



## MASTER DE MATHÉMATIQUES DE L'INFORMATION ET DE LA DÉCISION

### **Modélisation du système manguier – cécidomyies des fleurs pour une évaluation de modes de gestion du ravageur et de ses dégâts**

Bastien Reyné

Encadré par : Isabelle Grechi (Cirad, UPR HortSys)  
Frédéric Boudon (Cirad, UMR AGAP)



Année universitaire 2018/2019



## **Résumé**

Résumé en français



# Remerciements

Merci !

# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Table des matières</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Connaissances biologiques</b>	<b>3</b>
2.1 Le mangueier . . . . .	3
2.2 La cécidomyie des fleurs . . . . .	5
<b>3 Données expérimentales disponibles</b>	<b>7</b>
3.1 Dispositif expérimental et suivis . . . . .	7
3.2 Données . . . . .	8
<b>4 Le modèle</b>	<b>13</b>
4.1 Approche retenue et hypothèses . . . . .	13
4.2 Formalisme . . . . .	14
<b>5 Calibration du modèle</b>	<b>19</b>
5.1 Fonction de coût . . . . .	19
5.2 Analyse de sensibilité . . . . .	20
5.3 Algorithme d'optimisation . . . . .	21
5.4 Exploration de l'ensemble des solutions . . . . .	23
<b>6 Mise en œuvre et résultats</b>	<b>25</b>
<b>7 Conclusion</b>	<b>27</b>
<b>Bibliographie</b>	<b>29</b>
<b>A Verger n°2</b>	<b>31</b>

# 1

## Introduction

Le Cirad — où j'ai effectué mon stage — est un organisme de recherche spécialisé dans l'agronomie des régions tropicales et subtropicales, et l'un de ses objectifs principaux est d'encourager le développement durable desdites régions. Cependant la notion de développement durable vient avec quelques contraintes. Notamment, la durabilité implique la limitation des pesticides ; et le développement induit la nécessité d'une production agricole efficiente, capable de nourrir dix milliards de personnes d'ici 2050. L'une des missions du Cirad est de trouver des alternatives aux pesticides afin de gérer les bioagresseurs qui sont responsables de dégâts sur les cultures et engendrent des pertes de production.

Ainsi, il est naturel que le sixième fruit le plus produit au monde, à savoir la mangue<sup>1</sup>, soit l'objet de recherches visant à rendre sa culture plus durable. C'est d'autant plus vrai que la culture du manguier (*Mangifera indica L.*) n'est pas toujours facile. En effet, les manguiers présentent de forts asynchronismes phénologiques, que ce soit à l'intérieur d'une même parcelle entre les différents arbres ou à l'intérieur même d'un arbre entre les différentes branches. Cela entraîne une floraison et une fructification étalée dans le temps, rendant la gestion des vergers plus difficile. Ce phénomène entraîne aussi, pour les fleurs et les fruits, une fenêtre d'exposition prolongée aux bioagresseurs, ce qui favorise leur prolifération. Et ils sont légion. On peut citer la cécidomyie des fleurs et la mouche des fruits chez les ravageurs ou encore l'oïdium et l'antracnose, des maladies fongiques.

Parmi ces ravageurs, la cécidomyie des fleurs (*Procontarinia mangiferae*) peut fortement impacter la production. Cette dernière pond ses œufs dans les inflorescences où elle réalise une partie de son cycle de développement, la fin du cycle se déroulant dans le sol. Le développement des larves dans les inflorescences provoque des dommages potentiellement importants voire la mort des inflorescences. Et qui dit pas d'inflorescences, dit pas de mangues ! Dans le contexte de développement durable, il s'agit donc de trouver des solutions alternatives aux pesticides pour gérer ce ravageur et ses dégâts.

Pour ce faire, deux pistes sont envisagées. La première serait la synchronisation de la floraison, grâce à des pratiques culturales appropriées (la taille, par exemple), ce qui réduirait la fenêtre d'exposition des inflorescences aux ravageurs et limiterait par conséquent leur nombre, en limitant le nombre de cycle de développement du ravageur à l'intérieur du verger. La seconde repose sur le fait qu'une partie du cycle de développement du ravageur

