t) > fw_crit \end{cases} \quad (3.2)$$

$$W(t) = \frac{w0 * wf}{w0 + (wf - w0) * exp(-k * (t - t_bloom))} \quad (3.3)$$

Les 8 paramètres étudiés sont le poids du fruit à floraison ($w0$), le poids maximum du fruit (wf), le taux de conversion de ressource en masse de fruit (k), la masse fraîche de fruit qui représente le seuil de fissuration de la cuticule (fw_crit), la durée moyenne de la viabilité des spores à la surface du fruit (η), le taux de transmission par unité infectieuse (λ), une constante d'infection (γ_m) et une période infectieuse moyenne (ρ). Le modèle simule l'évolution de la maladie pendant une période donnée. (Voir annexe 1)

Pour lancer les simulations, le modèle a besoin de la période où l'on veut simuler. Du jour de pleine floraison (t_bloom) et enfin des valeurs initiales pour la densité S , E et I . Ces valeurs initiales sont calculées comme suit :

$$\begin{cases} E(t_0) &= \frac{Nb \text{ total fruit}}{Area \text{ arbre}} * iFR E \\ I(t_0) &= \frac{Nb \text{ total fruit}}{Area \text{ arbre}} * iFR I \\ S(t_0) &= \frac{Nb \text{ total fruit}}{Area \text{ arbre}} * iFR S \end{cases} \quad (3.4)$$

Avec iFR correspondant à la répartition initiale des densités. Cette répartition est $iFR S=0.98$, $iFR E= 0.018$ et $iFR I = 0.002$.

3.2 Modèle adapté aux micro-fissures (cracks)

On reprend ici le modèle mis en place par Daniele et on va essayer de remplacer la courbe de croissance du fruit par la surface de crack au cours du temps. On fait cela car la monilinia va infecter le fruit en entrant par ces cracks et il est donc plus judicieux de considérer les cracks comme proxy de l'infection. On peut donc penser que cette surface va bien modéliser l'infection.

Pour se faire, on va modifier sigma qui est le taux auquel un fruit exposé à l'agent pathogène passe dans le comportement infectieux à l'instant t. On va le remplacer dans les calculs des exposés(E) et des infectés(I). On met à la place la surface de crack (Sc(t)) en fonction du temps multiplié par une nouvelle constante d'infection. On retire la condition sigma car le crack va avoir une valeur de 0 au début et augmenter avec le temps.

$$\left\{ \begin{array}{l} \frac{dS}{dt} = \eta * E - \lambda * I * S \\ \frac{dE}{dt} = \lambda * I * S - \eta * E - \gamma_c * Sc(t) * E \\ \frac{dI}{dt} = \gamma_c * Sc(t) * E - \rho * I \end{array} \right. \quad (3.5)$$

Le modèle modifié a ainsi quatre paramètres qui sont la durée moyenne de la viabilité des spores à la surface du fruit (eta), le taux de transmission par unité infectieuse (lambda), une constante d'infection (gamma c) et une période infectieuse moyenne (rho).

Le modèle adapté aux cracks s'exécute après une initialisation similaire à celle de modèle original. On résume les paramètres des deux modèles dans un glossaire.(Voir annexe 1)

3.3 Jeux de données

Afin de réaliser ce stage, j'ai eu à ma disposition différents jeux de données. Tous d'abord deux jeux de données contenant les données expérimental issues de manipulations réalisés sur le terrain par les chercheurs.

Le premier jeu contient les valeurs pour 2014 ainsi que les dates correspondantes en jour après floraison. On a 15 observations avec les valeurs pour les fruits « infectés » et les fruits sains qui correspondent à la somme des fruits « susceptibles » et « exposés ».

Le deuxième jeu de données contient les données pour 2015. On a ici 13 observations avec à chaque fois le temps en jour après floraison, les fruits « infectés » et les fruits sains correspondants. On a ici des dates différentes du fichier de 2014.

Les valeurs des paramètres du modèle original ont été obtenue dans la littérature ou grâce à la station météo de l'INRA ainsi qu'avec deux vergers (Alexandra qui a des fruits avec une maturation rapide et Suncrest qui a des fruits avec une maturation lente).

Chapitre 4

Méthodologie et approches

4.1 Analyse de sensibilité

Afin de me familiariser avec le modèle existant, ma première tâche fût de réaliser une étude de sensibilité sur les paramètres qui avait été trouvés pour le modèle. Pour ce faire, j'ai appliqué trois méthodes d'analyse de sensibilité qui sont la méthode de Morris, la méthode de Sobol et la méthode Fast.

Ces méthodes sont très répandue aujourd'hui et servent pour analyser ou simplifier des modèles. Grâce à elles, on va déterminer dans quelles mesures les facteurs d'entrée d'un modèle vont influencer sur les sorties issues de simulations.

On peut différencier trois groupes de méthodes d'analyse de sensibilité :

- Les méthodes de screening, qui vont réaliser une analyse qualitative de la sensibilité de la variable de sortie aux variables d'entrée
- Les méthodes locales qui vont s'intéresser aux petites perturbations autour d'une valeur de variable d'entrée qui se répercutent sur la variable de sortie
- Les méthodes globales qui vont étudier la variabilité de la sortie du modèle dans l'intégralité de son domaine de variation

Les valeurs de référence pour ces paramètres ainsi que leurs intervalles de variation sont ceux estimés par (Bevacqua et al., 2017).

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeurs min	0.12	193	6.4*10-2	52	0.004	3.9*10-2	0.4*10-3	4.0*10-2
Valeur	0.17	201	6.7*10-2	61	0.13	5.8*10-2	1.5*10-3	5.2*10-2
Valeurs max	0.21	210	7.1*10-2	70	0.468	37*10-2	2.4*10-3	36.2*10-2

FIGURE 4.1 – Tableau des valeurs de référence et intervalles de variation des paramètres

4.1.1 Méthode de Morris

Cette méthode a été introduite par Morris (1991) et s'appuie sur une discrétisation de l'espace des variables, c'est-à-dire que l'exploration de la variable de sortie est réalisée sur une grille régulière des variables d'entrées.

Le but de cette méthode est de faire varier la valeur d'un facteur en laissant les autres à une valeur fixe afin de détecter l'effet de ce paramètre sur la sortie du modèle. Ce genre de méthode est appelé One-At-a-Time (OAT). La sensibilité de la sortie à une des variables X_i est mesurée en comparant les

résultats où seule cette variable a été variée.

On peut aussi réaliser sur différentes combinaisons de valeurs des différents facteurs incertains. Cela permet de voir les interactions entre facteurs.

Cette méthode va avoir différents objectifs qui vont être d'identifier les facteurs influents et de classer les facteurs selon leur influence. On va avoir pour cette méthode un modèle de la forme $Y = M(X_1, \dots, X_K)$ avec K facteurs.

On applique cela sur une grille régulière superposée sur le domaine de définition du modèle.

Des effets élémentaires, variations de Y , sont calculés au moyen de trajectoires définies sur les nœuds de la grille en faisant varier chaque facteur une seule fois sur chaque trajectoire

La trajectoire est définie par un ensemble de nœuds tel que seul un paramètre varie entre deux nœuds successifs. Ainsi, une trajectoire comporte $k + 1$ nœuds (en comptant le nœud de départ). Le nœud de départ est choisi aléatoirement ainsi que les paramètres qui varient entre deux nœuds successifs. Pour une construction de r trajectoires, le nombre de simulations nécessaire est de $r * (k + 1)$.

On construit ainsi différents scénarios balayant "toutes" les valeurs possibles pour les différents facteurs. On va calculer pour ces différents scénarios les effets élémentaires. Ainsi pour chaque facteur on obtient une distribution de valeurs pour chaque scénario, on va ainsi calculer des indicateurs de moyennes et d'écart type pour chaque facteur.

$X_i, i = 1, \dots, d$
 $\delta = \text{longueurdupasdanslagrille}$
 $P(k)_i = \text{Trajectoire}$
 $r = \text{nombredeTrajectoire}$
 $k = \text{numrodeTrajectoire}$
 $\pi_k = \text{permutationdel'ordredesfacteurspourlatrajectoirek}$

Effet élémentaire :

$$\Delta_i^k = \frac{M(P_{\pi_k(i-1)}^k) - M(P_{\pi_k(i)}^k)}{\delta} \quad (4.1)$$

Effet moyen :

$$\mu_i = \frac{1}{r} * \sum_{k=1}^r \Delta_i^k \quad (4.2)$$

On regarde ensuite μ^* (Campolongo et al 2007) qui est une mesure améliorée de μ et permet de déterminer la sensibilité de la variable de sortie aux variables d'entrée et de classer ces dernières par rapport à leurs importances sur la variable de sortie.

$$\mu_i^* = \frac{1}{r} * \sum_{k=1}^r \|\Delta_i^k\| \quad (4.3)$$

Ecart-type des effets élémentaires :

$$\sigma_i = \sqrt{\frac{1}{r} * \sum_{k=1}^r (\Delta_i^k - \mu_i)^2} \quad (4.4)$$

Δ_i^k est la variation de Y sur la trajectoire k ($i = 1, \dots, d$) relative à la variation du facteur X_i .

Une analyse du graphique de Morris (μ_i * en abscisse et σ_i en ordonnée) permet de distinguer :

- Les facteurs dont les effets sont négligeables : points proches de l'origine (0,0)
- Les facteurs dont l'effet linéaire est important : points situés à droite sur l'axe des abscisses
- Les facteurs dont les effets sont non-linéaires ou incluent de l'interaction avec d'autres facteurs : points situés en haut sur l'axe des ordonnées

Avec les indicateurs calculés, un facteur possédant une forte moyenne est considéré comme ayant une forte importance sur la sortie du modèle, une forte valeur de la variance montrera une interaction entre le facteur considéré et un autre ou alors un facteur dont l'effet est non linéaire.

4.1.2 Méthode de Sobol

La méthode de Sobol est basée sur une analyse fonctionnelle de la variance. Sobol (1993) propose de décomposer la fonction f du modèle en somme de fonctions de dimension croissante.

Pour cette méthode, on va avoir une copie indépendante de X_{-i} qui va être X'_{-i} . On va avoir un modèle de la forme $Y = M(X_i, X'_{-i})$.

A partir de cela, on va créer 2 échantillonnages de Monte Carlo ou LHS (Latin Hypercube Sampling) qui vont être appelé A et B. Ces deux échantillons seront de taille n.

$$A = \begin{pmatrix} X_{1,1}^A & \dots & X_{d,1}^A \\ \vdots & \ddots & \vdots \\ X_{1,n}^A & \dots & X_{d,n}^A \end{pmatrix} \quad B = \begin{pmatrix} X_{1,1}^B & \dots & X_{d,1}^B \\ \vdots & \ddots & \vdots \\ X_{1,n}^B & \dots & X_{d,n}^B \end{pmatrix} \quad (4.5)$$

On crée d matrices d'échantillonnage C_i à partir de A et B.

$$C_i = \begin{pmatrix} X_{1,1}^A & \dots & X_{i,1}^B & \dots & X_{d,1}^A \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{1,n}^A & \dots & X_{i,n}^B & \dots & X_{d,n}^A \end{pmatrix} \quad (4.6)$$

On évalue ensuite $(2 + d) * n$ combinaisons des facteurs sur le modèle M.

$$y^A = \begin{pmatrix} y_1^A \\ \vdots \\ y_n^A \end{pmatrix} \quad y^B = \begin{pmatrix} y_1^B \\ \vdots \\ y_n^B \end{pmatrix} \quad (4.7)$$

On calcule ensuite les variances permettant d'obtenir l'indice de Sobol.

$$\widehat{V}_i = \frac{1}{n} * \sum_{k=1}^{n-i} y_k^B * (y_k^{C_i} - y_k^A) \quad V - \widehat{V}_{-(i)} = \frac{1}{n} * \sum_{k=1}^n y_k^A * (y_k^A - y_k^{C_i}) \quad (4.8)$$

On utilise ensuite \widehat{V} comme estimateur de la variance totale $V(Y)$ pour obtenir l'indice de Sobol d'ordre 1 et celui d'indice total.

$$\widehat{S}_i = \frac{\widehat{V}_i}{\widehat{V}} \quad (4.9)$$

$$\widehat{S}_{Ti} = \frac{V - \widehat{V}_{-(i)}}{V} \quad (4.10)$$

4.1.3 Méthode FAST

La méthode FAST (Fourier Amplitude Sensitivity Testing) est basée sur les principes de l'analyse de Fourier (Saltelli et al, 1998). Similaire à la méthode de Sobol, la méthode FAST décompose la variance de Y et utilise des indices de sensibilité. L'idée de cette méthode est d'utiliser la transformée de Fourier pour effectuer la décomposition.

On va ici utiliser $G(x)$ défini et intégrable sur le domaine $D = D_1 * \dots * D_d$ avec :

$$G(x) = f_0 + f_1 * (x_1) + \dots + f_d * (x_d) + f_{1,2} * (x_1, x_2) + \dots + f_{d-1,d} * (x_{d-1}, x_d) + \dots + f_{1,\dots,d} * (x_1, \dots, x_d) \quad (4.11)$$

On a pour la variance :

$$Var(Y) = \sum_{u \in \{1, \dots, d\}} V_u \quad (4.12)$$

où

$$V_u = \sum_v |c_v(G)| \quad (4.13)$$

Avec :

$$|c_v(G)| = \int_D G(x) * \exp(-i * 2 * \pi * x_j * \sum_{j=1}^d V_j) * dx \quad (4.14)$$

L'échantillonnage est effectuée dans $D = [0, 1]^d$ en parcourant le champ des fréquences possibles, avec $v = (v_1, \dots, v_d)$ vecteur de fréquences, pour les entrées X_j .

Pour un v donné, on définit la composante spectrale d'une trajectoire (D_v), composée d'une suite de points $x_j = (x_{j,1}, \dots, x_{j,d})$, $j = 1, \dots, N$:

$$D_v = \left| \frac{1}{n} \sum_{j=1}^N G(x) * \exp(-i * v * u_j) \right|^2 \quad (4.15)$$

Avec : $u_j = -\pi + 2 * \pi * (j - 1/2)/N$: Suite de valeurs dans $[-\pi, \pi]$ équi-espacées.
Les composantes des points x_j sont choisies telles que $x_{j,i} = g(\sin(w_i * u_j + \phi_i))$

w_i : fréquence des répétitions des $x_{j,i}$

ϕ_i : déphasage permettant de varier les trajectoires

g : fonction à déterminer selon la distribution. Pour la loi $U(0, 1)$, g est la fonction :

$$x \rightarrow \frac{1}{2} + \frac{1}{\pi} * \arcsin(x) \quad (4.16)$$

On obtient donc les indices.

$$\widehat{S * I}_i = 1 - \frac{\sum_{m=1}^M D_{mwk}}{Var(Y)} \quad (4.17)$$

$$T * \widehat{S * I}_i = 1 - \frac{\sum_v D_v}{Var(Y)} \quad (4.18)$$

4.2 Calibration des modèles

4.2.1 Formulation du problème

L'objectif principal ici est d'estimer les paramètres du modèle pour les rapprocher de nos courbes de données réelles. Pour ce faire, on cherche généralement à minimiser le RMSE (Root Mean Square Error) qui calcule la différence entre valeurs observées et simulées. Plus la valeur de RMSE se rapproche de 0, plus la distance entre les observations et les prédictions sera petite et donc plus nos prédictions seront précises. Le RMSE sera calculé avec la formule :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (observes_i - simules_i)^2}{n}} \quad (4.19)$$

On cherche un ensemble de paramètres qui minimisent séparément les RMSE I et RMSE SE :

$$\min_X RMSE_I \quad (4.20)$$

$$\min_X RMSE_{SE} \quad (4.21)$$

On veut comparer des RMSE avec différentes unités. On va donc utiliser le NRMSE (Normalized root-mean-square error). Il existe deux approches : normaliser le RMSE à la gamme des données observées ou normaliser à la moyenne des données observées.

$$NRMSE = \frac{RMSE}{observes_{max} - observes_{min}} \quad (4.22)$$

On cherche donc à estimer les valeurs des paramètres X pour minimiser le NRMSE pour les fruits infectés d'une part et les fruits susceptible et exposés d'autre part :

$$\min_X (RMSE_I, RMSE_{SE}) \quad (4.23)$$

4.2.2 Choix algorithmique

Les erreurs NRMSE ci-dessous sont fortement non linéaires en fonction des paramètres à estimer. Les modèles utilisés sont numériques, ce qui signifie que l'on ne dispose pas des formules mathématiques analytiques explicites de nos différents NRMSE en fonction de nos paramètres. Toutes ces difficultés font de ce problème un problème d'optimisation dit « difficile » ce qui veut dire qu'on ne connaît pas d'algorithme exact permettant de repérer un optimum global (i.e. la meilleure solution possible) à coup sûr et en un nombre fini de calculs. Deux grandes familles d'approches d'optimisation peuvent être distinguées : déterministes et stochastiques. Bien que les approches déterministes aient été utilisées avec succès ; Ils montrent des performances médiocres pour de nombreux problèmes. Les approches stochastiques sont considérées comme plus souples et plus efficaces pour résoudre des problèmes d'optimisation importants et complexes qui sont non différentiables, multimodaux. Deux principaux types d'approches stochastiques pourraient être distingués : les heuristiques et les métaheuristiques. Les métaheuristiques n'assument aucune propriété mathématique du problème traité. Un panel très large de métaheuristiques a été proposé dans la littérature. Nous pouvons citer à titre d'exemple les algorithmes génétiques, le recuit simulé, la recherche taboue, ou les essais particuliers. Ces algorithmes peuvent fournir des solutions de bonne qualité en un temps calcul raisonnable sans pour autant garantir l'optimalité.

Ainsi, pour calibrer nos deux modèles, j'ai eu recours aux algorithmes génétiques qui sont des algorithmes stochastiques s'inspirant de la théorie de l'évolution. J'ai donc utilisé deux algorithmes génétiques différents bien adaptés à nos besoins (Quilot et al. 2012). Ces algorithmes imitent bien les processus biologiques de la reproduction et la sélection naturelle pour résoudre et trouver les solutions les plus adaptées. Ils utilisent des techniques dérivées de la génétique et des mécanismes d'évolution de la nature : croisements, mutations, sélections, etc. (Kikuchi et al. 2003) (Holland 1975)

Les deux algorithmes que j'ai utilisés sont `ga` et `rbga` disponibles respectivement sous R dans les packages `GA` et `genalg` (Willighagen 2015).

4.2.3 Calibration des modèles sur les données des fruits infectés et sains simultanément

Dans un second temps, pour améliorer la prédiction de nos modèles, on a essayé de calibrer sur les données observées 2014 et 2015 des fruits infectés et sains en même temps.

Ceci consiste à minimiser simultanément les erreurs quadratiques moyennes normalisées pour les fruits infectés ($NRMSE_I$) d'une part et les fruits sains ($NRMSE_{SE}$) d'autre part.

On peut donc formuler le problème comme suit :

$$\min_X(NRMSE_I, NRMSE_{SE}) \quad (4.24)$$

Pour résoudre ce problème d'optimisation difficile, on utilise l'algorithme génétique multicritère. Il s'agit de NSGA-II (Nondominated Sorting Genetic Algorithm 2), un algorithme génétique multi-objectif développé par K. Deb (2002). Cet algorithme a été choisi car il est considéré comme l'un de meilleurs algorithmes évolutionnaire multi-objectif existants. Une implémentation de cet algorithme est disponible sous R dans le package `mco`. Le pseudo-code de NSGA-2 est proposé ci-après :

- Initialiser aléatoirement la population
- Les chromosomes sont triés et mis en avant sur un ensemble de Pareto non dominés. Dans le front de Pareto, les chromosomes sont classés sur la base euclidienne entre les solutions ou I-dist (terme utilisé dans NSGA-II). Généralement, les solutions éloignées des autres solutions reçoivent une préférence plus élevée lors de la sélection. Ceci est fait afin de créer une solution diversifiée et éviter un ensemble de solutions encombrées
- Les meilleurs chromosomes N sont prélevés dans la population actuelle et mis dans une réserve d'accouplement
- Dans la réserve d'accouplement, la sélection du tournoi, la traversée et l'accouplement sont effectués
- La réserve d'accouplement et la population actuelle sont combinées. L'ensemble résultant est trié, et les meilleurs chromosomes N font la nouvelle population

- Passez à l'étape 2, sauf si le nombre maximal de générations a été atteint
- L'ensemble de solutions est l'ensemble non dominé de Pareto le mieux classé de la dernière population

On peut résumer cela dans le graphique suivant.

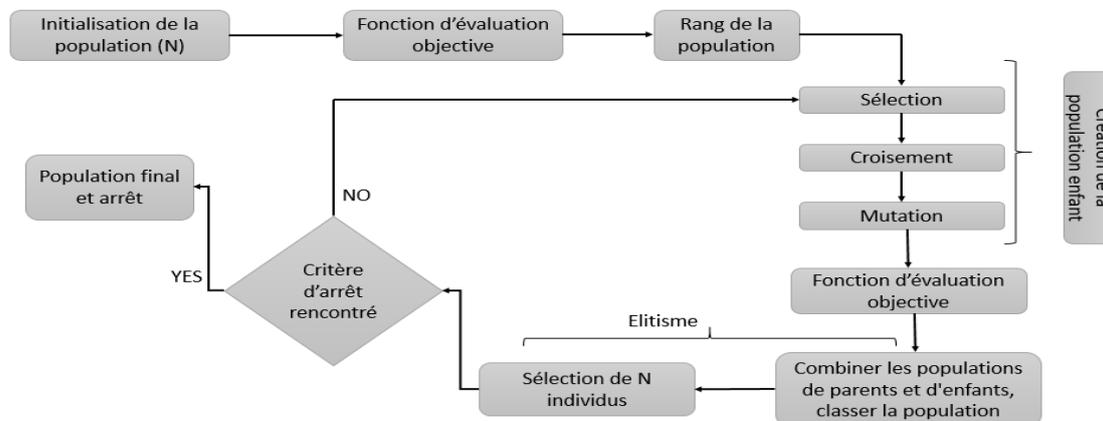


FIGURE 4.2 – Diagramme du fonctionnement de NSGA-2

4.3 Intégration des modèles dans Qualitree

d'intégrer le modèle épidémiologique comme module élémentaire dans le modèle QualiTree et de faire l'optimisation multicritère pour concevoir des scénarios de conduite de vergers économes en intrants, économiquement viables, et respectant les contraintes particulières de la filière.

Après avoir fait cela, on va intégrer le modèle de Daniele à Qualitree afin d'améliorer Qualitree pour qu'il puisse sortir les densités d'infection pour les pêches au cours du temps.

4.3.1 Présentation de Qualitree

QualiTree est un modèle écophysiological guidé par les processus d'assimilation et de gestion du carbone, qui prend en compte l'effet de techniques culturales (taille, irrigation et éclaircissage). Il prédit la variabilité intra-arbre de la qualité des fruits (calibre, teneurs en sucres, taux de matière sèche, surface de fissures sur l'épiderme du fruit (porte d'entrée des spores de monilia), etc) (Lescourret et al. 2010).

Qualitree va décrire l'arbre comme un ensemble d'objets composé des fruits, des branches et du tronc, des feuilles, etc.. Dans notre cas, on va surtout regarder la partie sur les fruits.

Chaque exécution de Qualitree génère quatre fichiers :

- Un fichier log pour pouvoir déterminer quels paramètres, temps ou encore fonctions ont été utilisés pour les simulations
- Un fichier tree avec toutes les données obtenues sur l'arbre simulé
- Un fichier br avec toutes les données obtenues sur les branches simulées
- Un fichier pu avec toutes les données obtenues sur les unités de production simulées donc les fruits

4.3.2 Insertion dans Qualitree

Une fois les deux modèles épidémiologiques bien calibrés, la dernière tâche de mon stage devrait s'intéresser à l'intégration de ces deux modèles dans Qualitree. Cela devrait améliorer le modèle Quali-tree dans le sens où les fruits tombés après infection seront pris en compte. La charge en fruits sera mise à jour à chaque fois où un fruit infecté tombe. En outre, les sorties de Qualitree qui étaient nécessaires

pour faire tourner les modèles épidémiologiques ne le seront plus car les deux modèles deviennent partie intégrante de Qualitree.

Pour les simulations, on va ainsi avoir trois temps de stress hydrique où l'on peut faire un stress ou non selon trois charges de fruit sur les arbres (low = 53 fruits par arbre, com = 157 fruits par arbre et high = 471 fruits par arbre). On va ainsi simuler selon les différentes combinaisons (54) et étudier les résultats.

Nos deux modèles ont été codés à l'intérieur. On a ainsi utilisé le modèle Qualitree ainsi enrichie sur les serveurs de l'INRA car le modèle Qualitree est très complexe et demande beaucoup de puissance pour une simulation. On a lancé 54 simulations en précisant à chaque fois les paramètres utilisés, le modèle que l'on veut (masse ou crack), la période de temps, les méthodes utilisées et enfin les paramètres sur l'arbre étudié (M.Miras-Avalos et al.2011).

On a ensuite étudié et comparer les résultats obtenus.

4.3.3 Optimisation des scénarios

Dans la partie précédente, nous avons simulé sur Qualitree 54 scénarios prédéfinis dans les bases de données de Qualitree. Ces scénarios résultent de la combinaison de trois niveaux de charges en fruits et 18 régimes d'irrigation différents. Dans cette partie, on va créer des scénarios virtuels en gardant les mêmes niveaux de charge en fruit mais en jouant sur le régime d'irrigation.

Pour cela, on crée une fonction permettant de générer une courbe de stress hydrique. A partir d'une valeur de potentiel hydrique de départ (au collet tôt le matin), on calcule le potentiel hydrique minimal (mi-journée) en multipliant la valeur de départ par un ratio qui permet de contrôler la sévérité du stress hydrique. Ceci nous permet de générer un ensemble de courbes sinusoïdales en faisant varier notre potentiel de départ et notre valeurs de ratio qui varient chacun dans un intervalle donné.

On a couplé l'algorithme NSGA-2 et Qualitree pour trouver les meilleurs scénarios hydriques permettant à la fois de minimiser la densité de fruits infectés et maximiser la densité des fruits sains.

Chapitre 5

Résultats

5.1 Résultats de l'analyse de sensibilité

La première tâche que j'ai réalisé dans le cadre de mon stage est l'analyse de sensibilité du modèle original. Pour cela, on a utilisé 3 méthodes différentes d'analyse de sensibilité (Morris, Sobol et Fast). L'utilisation de ces trois méthodes a pour objectif de supprimer le biais lié au choix de la méthode d'analyse. Partant du constat que l'effet de certains paramètres sur la dynamique de sortie de modèle (masse ou crack) pourrait être plus ou moins important suivant la date à laquelle l'analyse est menée, nous avons choisis d'analyser le modèle à trois dates (11 ;22 ;33 jours après le début de simulation). Les paramètres influants sur une sortie du modèle ou une autre peuvent être différents. Nous avons ainsi analysé trois sorties différentes correspond aux densités de fruits infectés, susceptibles, exposés. L'objectif de notre démarche est d'identifier le nombre maximum de paramètres ayant un effet important sur les sortie d'intérêt de notre modèle.

Les trois méthodes utilisés ont été paramétrées comme suit :

- 1) Morris : Le nombre de répétitions est de 8 (égal au nombre de paramètres)
- 2) Sobol : La taille de l'échantillon est de 1000 et le nombre de répétitions est de 20
- 3) Fast : La taille de l'échantillon est de 1000

Ces trois méthodes ont été appliquées au modèle via le package sensitivity de R.

On observe ainsi différents résultats :

On observe que pour les trois méthodes les paramètres lambda et gamma ont un effet très fort sur les trois sorties. On voit aussi que eta a un effet fort dans la plupart des cas.

Lors de l'analyse de morris, on observe que k va avoir un effet assez fort surtout pour les fruits susceptibles. Cependant, on ne retrouve pas cet effet lors des autres analyses. On a donc décidé d'exclure ce paramètre.

S temps 11

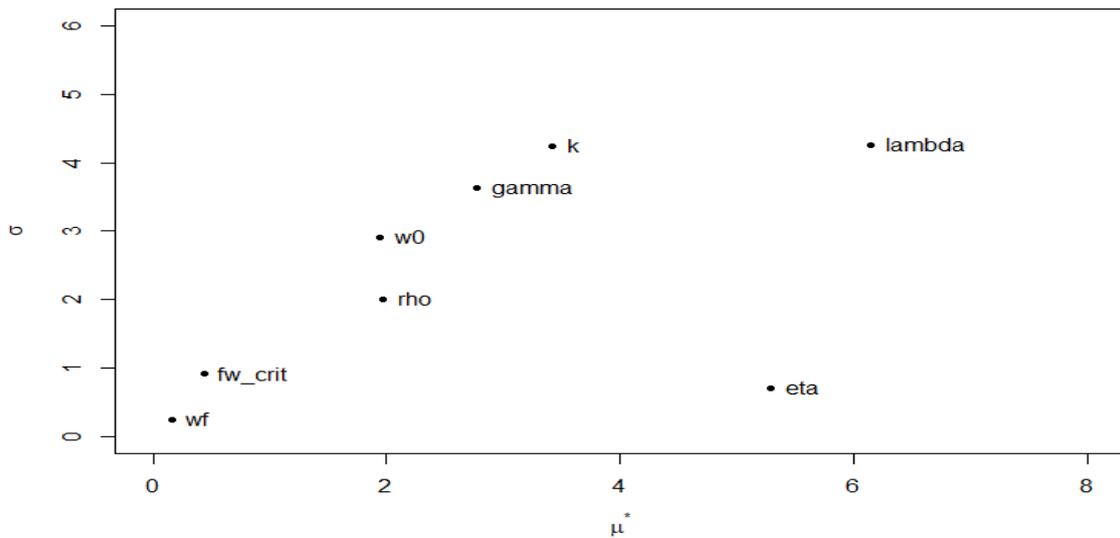


FIGURE 5.1 – l’analyse de sensibilité sur les « susceptible » au temps 11 avec morris

Enfin, on observe dans les trois analyses que rho va avoir un effet de plus en plus fort au cours du temps pour les fruits infectés.

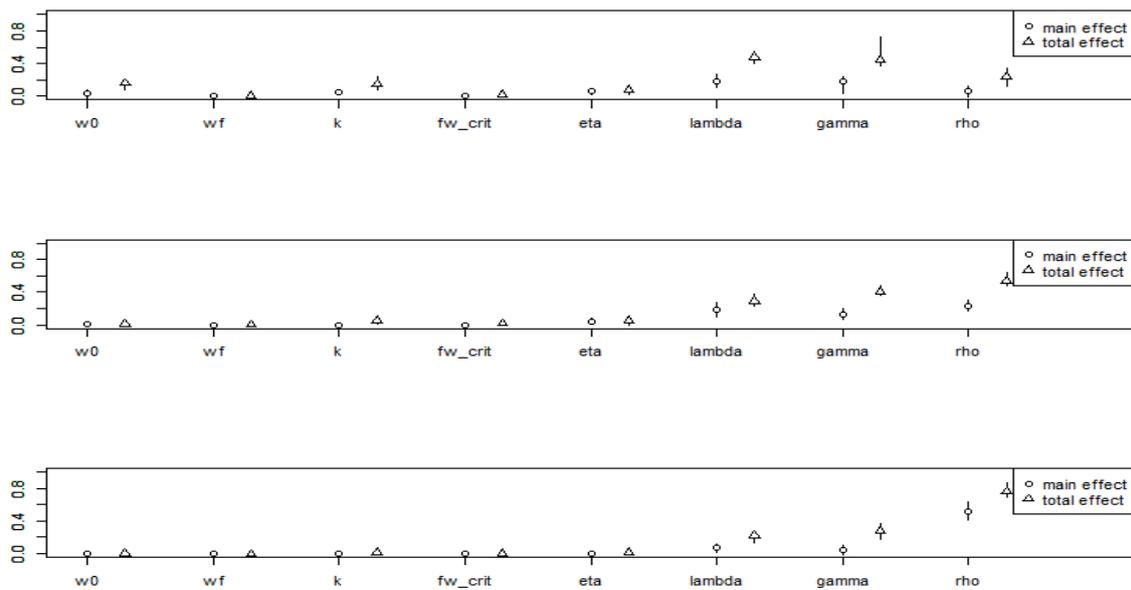


FIGURE 5.2 – l’analyse de sensibilité sur les « infectés » avec Sobol

5.2 Résultats des estimations et optimisation

Les jeux de données disponibles consistent en comptages de fruits infectés d’une part et de fruits susceptibles ou exposés d’autre part. Donc, on a deux ensembles de données par année (fruit infecté I et fruit susceptible ou exposé SE).