---

1. La sixième production fruitière mondiale est en réalité le groupement des mangues, mangoustans et goyaves [Food and Agriculture Organization, 2017]

se passe dans le sol. Restreindre l'accès au sol (*e.g.* avec un enherbement haut, qui augmente le trajet des larves pour atteindre la terre et favorise la présence de prédateurs) ou l'empêcher (*e.g.* en utilisant une bâche) permet d'altérer ou rompre le cycle de développement du ravageur, et devrait *a priori* permettre de réduire sa présence. Afin de pouvoir vérifier cette hypothèse, une expérimentation sur un verger a été conduite en 2017 ; une parcelle a été divisée en trois sous-parcelles pour tester trois modalités de couverture du sol différentes : un enherbement ras, un paillage synthétique et un enherbement haut. De cette expérimentation, des données ont été acquises. Pour aller plus loin dans la compréhension du fonctionnement du système manguier – cécidomyies des fleurs et de sa gestion par les deux pistes envisagées, une démarche de modélisation de ce système a été engagée. Une première version du modèle a été réalisée lors du stage de [Saint-Criq \[2018\]](#). Mon stage en est la suite et a pour objectifs :

1. modéliser le système manguier – cécidomyies des fleurs, en utilisant les données expérimentales pour le calibrer ;
2. analyser le fonctionnement du système manguier – cécidomyies des fleurs ;
3. tester *in silico* l'effet de la dynamique de floraison et de conditions environnementales, telles que la pression exogène du ravageur, sur l'infestation du verger par les cécidomyies des fleurs.

[BALANCER LE PLAN]

# 2

## Connaissances biologiques

On s'intéresse dans ce chapitre au manguier et à la cécidomyie des fleurs. On s'intéresse en particulier à leurs descriptions d'un point de vue biologique, leurs développements et leurs interactions. Autrement dit, on recense ici les connaissances biologiques nécessaires à la compréhension de notre sujet.

### 2.1 Le manguier

Chaque manguier est un arbre qui a un *cycle phénologique* qui lui est propre. On désigne par cycle phénologique les phénomènes périodiques qui rythment le développement du monde vivant en fonction des variations climatiques saisonnières. Si l'on observe toujours les mêmes étapes, à savoir une croissance végétative suivie d'un repos végétatif puis d'une floraison et enfin d'une fructification (avant de recommencer), il est nécessaire de noter qu'il y a chez le manguier, en sus des conditions climatiques, d'autres facteurs propres à l'arbre qui influent sur la phénologie tels que la position de l'unité de croissance ou la charge en fruit de l'année précédente [Magne, 2004; Normand *et al.*, 2009]. Il en résulte que le cycle phénologique est très variable d'un manguier à l'autre. On observe même des différences au niveau d'un même arbre. Cela contribue à une phénologie étalée dans le temps à l'échelle d'un verger, on parle alors d'asynchronismes phénologiques. Il faut aussi noter que le cycle phénologique peut être modifié (dans une certaine mesure) grâce à des opérations techniques, comme la taille par exemple. Quand cet étalement concerne des stades sensibles aux ravageurs, comme la floraison, cela implique une disponibilité en ressources assez longue pour les ravageurs, et peut expliquer leur forte présence.

La croissance végétative du manguier est dite rythmique. Cela signifie que sa croissance est entrecoupée par des périodes de repos. Durant les périodes de croissance, les branches sont prolongées à leur extrémité par des axes feuillés. Ces axes sont appelées *unités de croissance*. Au fil du temps, les unités de croissance perdent leurs feuilles, durcissent et grossissent jusqu'à faire partie intégrante de la branche. On peut alors voir le manguier comme un empilement d'unités de croissance ; la base du tronc étant l'unité de croissance la plus ancienne et celles qui portent des feuilles les plus récentes [Normand, 2009]. Des unités de croissance sont visibles sur la figure 2.1.