Après analyse de sensibilité du modèle original et identification des paramètres influents sur les sorties d’intérêt, nous avons essayé de calibrer nos deux modèles sur ces deux ensembles de données. Pour ce faire, on a procédé en trois étapes :

1) Minimisation des erreurs quadratiques pour les fruits infectés et pour les fruits sains séparément (comme dans 4.20 4.21) afin de trouver les meilleurs paramètres pour le modèle prenant en entrée la courbe de croissance. On réalise cela avec les algorithmes génétiques ga et rbga respectivement des packages GA et genalg de R.

2) On a cherché à trouver les paramètres permettant d'obtenir le meilleur compromis entre la minimisation du RMSE des fruits infectés et du RMSE des fruits sains (comme dans 4.23). On utilise ici l'algorithme NSGA-2 et on fait cela pour le modèle original (courbe de croissance en entrée).

3) On regarde ensuite pour trouver les paramètres permettant d'obtenir le meilleur compromis entre la minimisation du NRMSE des fruits infectés et du NRMSE des fruits sains (comme dans 4.24). On utilise ici l'algorithme NSGA-2 et on fait cela pour les deux modèles. L'utilisation de NRMSE à la place de RMSE permet de s'affranchir de la différence d'échelles de variations de RMSE observées dans les deux étapes précédentes.

On prend pour paramètre initial le temps de jours 153 au 196. On a le jour de pleine floraison qui est le 59 et enfin des valeurs initial pour la densité S, E et I qui sont respectivement 14, 0.1639 et 0.0281.

5.2.1 Test préliminaire

Dans cette étape, on a cherché à minimiser les RMSE pour les fruits infectés et les fruits susceptibles ou exposés séparément. On a commencé par regarder les résultats obtenus avec les données de l'année 2014 puis de l'année 2015 et enfin avec les deux jeux de données (2014+2015).

On a lancé les deux algorithmes ga et rbga avec les mêmes valeurs de taille de population (100) et de nombre de générations (200). Les résultats d'estimation obtenus par chaque algorithme étaient meilleurs pris par année que par jeux de données fusionnés (2014+2015). Notre choix était cependant de garder le paramétrage obtenu avec les données fusionnées pour prendre en compte la variabilité climatique de plus en plus préoccupante. Une autre raison de ce choix est l'écart entre les valeurs estimées par année (deux jeux de valeurs différents). Or les paramètres se doivent d'affranchir un peu de l'année considérée. Le choix des valeurs estimées avec données fusionnées est plus judicieux se rapprochant plus précisément de la réalité.

Les résultats résumés dans le tableau ci-dessous montrent bien que l'algorithme rbga a de meilleures performances :

Algorithme	ga	rbga
RMSE I	0.80	0.69
RMSE SE	5.60	5.08

TABLE 5.1 – Résultats ga et rbga pour RMSE I et RMSE SE

La figure ci-dessous montre l'évolution des RMSE au cours de génération de rbga pour les fruits infectés en considérant les données 2014+2015 (Voir annexe 7 pour plus de détails sur les performances de rbga et ga)

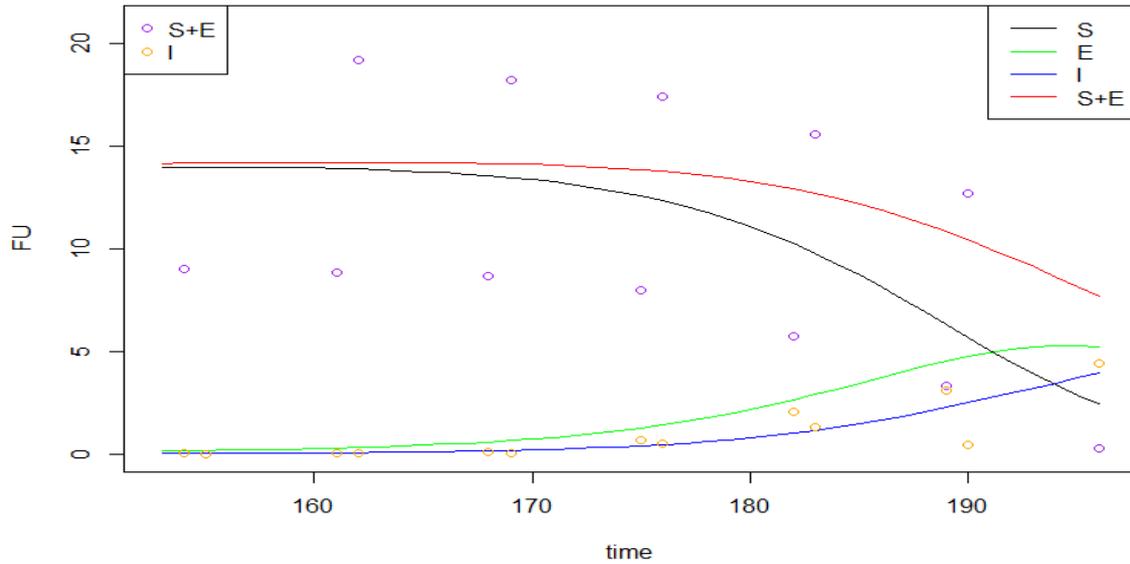


FIGURE 5.3 – Représentation des courbes obtenues pour les paramètres de rbga pour les fruits infectés

La meilleure estimation obtenue pour les paramètres est la suivante :

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeur	0.20694	194.38854	0.06476	55.39013	0.01945	0.04514	0.00044	0.04248

FIGURE 5.4 – Paramètres pour RMSE minimum

Une autre conclusion de cette étape concerne le constat qu'un antagonisme entre RMSE I et RMSE SE est assez clair ce qui veut dire que minimiser l'un revient dans certains cas à maximiser l'autre.

5.2.2 Minimisation de deux RMSE simultanément

Après avoir fait le constat ci-dessus, on a tenté d'optimiser les paramètres pour les deux en même temps pour plus de précision sur les courbes finales et surtout pour trouver l'ensemble de valeurs de paramètres correspondant à des compromis intéressants entre les deux erreurs (RMSE I et RMSE SE). Pour cela, nous avons utilisé l'algorithme génétique NSGA-2 avec taille de population de 100, un nombre de générations de 100, et on répété nos optimisations 10 fois. NSGA-2 nous renvoie un ensemble d'estimations (100) par répétition. La figure ci-dessous illustre ceci avec en abscisses les RMSE SE et en ordonnées les RMSE I. Chaque point correspond à une jeu de paramètres estimé différent.

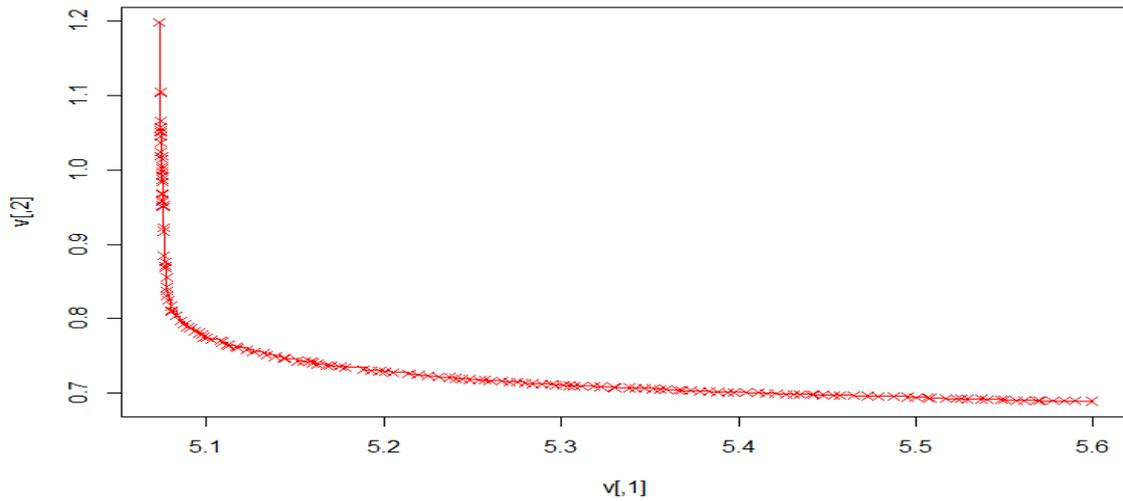


FIGURE 5.5 – RMSE obtenus pour les itérations avec les fruits « sains » en abscisses et « infectés » en ordonnées

Pour identifier la meilleure estimation de paramètres par répétition parmi les 100 obtenus par NSGA-2, on somme les couples de RMSE I et RMSE SE et on sélectionne la ou les solution(s) correspondant à la valeur minimale de la somme.

Par exemple, pour la répétition ci-dessus on a identifié la solution suivante :

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeur	0.12746	201.36260	0.06616	57.16601	0.00474	0.04379	0.00122	0.14601

FIGURE 5.6 – Paramètres pour RMSE minimum

Nous avons répété la procédure ci-dessus pour les 10 résultats des répétitions. L’histogramme des 10 solutions ainsi identifiées est le suivant :

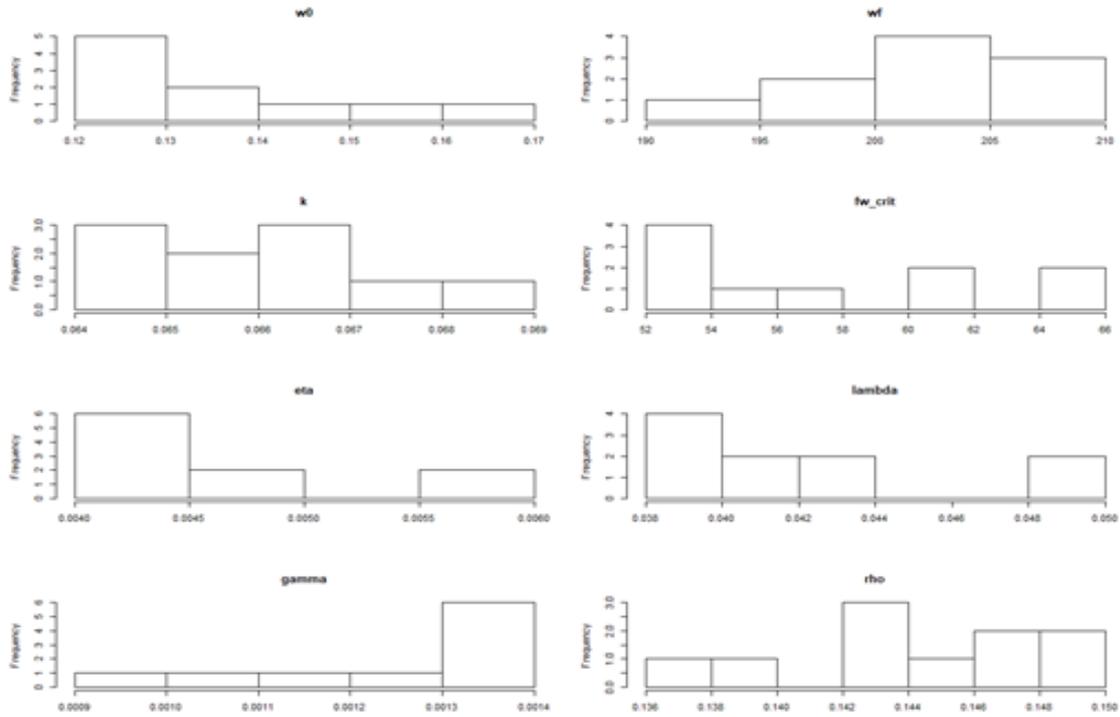


FIGURE 5.7 – Histogramme des paramètres trouvés lors des répétitions

Sur cet histogramme, on observe que certains paramètres tendent vers le minimum ou maximum de leurs intervalles de variations.

On a donc modifié leurs bornes pour élargir leurs domaines de variations. Nous avons ainsi revu à la baisse des bornes inférieures de w_0 , η , fw_crit , λ . La borne supérieure de γ a été revue à la hausse.

Le nouveau domaine de variation de nos paramètres devient donc :

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Minimum	0.08000	193.00000	0.06400	35.00000	0.000004	0.00390	0.00040	0.04000
Maximum	0.21000	210.00000	0.07100	70.00000	0.46800	0.37000	0.02400	0.36200

FIGURE 5.8 – Nouveaux intervalles de variation de paramètre

On a refait ensuite les 10 répétitions de NSGA-2 avec le même paramétrage (taille de population et nombre de génération) et on a constaté que les valeurs des paramètres obtenus sont plus centrées. (Voir annexe 8)

On a réutilisé la procédure de sélection des meilleurs solutions décrite plus haut (somme des RMSE I et RMSE SE). La figure suivante illustre la meilleure calibration obtenue.

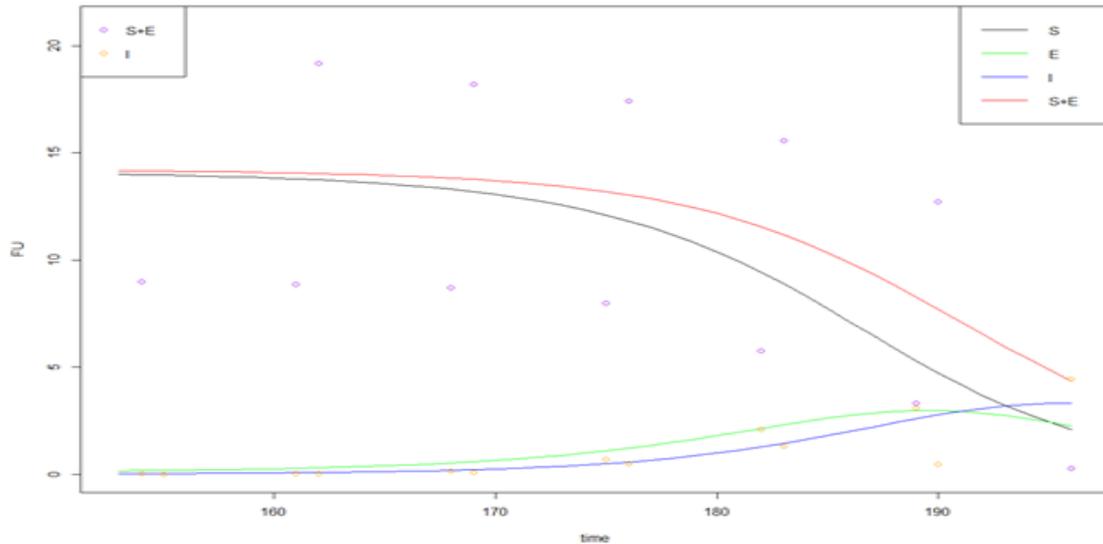


FIGURE 5.9 – Représentation des courbes obtenus avec les paramètres

Cet calibration a été obtenue avec une somme de RMSE de 5.869065 et les valeurs des paramètres correspondant seront donc ceux en annexe.(Voir annexe 8)

5.2.3 Minimisation de deux NRMSE simultanément

On a ensuite repris les études sur les modèles avec pour entrée la courbe de croissance et le modèle avec la surface de crack comme entrée. Cependant, on va cette fois-ci minimiser NRMSE I et NRMSE SE. Comme précédemment, on va regarder sur 2014, 2015 et 2014+2015.

Pour faire cela, on utilise NSGA-2 avec une population de 200 et 150 générations. On va répéter l'algorithme NSGA-2 15 fois. Après cela, on cherche à retirer les points qui sont dominés.

Pour ce faire, on commence par concaténer les matrices des NRMSE et des paramètres. On avait ici 4 paramètres, deux NRMSE et une population de 200. On va ainsi avoir une matrice de dimension $6 * 3000$.

On applique un test de dominance afin d'identifier les points dominés. Un point sera dominé si l'on trouve un autre point qui a des valeurs inférieures pour les deux NRMSE. On applique ce test grâce à la fonction *is_dominated* du package *emoa* de R qui va vérifier si des points sont dominés dans le graphique du front de Pareto.

Ainsi, on applique l'algorithme *is_dominated* sur les 3000 points obtenus lors des simulations et en retirant les points dominés, on observe qu'il reste 1445 points non-dominés.(Voir annexe 9)

On peut maintenant étudier la distribution des paramètres sur les 1445 points restants.

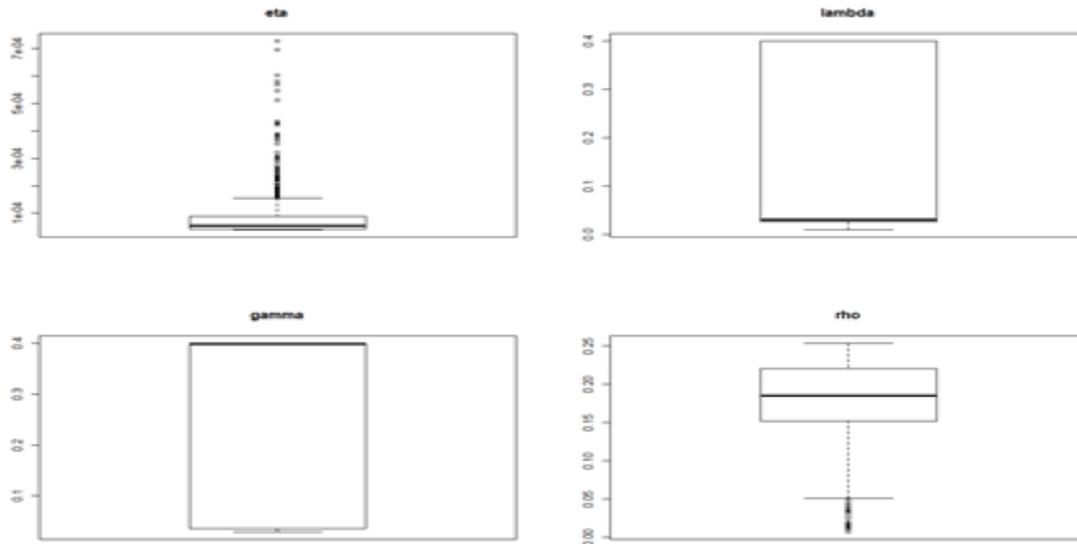


FIGURE 5.10 – Boxplot des paramètres restants

On observe des points isolés surtout pour eta. (Voir annexe 9 pour les résultats numérique)

On s'intéresse ensuite à trois cas :

- 1) On regarde le NRMSE I minimum pour voir quel paramètres se rapprochent le plus des courbes des fruits infectés.
- 2) Le NRMSE SE minimum pour voir quel paramètres se rapprochent le plus des courbes des fruits sains.
- 3) Le minimum de (NRMSE I + NRMSE SE) pour se rapprocher le plus des deux courbes.

Minimiser	NRMSE I	NRMSE SE	NRMSE I + NRMSE SE
Valeur	49.9	82.7	145.7

TABLE 5.2 – Résultats des NRMSE minimum obtenus dans les 1445 points

On observe qu'en minimisant le NRMSE I seulement ou le NRMSE SE seulement, on se rapproche de l'une des deux courbes mais on s'éloigne de l'autre (Paramètres et courbes en annexe 10).

On va donc regarder les paramètres ainsi que les courbes pour la minimisation des deux. On retrouve 176 points dont la somme des NRMSE est sera égales et minimums. Ces 176 correspondent à 176 ensembles de paramètres (décrit numériquement en annexe 10). Pour choisir lequel utiliser, on a regarder lequel se rapproché de la médiane. Après avoir sélectionné nos paramètres :

	<i>eta</i>	<i>lambda</i>	<i>gamma</i>	<i>rho</i>
<i>Valeurs</i>	4.066049e-05	2.737425e-02	3.998037e-01	1.687327e-01

FIGURE 5.11 – Paramètres minimisant SE et I

On regarde les courbes obtenues :

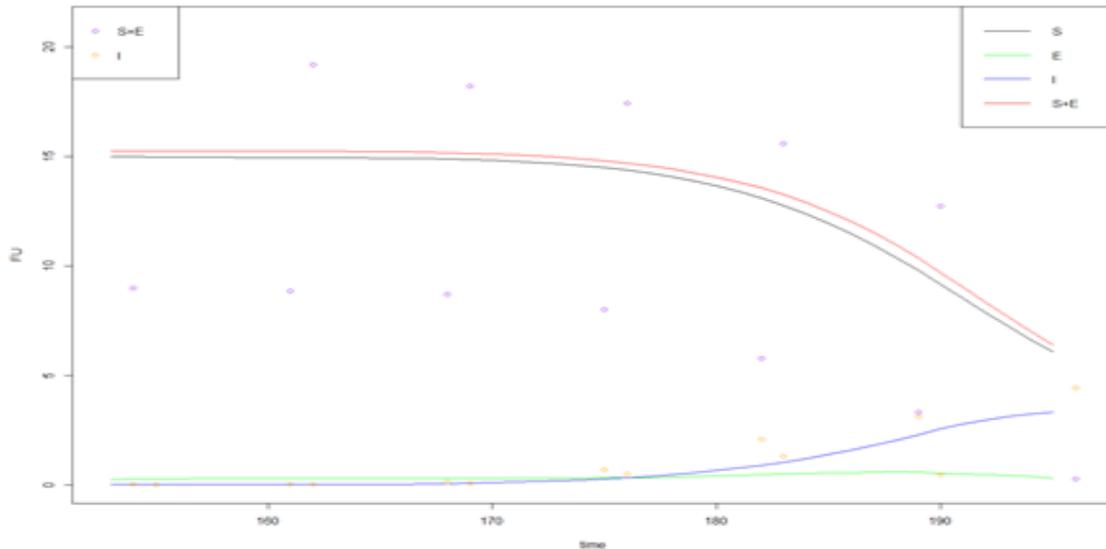


FIGURE 5.12 – Courbes obtenues avec les paramètres pour les I et S+E

On observe que les courbes passent bien au milieu des courbes des fruits sains et celles des fruits infectés et suivent les mêmes trajectoires.

Au final, pour notre modèle, on sélectionne les valeurs des paramètres obtenues ci-dessus pour la minimisation de $\text{NRMSE I} + \text{NRMSE SE}$. On sélectionne aussi les bornes correspondantes au minimum et maximum obtenus pour chaque paramètre dans les 1445 points que l'on a retenus.

	Eta	Lambda	Gamma	Rho
Valeur min	0.0000400	0.01225	0.01035	0.04005
Valeur	0.0000406	0.02737	0.39980	0.16873
Valeur max	0.0060940	0.40000	0.40000	0.18177

FIGURE 5.13 – Tableau des paramètres avec leurs intervalles

5.3 Intégration dans Qualitree

5.3.1 Etude des simulations

On réalise les simulations et on étudie les résultats obtenus. On a ainsi des simulations pour le modèle avec les paramètres de Jose (avant optimisation), le modèle avec la masse et le modèle avec le crack. Pour chacune de ces catégories, on va avoir trois catégories qui vont être low, com et high qui corresponde à la valeurs du nombres de fruits sur l'arbre qui va être soit peu de fruits, des fruits au niveau commercial et beaucoup de fruits. On notera que le modèle avec les paramètres précédent est calibré pour 2014 uniquement. On aura certainement es différence plus importante entre celui-ci et les autres.

On commence par regarder les valeurs obtenues pour les fruits susceptibles :

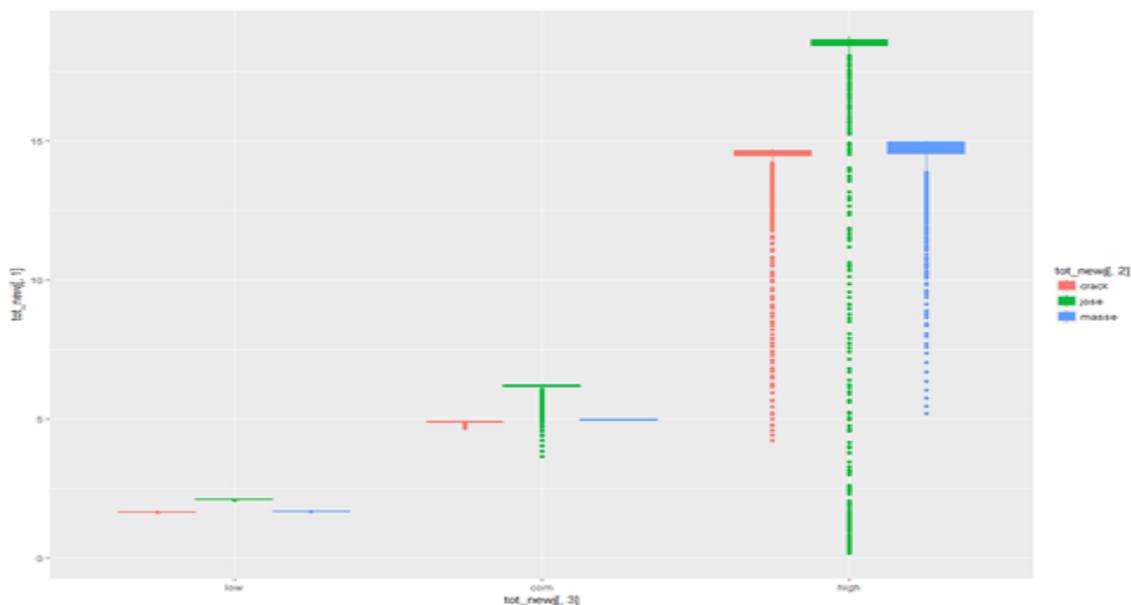


FIGURE 5.14 – Boxplot des fruits susceptibles par nombre de fruit et par simulation

On peut aussi regarder un à un les boxplots pour pouvoir mieux voir les variations. (Annexe 11)

On regarde ensuite les valeurs exactes.

		Low	Com	High
Original	Min	2.075	3.647	0.1815
	1st Qu	2.102	6.169	18.4137
	Median	2.108	6.215	18.5006
	Mean	2.105	6.139	16.8087
	3rd Qu	2.111	6.237	18.6426
	Max	2.113	6.255	18.7568
<hr/>				
Modifié	Min	1.664	4.928	5.20
	1st Qu	1.683	4.968	14.53
	Median	1.686	4.989	14.90
	Mean	1.684	4.980	14.38
	3rd Qu	1.686	4.994	14.98
	Max	1.687	4.996	14.99
<hr/>				
Crack	Min	1.644	4.668	4.235
	1st Qu	1.654	4.877	14.478
	Median	1.655	4.898	14.646
	Mean	1.654	4.874	14.065
	3rd Qu	1.655	4.899	14.651
	Max	1.656	4.905	14.708

FIGURE 5.15 – Indicateurs des fruits susceptibles

On observe des valeurs supérieures pour les simulations de jose. On voit aussi que les valeurs pour le high baissent plus fortement avec le modèle de jose.

On regarde ensuite les fruits exposés et infectés. (Voir annexe 11)

On voit que les valeurs pour les exposés de jose pour high vont augmenter plus fortement que les autres. Au contraire les valeurs pour la masse vont être inférieures aux autres.

Et enfin, on étudie les fruits infectés, on observe que les valeurs de jose pour high vont augmenter plus fortement que les autres. Elles vont aussi être plus fortes pour les low et com.

5.3.2 Optimisation

On cherche à créer de nouvelle courbe de stress hydrique avec des valeurs de départ comprise entre $[-2.4 ; -1]$ et avec un ratio pour obtenir la valeurs minimale compris entre $[1 ; 2]$. On utilise ces bornes pour répondre à des contraintes réaliste comme par exemple ne peut pas atteindre -5.

On utilise ensuite NSGA-2 avec 20 génération et une population de taille 20. On réalise cela pour les trois cas de nombres de fruits. On utilise NSGA-2 pour à la fois minimiser la densités de fruits infectées et maximiser la densité de fruits sains(ou minimiser - la densité de fruits sains). On commence par les arbres avec peu de fruits (low).

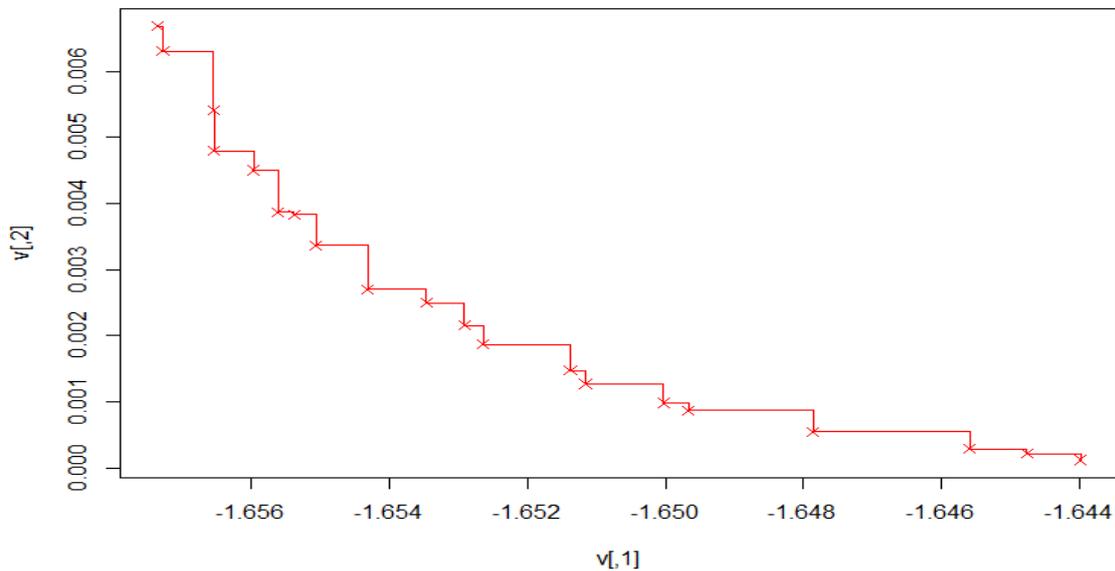


FIGURE 5.16 – Valeurs des densités finales pour les low

On regarde pour obtenir la somme minimal des deux et on a une densité de sain égale à 1.6556 et une densité d'infectés égale à 0.0039. Ces valeurs sont obtenue pour une valeur de départ de -1.0339 et un ratio de 1.9604.

On regarde ensuite pour un nombre de fruit "commercial" (com) (Voir annexe 12). La somme minimum est atteinte pour les valeurs 4.9170 et 0.02020. Et en partant -1.0001 avec un ratio de 1.87.

Enfin, on regarde avec un nombre de fruit élevé (high) (Voir annexe 12). On observe le minimum pour 14.9674 et 0.0082. On les obtient en partant de -1.0022 avec un ratio de 1.5030.

Au final, on voit que dans les trois cas on va partir proche du minimum que l'on pris donc 1. On voit cependant que plus le nombre de fruit sera élevé, moins le ratio sera grand.

Chapitre 6

Conclusion

Pour conclure, ce stage s'est déroulé en plusieurs étapes avec au début une phase de compréhension du sujet et des méthodes ainsi qu'une première analyse afin de maîtriser le modèle et les variables. J'ai ensuite approfondi mes connaissances sur le modèle afin de l'optimiser et de pouvoir le modifier pour en créer un nouveau. Et enfin, j'ai pu insérer les deux modèles dans Qualitree et travailler sur les résultats qu'ils donnaient pour étudier les modèles et pouvoir regarder l'effet des courbes de stress hydriques.

De ce fait, à la fin de ce stage, j'ai ainsi pu obtenir deux modèles qui vont donner selon deux méthodes les densités de fruit pour étudier la monilia. Ces modèles sont utilisables avec l'algorithme Qualitree. J'ai aussi pu voir pour quelles paramètres les courbes de stress hydrique aller minimiser la propagation de la monilia.

On peut cependant noter une limite à la modélisation que l'on a effectuée car pour la réaliser on a recherché une valeur moyenne entre les deux années et on ne retrouve pas les valeurs exactes ni 2014 ni 2015. On obtient ainsi des valeurs intermédiaires mais non exactes. On peut aussi voir une limite sur la charge de fruit qui en réalité sera souvent supérieure à ce que l'on a estimé dans les expérimentations.

Lors de ce stage, j'ai pu appliquer différentes connaissances acquises sur R et les bases de données. J'ai aussi appris de nombreuses choses comme les fonctions R permettant de réaliser des analyses de sensibilité ou de l'optimisation. J'ai aussi pu apprendre la méthodologie pour réaliser une optimisation dans un milieu réel.

? Ainsi, ce stage m'aura beaucoup apporté, j'ai pu expérimenter le monde du travail dans le domaine de la recherche. Et j'espère avoir apporté à l'INRA et l'unité PSH une aide et un travail efficaces. J'ai pu m'intégrer et œuvrer dans une bonne ambiance grâce au personnel du service. De ce fait, pour moi, ce stage a été une très bonne expérience.

Bibliography

Chapitre 7

Annexes

On va voir ici les différentes annexes.

Annexe 1

1.1

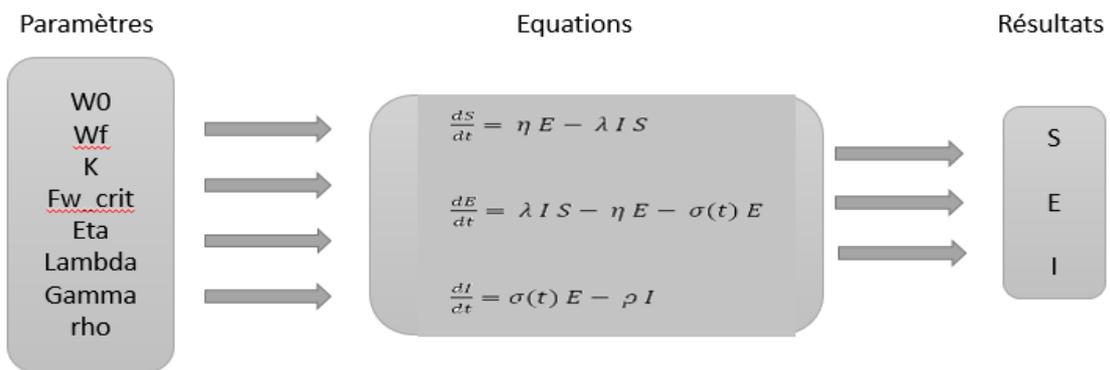


FIGURE 7.1 – Schéma du modèle

1.2

Nom	Description	Unité
i FR I	fraction initial de fruits infectés	0 - 100
i FR E	fraction initial de fruits exposés	0 - 100
i FR S	fraction initial de fruits susceptibles	0 - 100
w0	poids du fruit à floraison	g
wf	poids max du fruit	g
k	taux de conversion de ressource en masse de fruit	day ⁻¹
fw_crit	masse fraiche de fruit qui représente le seuil de fissuration de la cuticule	g
eta - η	durée moyenne de la viabilité des spores à la surface du fruit	day ⁻¹
lambda - λ	taux de transmission par unité infectieuse	fruit ⁻¹ .m ² .day ⁻¹
gamma m - γ_m	constante d'infection pour la masse	g ⁻¹ .day ⁻¹
gamma c - γ_c	constante d'infection pour le crack	cm ⁻²
rho - ρ	période infectieuse moyenne	day ⁻¹

FIGURE 7.2 – Glossaire des paramètres

Annexe 2

2.1

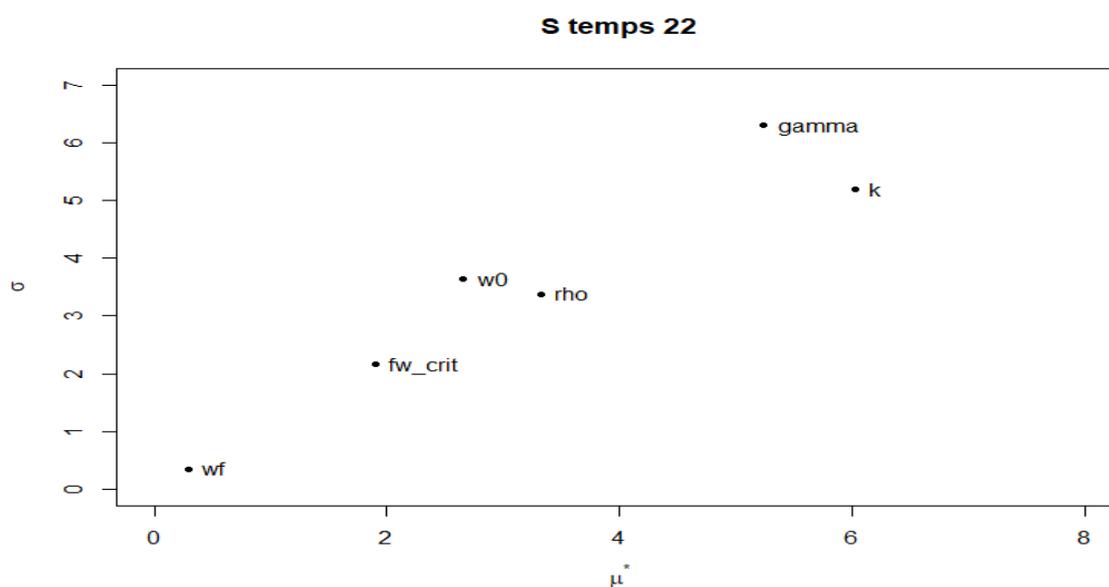


FIGURE 7.3 – l'analyse de sensibilité sur les « susceptible » au temps 22

2.2

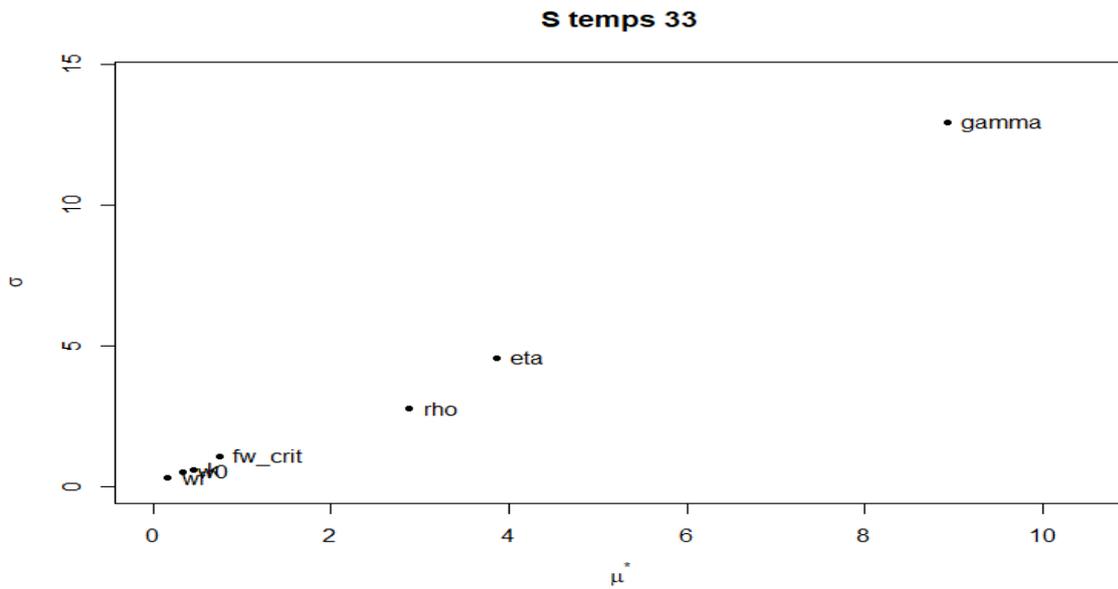


FIGURE 7.4 – l'analyse de sensibilité sur les « susceptible » au temps 33

Annexe 3

3.1

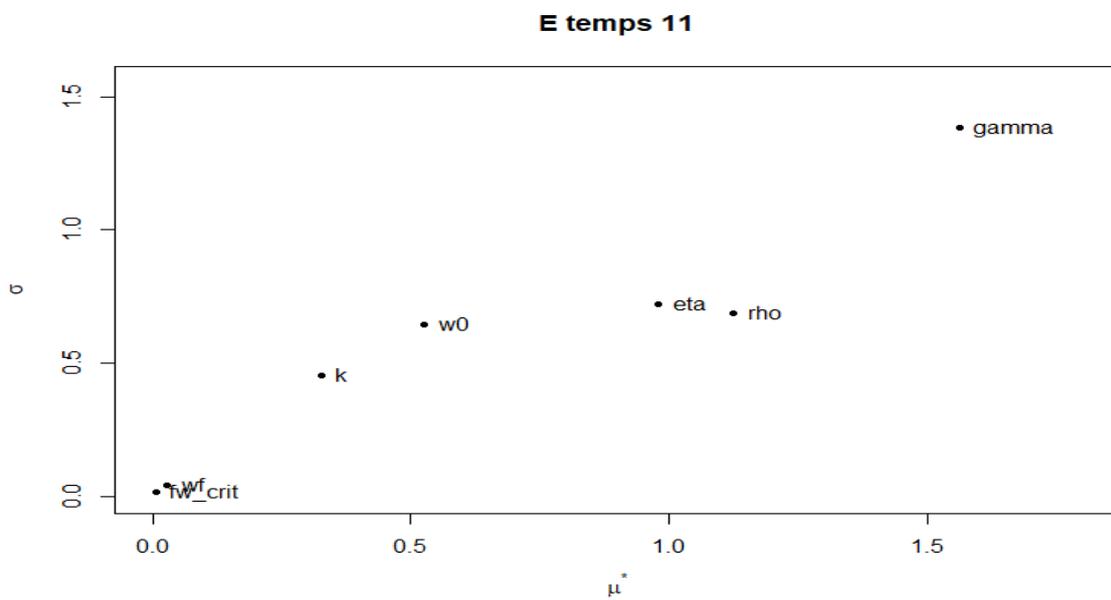


FIGURE 7.5 – l'analyse de sensibilité sur les « exposés » au temps 11

3.2

E temps 22

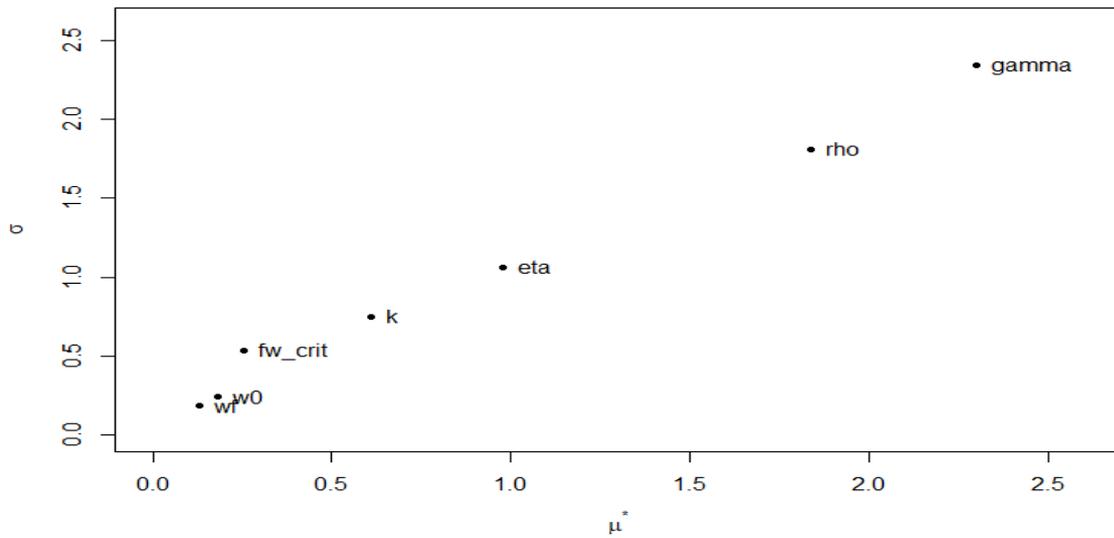


FIGURE 7.6 – Analyse de sensibilité sur les « exposés » au temps 22

3.3

E temps 33

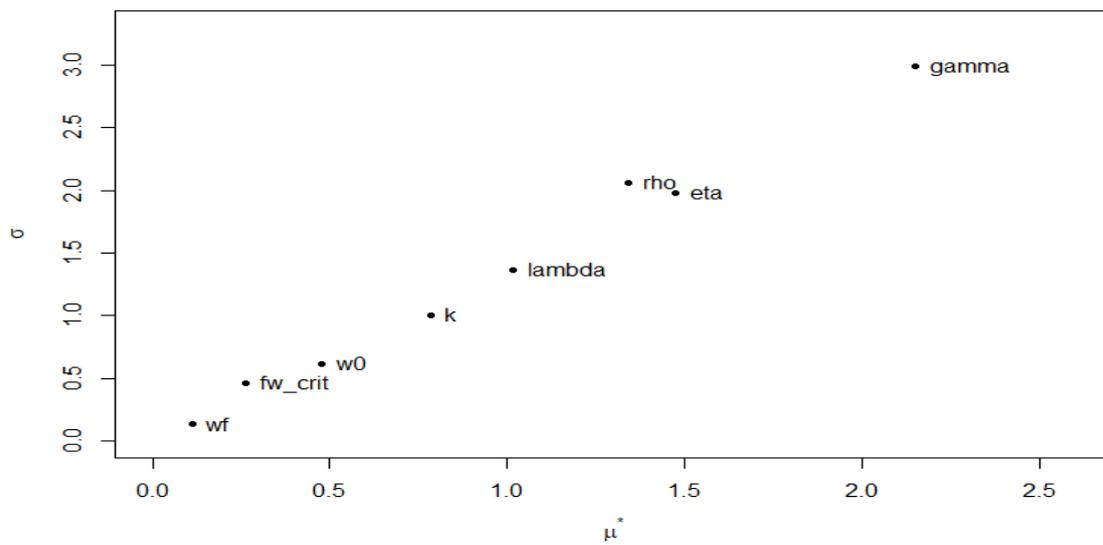


FIGURE 7.7 – Analyse de sensibilité sur les « exposés » au temps 33

Annexe 4

4.1

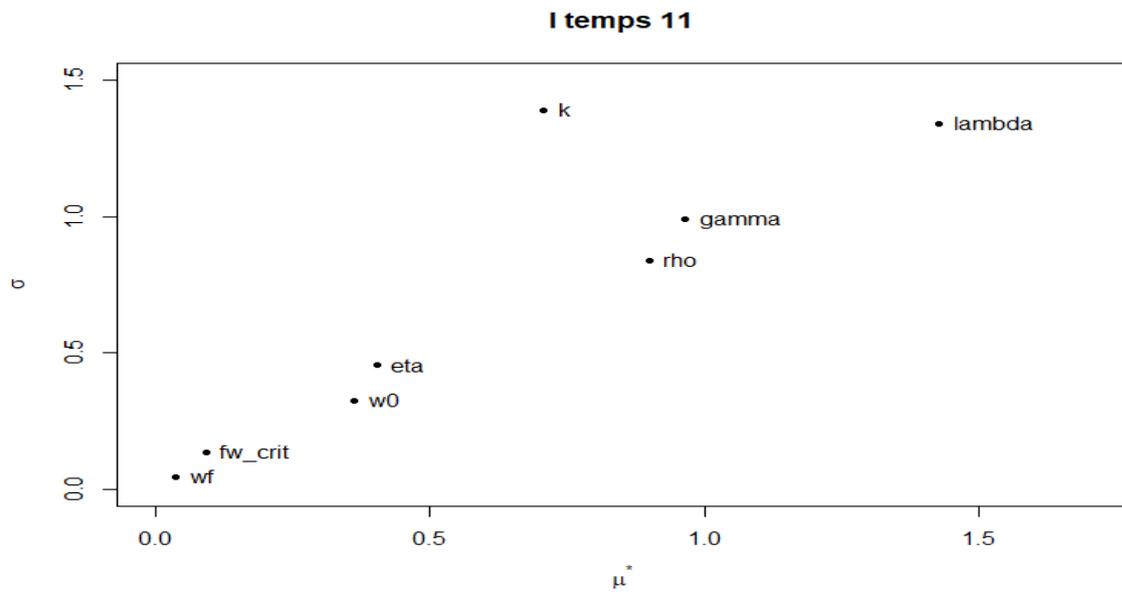


FIGURE 7.8 – l'analyse de sensibilité sur les « infectés » au temps 11

4.2

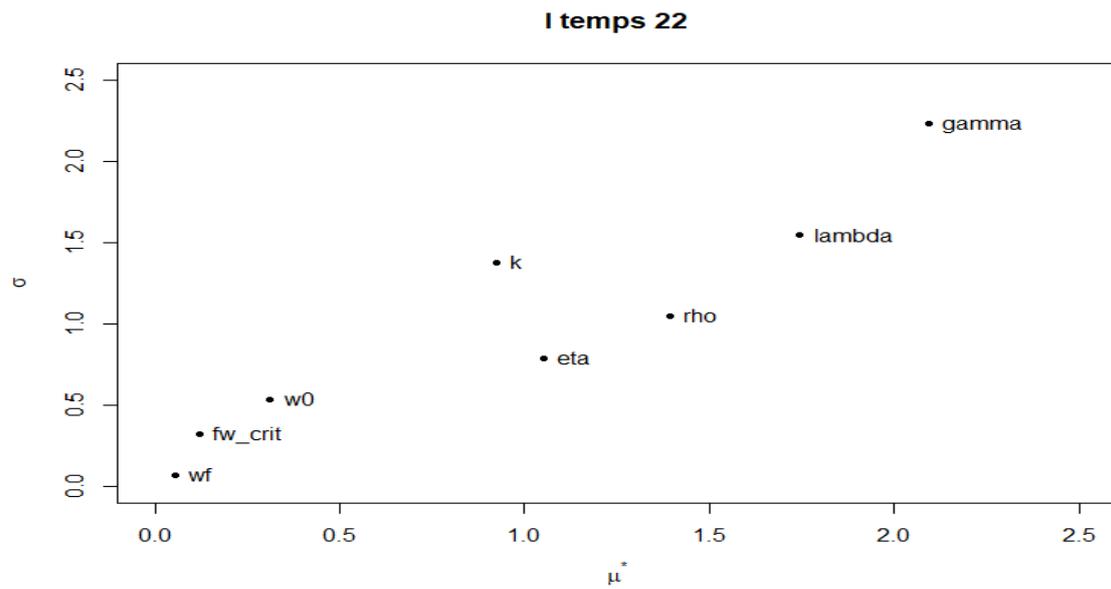


FIGURE 7.9 – Analyse de sensibilité sur les « infectés » au temps 22

4.3

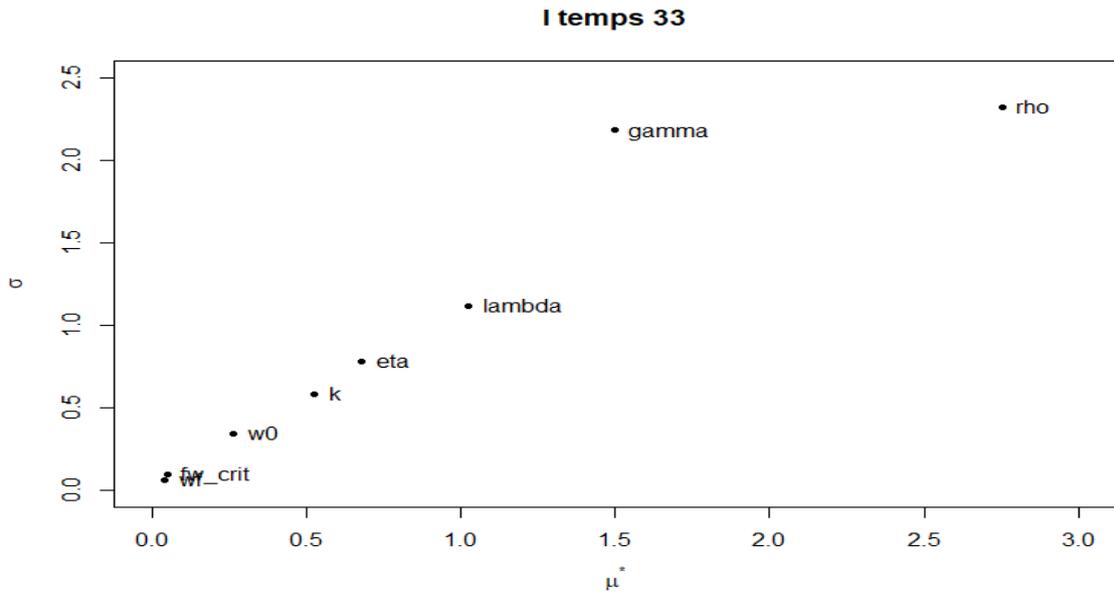


FIGURE 7.10 – Analyse de sensibilité sur les « infectés » au temps 33

Annexe 5

5.1

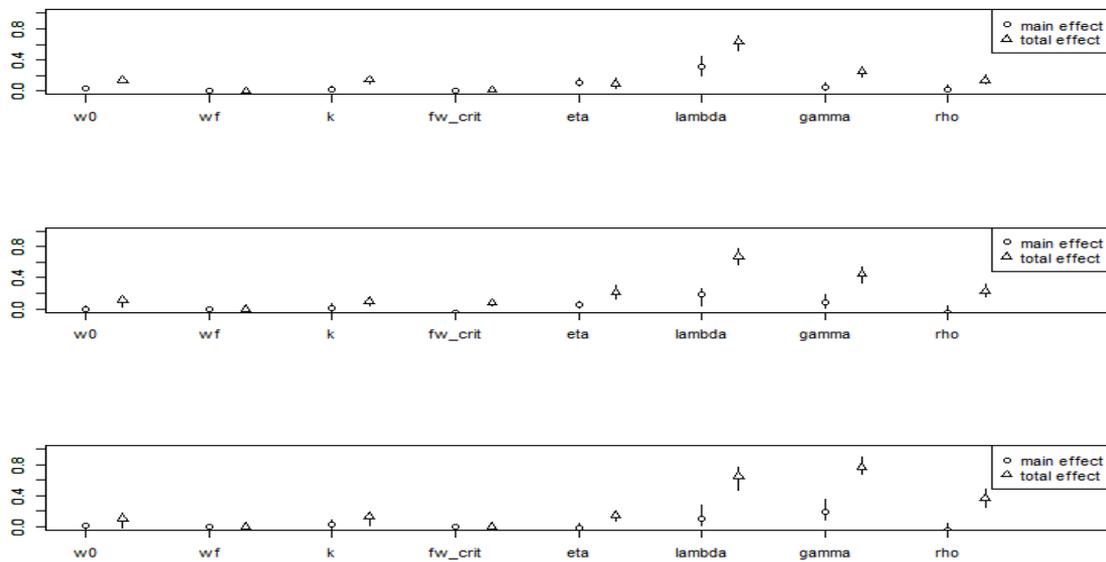


FIGURE 7.11 – l'analyse de sensibilité sur les « exposés »

5.2

Annexe 6

6.1

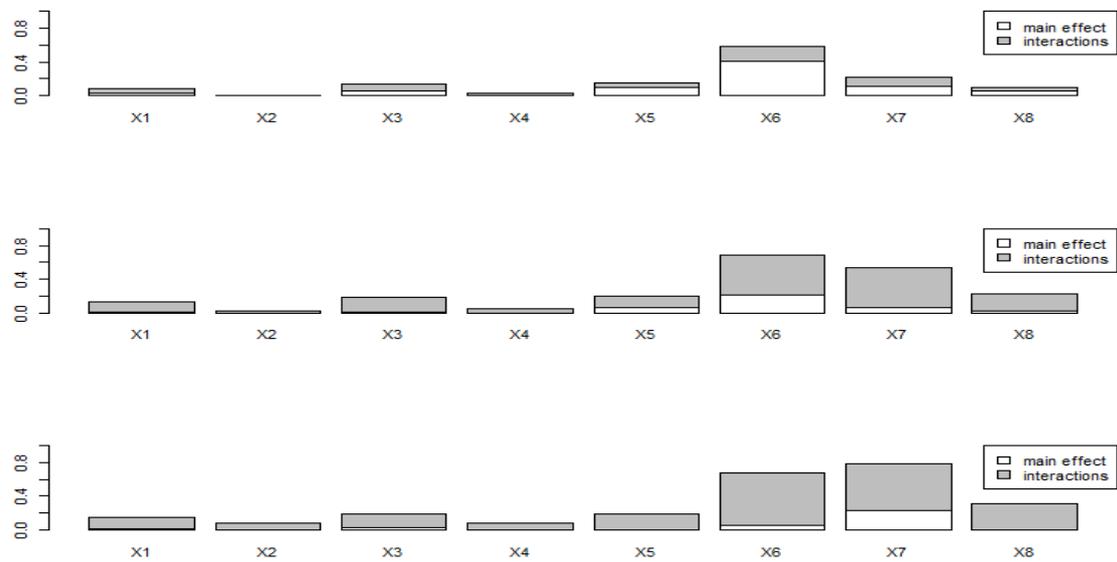


FIGURE 7.12 – l'analyse de sensibilité sur les « exposés »

6.2

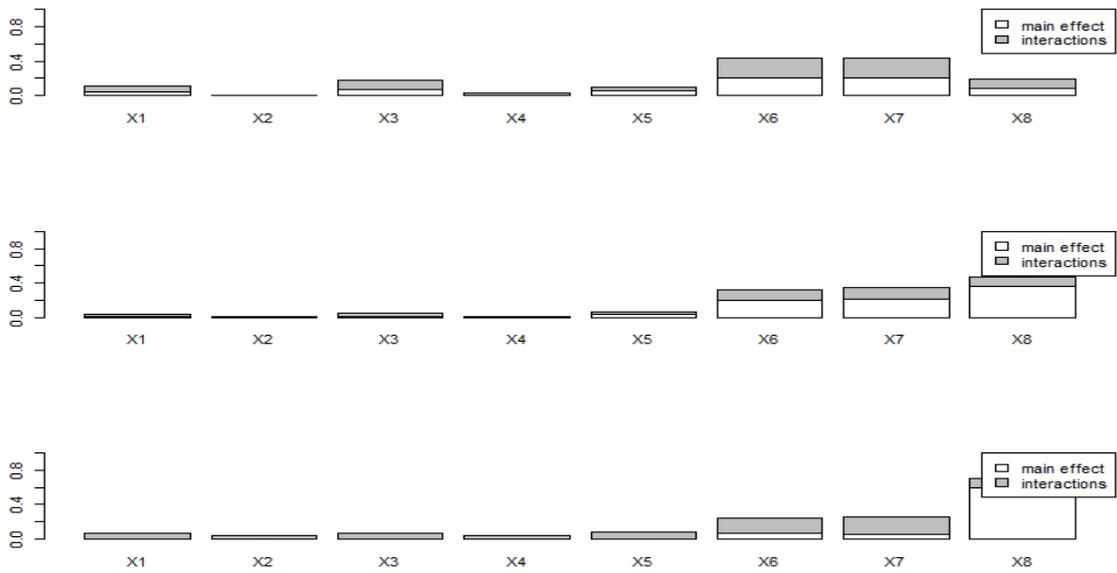


FIGURE 7.13 – l'analyse de sensibilité sur les « infectés »

Annexe 7

7.1

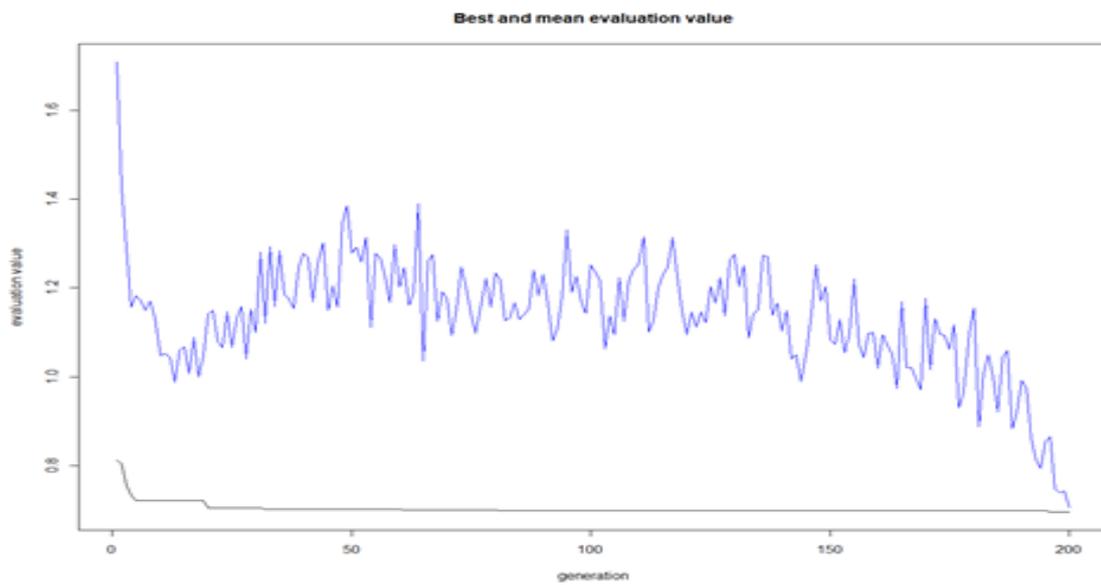


FIGURE 7.14 – RMSE obtenus pour les itérations avec les fruits « infectés »

7.2

7.3

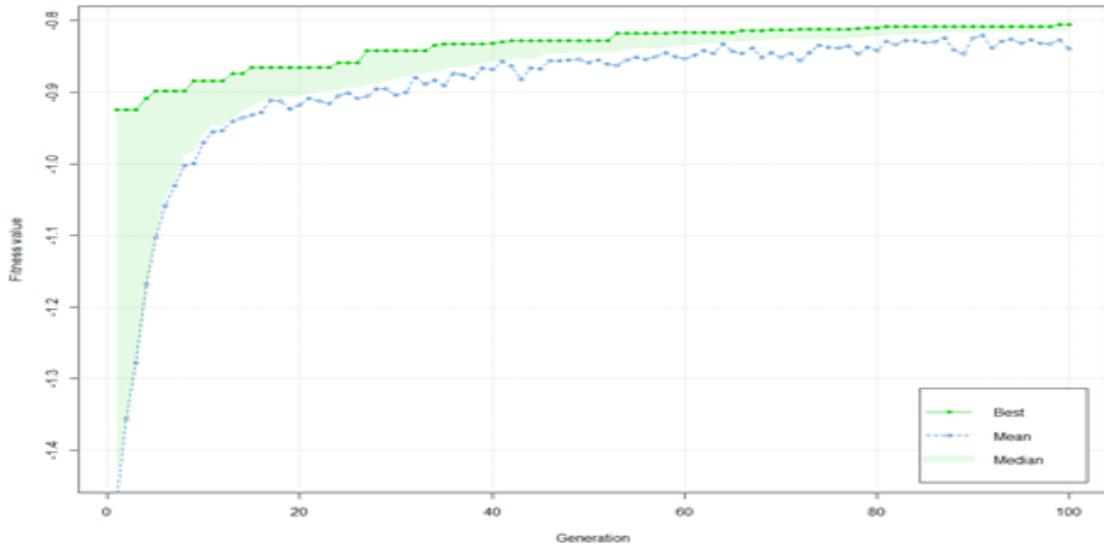


FIGURE 7.15 – RMSE obtenus pour les générations avec les fruits « infectés »

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeur	0.14906	201.1702	0.06628	63.35345	0.25592	0.09178	0.00096	0.08100

FIGURE 7.16 – Paramètres pour RMSE minimum

7.4

7.5

7.6

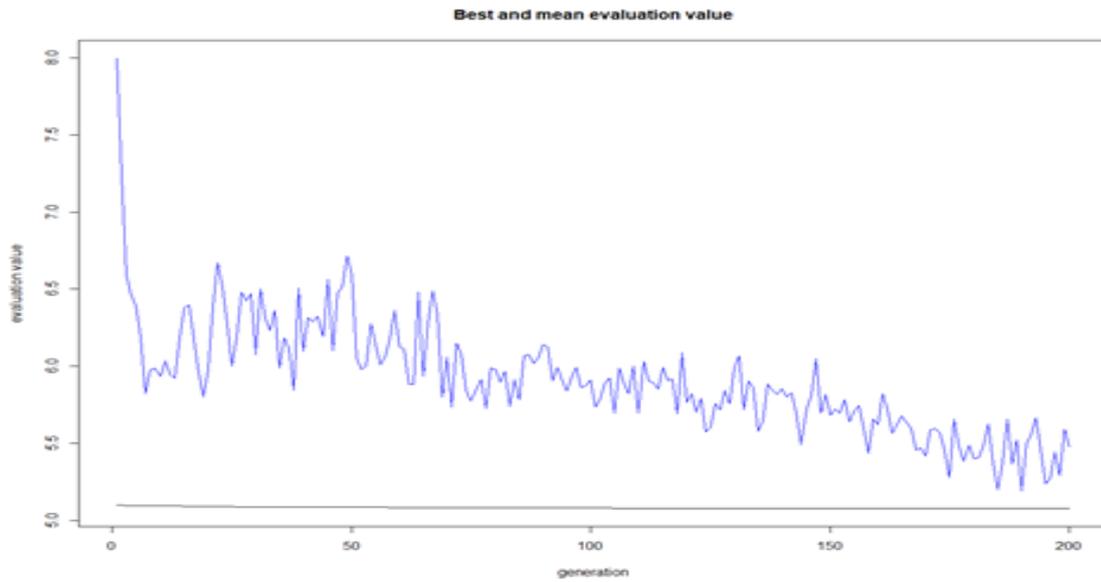


FIGURE 7.17 – RMSE obtenus pour les itérations avec les fruits « sains »

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeur	0.12349	209.76172	0.06456	55.46809	0.00544	0.05049	0.00116	0.12429

FIGURE 7.18 – Paramètres pour RMSE minimum

7.7

Annexe 8

8.1

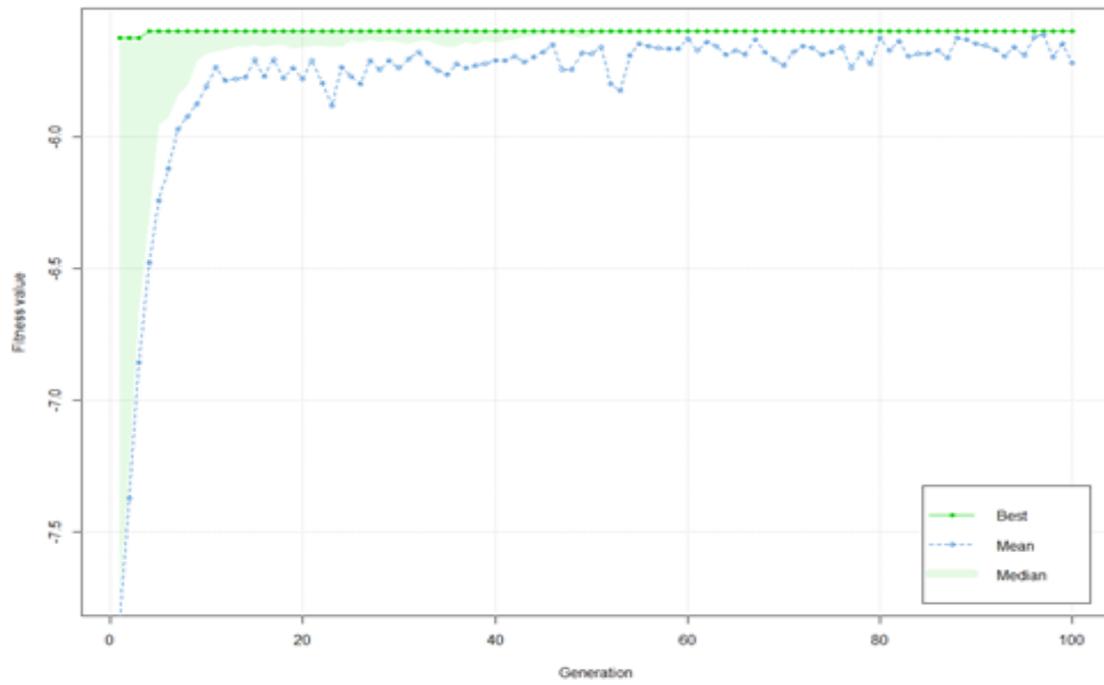


FIGURE 7.19 – RMSE obtenus pour les générations avec les fruits sains

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeur	0.19072	198.3096	0.06705	62.40078	0.34340	0.18260	0.00063	0.19495

FIGURE 7.20 – Paramètres pour RMSE minimum pour ga

8.2

Annexe 9

9.1

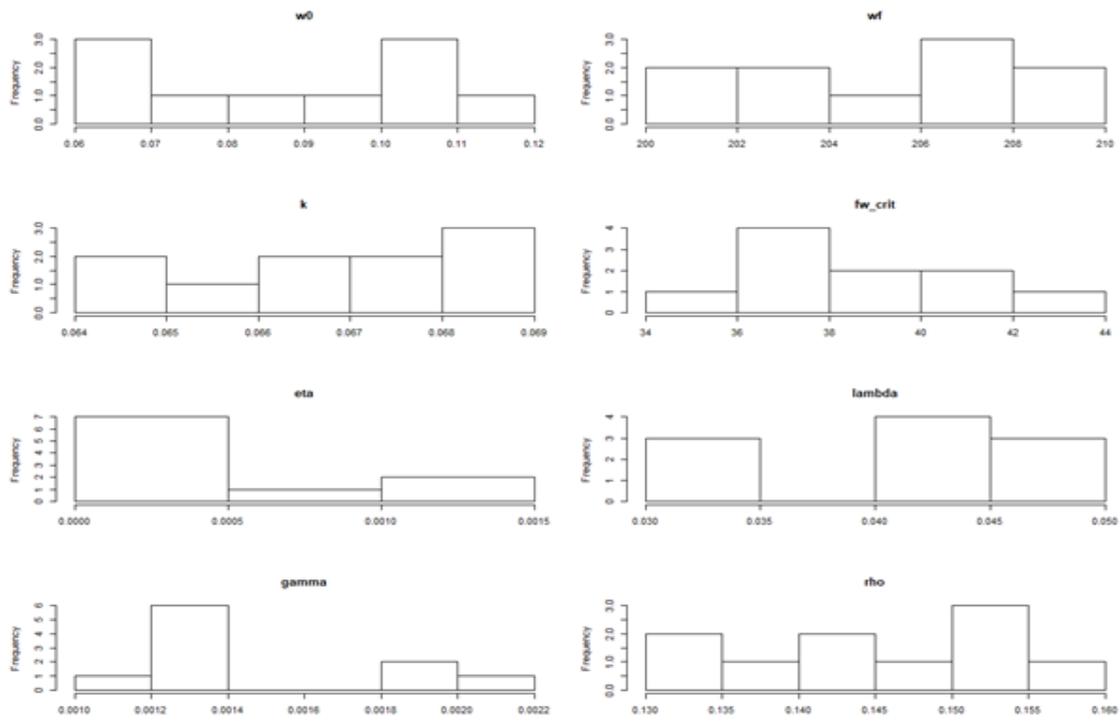


FIGURE 7.21 – Boxplot des paramètres trouvés lors des nouvelles répétitions

	W0	Wf	K	Fw_crit	Eta	Lambda	Gamma	Rho
Valeurs min	0.0800	193	$6.4 \cdot 10^{-2}$	35	0.000004	$3.9 \cdot 10^{-3}$	$0.4 \cdot 10^{-3}$	$4.0 \cdot 10^{-2}$
Valeur	0.10648	204.0464	0.06431	37.6769	0.00083	0.04301	0.00137	0.14694
Valeurs max	0.21	210	$7.1 \cdot 10^{-2}$	70	0.468	$37 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	$36.2 \cdot 10^{-2}$

FIGURE 7.22 – Paramètres qui minimise le RMSE

9.2

Annexe 10

10.1

10.2

10.3

10.4

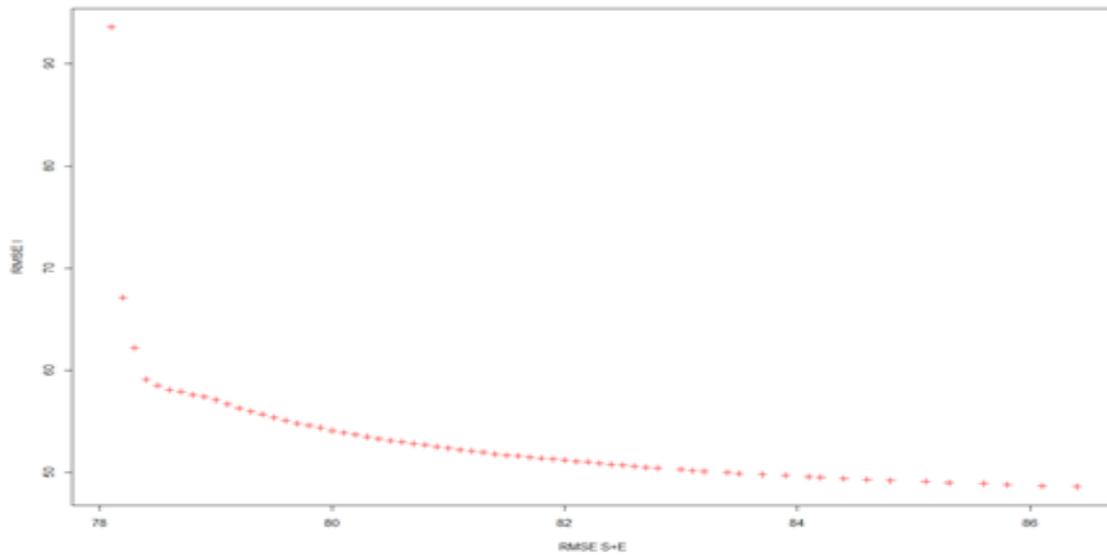


FIGURE 7.23 – NRMSE obtenus lors des simulations sans points dominés

	Eta	Lambda	Gamma	Rho
Min	0.0000400	0.01225	0.01035	0.04005
1er quartile	0.0000758	0.01701	0.30317	0.12771
Médiane	0.0001884	0.01889	0.36773	0.17004
Moyenne	0.0004196	0.02971	0.33617	0.16294
Ecart-type	0.0006485	0.05990	0.08028	0.05659
3eme quartile	0.0004689	0.39988	0.39179	0.20823
Max	0.0060940	0.40000	0.40000	0.24614

FIGURE 7.24 – Valeurs des différents paramètres pour 2014+2015

10.5

10.6

	<i>eta</i>	<i>lambda</i>	<i>gamma</i>	<i>rho</i>
<i>Valeurs</i>	9.929290e-05	0.3998802	0.03856301	0.01825499
	9.958724e-05	0.3998802	0.03856301	0.01825499

FIGURE 7.25 – Paramètres minimisant SE

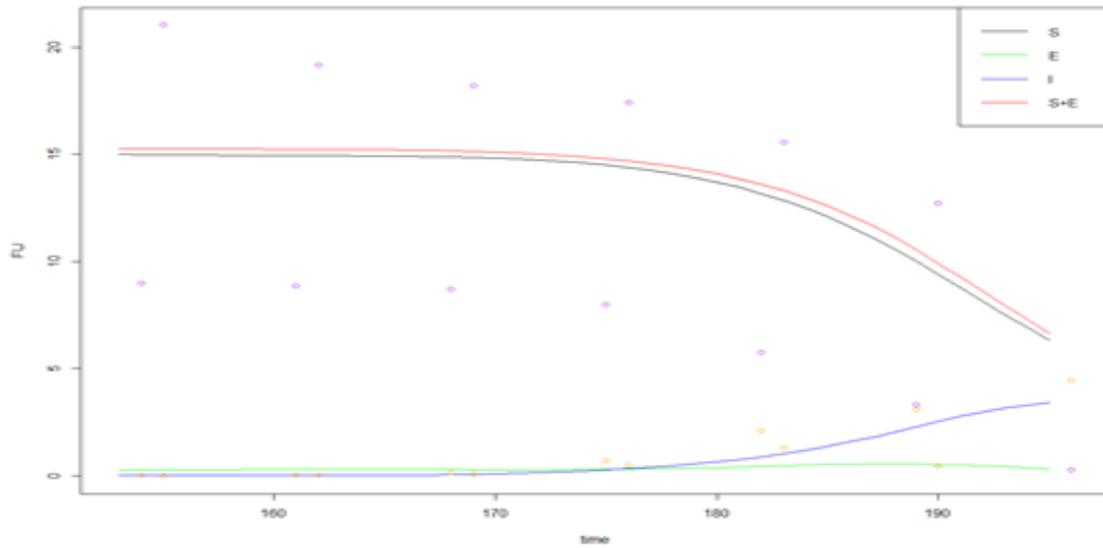


FIGURE 7.26 – Courbe avec les paramètres minimisant SE

Annexe 11

11.1

11.2

	<i>eta</i>	<i>lambda</i>	<i>gamma</i>	<i>rho</i>
Valeurs	4.839796e-05	9.637063e-03	3.988985e-01	6.845104e-03

FIGURE 7.27 – Paramètres minimisant I

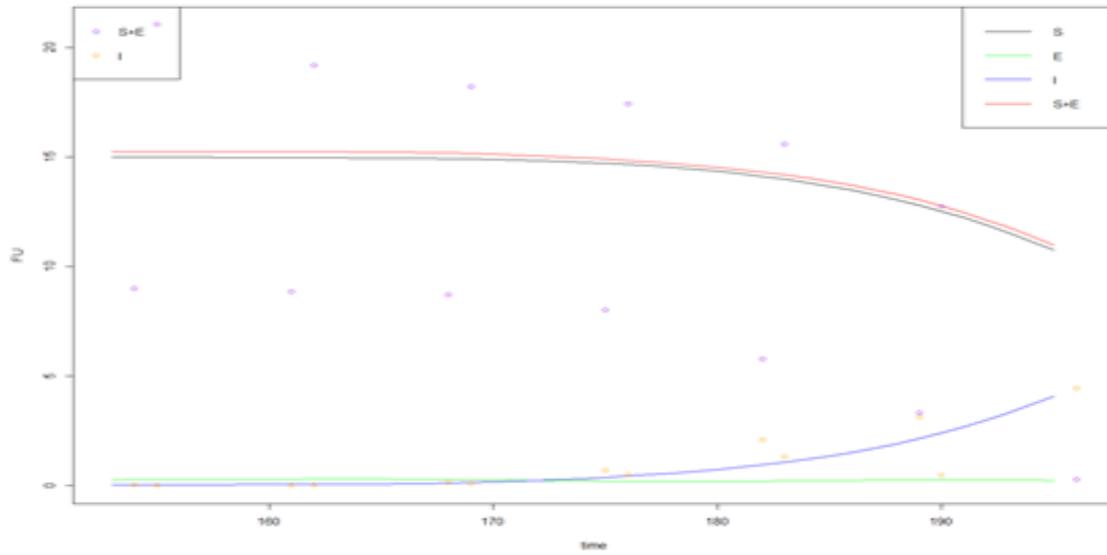


FIGURE 7.28 – Courbes obtenues avec les paramètres pour les I

11.3

11.4

	<i>eta</i>	<i>lambda</i>	<i>gamma</i>	<i>rho</i>
Valeurs	4.066049e-05	2.737425e-02	3.998037e-01	1.687327e-01

FIGURE 7.29 – Paramètres minimisant SE et I

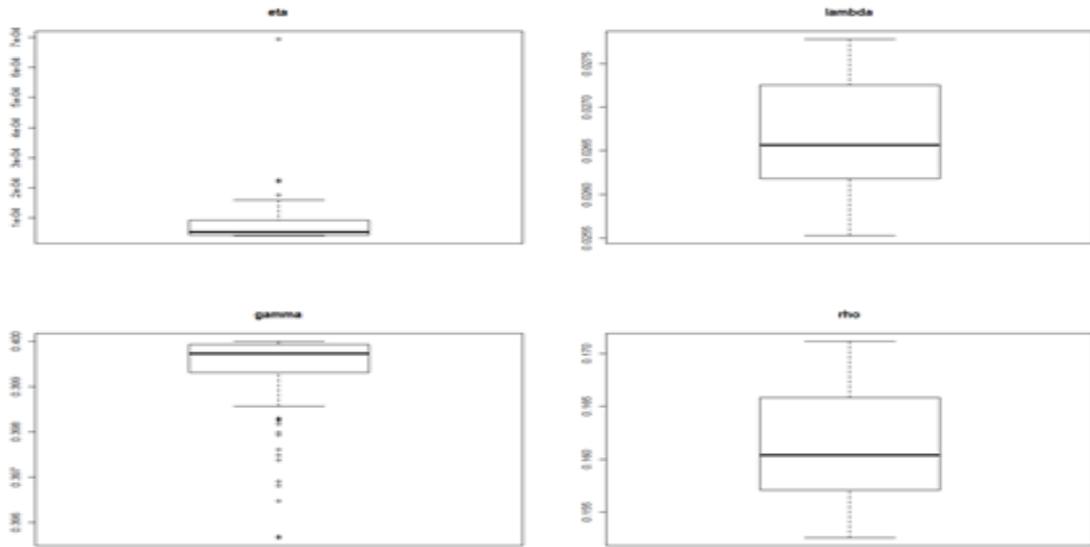


FIGURE 7.30 – Boxplot des paramètres minimisant

11.5

11.6

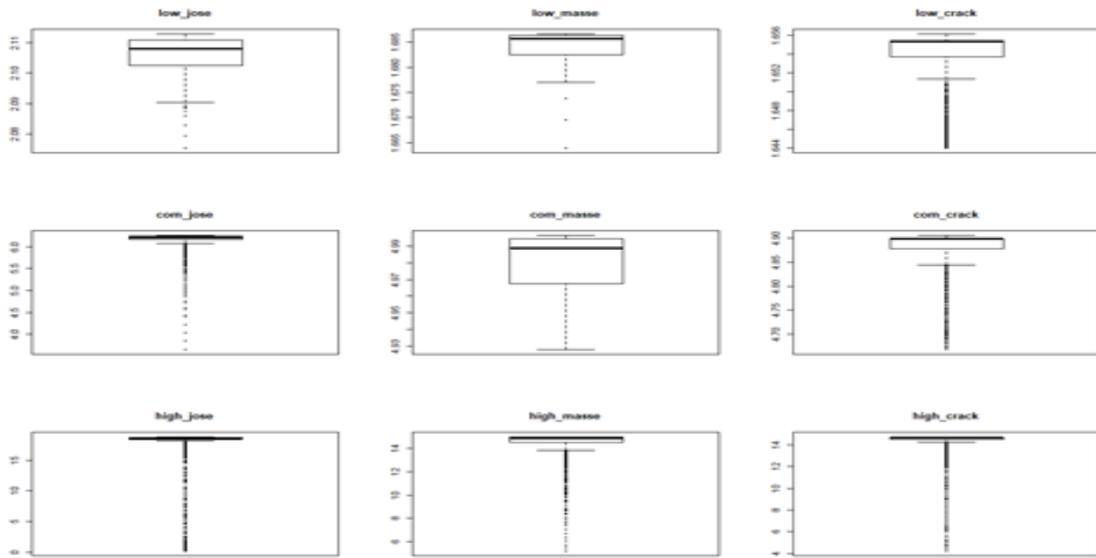


FIGURE 7.31 – Boxplot des fruits susceptibles par nombre de fruit et par simulation un à uns

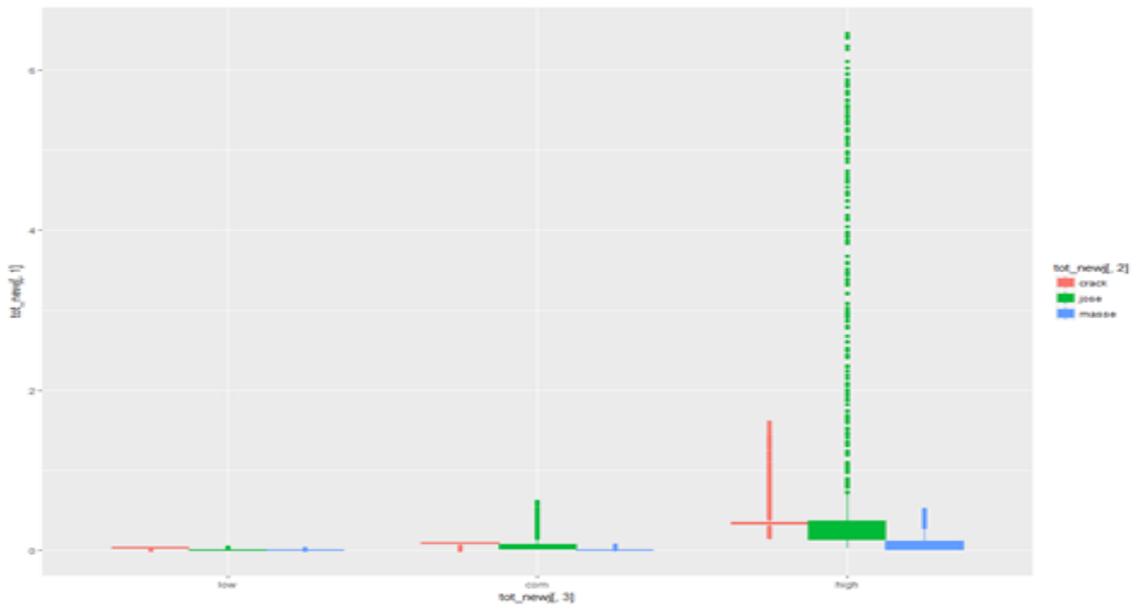


FIGURE 7.32 – Boxplot des fruits exposés par nombre de fruit et par simulation

11.7

Annexe 12

12.1

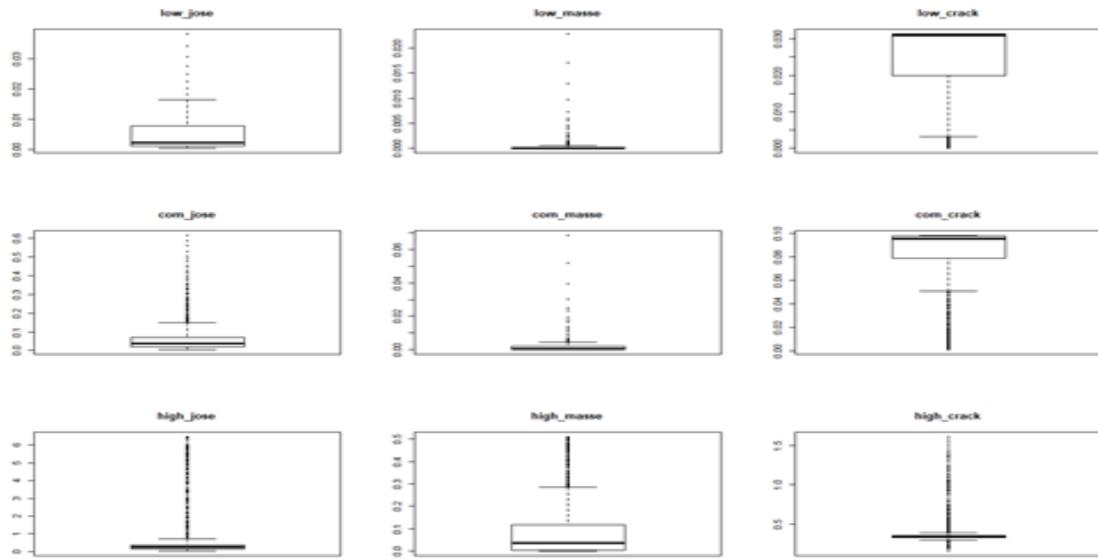


FIGURE 7.33 – Boxplot des fruits exposés par nombre de fruit et par simulation un à uns

12.2

		Low	Com	High
Original	Min	0.0004267	0.004651	0.02669
	1st Qu	0.0011390	0.018351	0.13424
	Median	0.0023221	0.036423	0.27255
	Mean	0.0065144	0.058660	0.76008
	3rd Qu	0.0078456	0.070448	0.36784
	Max	0.0381600	0.614392	6.46086
<hr/>				
Modifié	Min	1.610e-07	1.017e-05	0.0003095
	1st Qu	2.153e-06	1.198e-04	0.0049118
	Median	1.472e-05	4.352e-04	0.0376396
	Mean	1.356e-03	4.592e-03	0.0887485
	3rd Qu	2.064e-04	1.925e-03	0.1172772
	Max	2.281e-02	6.830e-02	0.5085764
<hr/>				
Crack	Min	6.939e-06	0.001043	0.1609
	1st Qu	2.002e-02	0.078849	0.3302
	Median	3.091e-02	0.095054	0.3412
	Mean	2.351e-02	0.077174	0.3955
	3rd Qu	3.126e-02	0.097551	0.3536
	Max	3.130e-02	0.098042	1.6054

FIGURE 7.34 – Indicateurs des fruits exposés

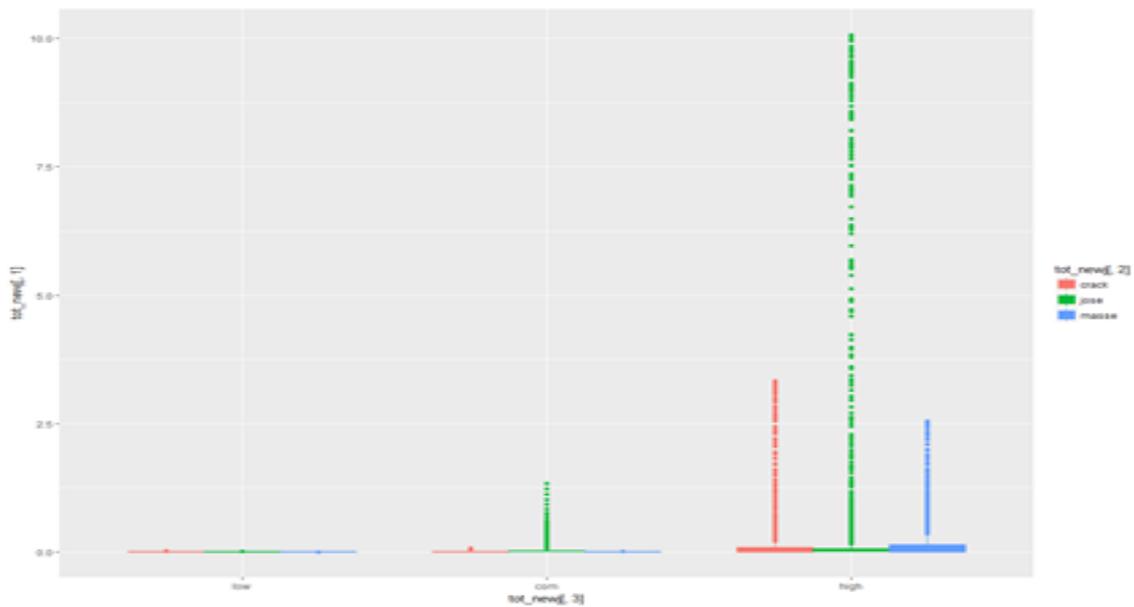


FIGURE 7.35 – Boxplot des fruits infectés par nombre de fruit et par simulation

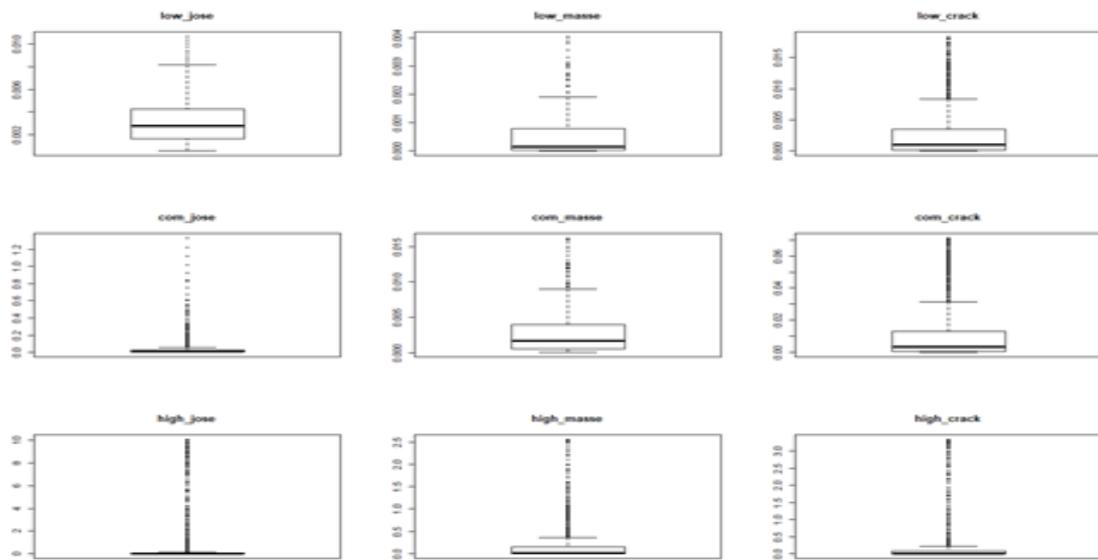


FIGURE 7.36 – Boxplot des fruits infectés par nombre de fruit et par simulation un à uns

		Low	Com	High
Original	Min	0.0005821	0.001322	0.001924
	1st Qu	0.0016529	0.005973	0.011426
	Median	0.0027678	0.011742	0.027292
	Mean	0.0030674	0.048348	0.838857
	3rd Qu	0.0042793	0.026699	0.067070
	Max	0.0106420	1.328565	10.065853
Modifié	Min	2.817e-06	4.773e-05	0.0002205
	1st Qu	3.380e-05	5.543e-04	0.0041189
	Median	1.577e-04	1.715e-03	0.0233465
	Mean	6.015e-04	3.010e-03	0.1890011
	3rd Qu	7.844e-04	3.948e-03	0.1450460
	Max	4.040e-03	1.618e-02	2.5490480
Crack	Min	1.730e-06	9.640e-06	0.000002
	1st Qu	1.445e-04	6.649e-04	0.001978
	Median	9.560e-04	3.385e-03	0.008496
	Mean	2.978e-03	1.482e-02	0.223224
	3rd Qu	3.431e-03	1.300e-02	0.084679
	Max	1.837e-02	7.102e-02	3.335081

FIGURE 7.37 – Indicateurs des fruits infectés