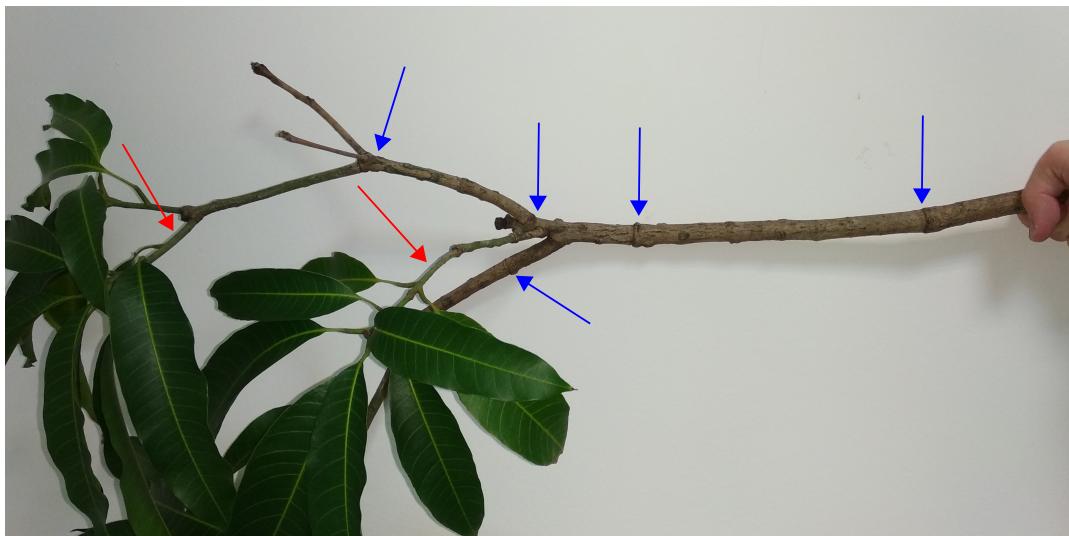


FIGURE 2.1 – Photographie d’une branche portant des unités de croissance. Les flèches rouges montrent des unités de croissance feuillés. Les flèches bleues montrent les délimitations des unités de croissance antérieures ; la portion de la branche entre deux flèches bleues correspond à une ancienne unité de croissance qui a perdu ses feuilles, lignifiée et qui fait désormais partie intégrante de la branche.

Et ce sont ces unités de croissance au bout des branches qui portent les *inflorescences*. Les inflorescences désignent des groupements de fleurs. De façon plus formelle, ce sont des «panicules pyramidales pouvant mesurer jusqu’à 30 cm». Ces fleurs se comptent par centaines voire milliers pour chaque inflorescence. Elles comprennent des fleurs mâles ou hermaphrodites. Ces dernières deviennent des fruits, si elles sont fécondées. En théorie. En pratique, une très faible proportion donnera des fruits. Même sans bioagresseurs et avec de bonnes conditions climatiques, il y a rarement plus de cinq fruits par inflorescence. Par contre si les conditions sont défavorables, beaucoup d’inflorescences peuvent ne pas produire de fruit du tout. Une photographie d’inflorescence est visible sur la figure 2.2.



FIGURE 2.2 – Une inflorescence de mangue (photo : F. Normand)

Le niveau de développement des inflorescences se décompose en différents *stades phénologiques*. On considère ici les stades phénologiques allant de C à F (voir figure 2.3). Le stade C correspond au débourrement (l'éclosion des bourgeons) de l'inflorescence. Le stade F s'étend entre l'apparition de la première fleur jusqu'à la disparition de la dernière. Et les stades D et E représentent les étapes intermédiaires du développement qui mènent du débourrement à la floraison d'une inflorescence.

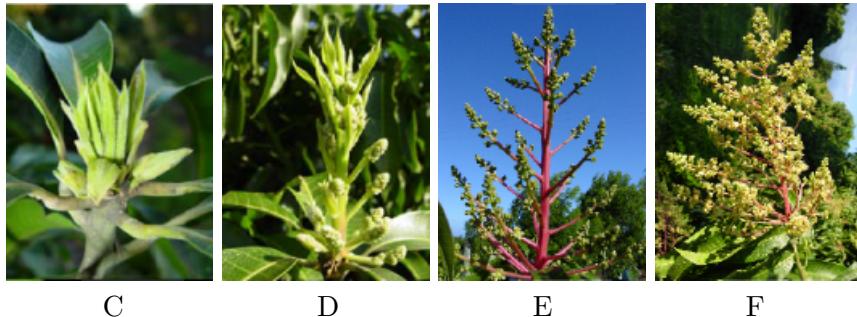


FIGURE 2.3 – Les stades phénologiques C à F d'une inflorescence de manguier (photos : F. Normand)

Les durées des stades phénologiques sont les suivantes :

Stade phénologique	C/D	E	F
Durée (en jours)	7	9	34

Les inflorescences ont ainsi une durée de vie théorique de 50 jours [Saint-Criq, 2018]. Cette durée théorique peut être réduite en cas d'attaque de cécidomyies des fleurs, surtout lorsque l'inflorescence se fait attaquer lors des premiers stades, moment où elle est la plus vulnérable.

À la Réunion, les inflorescences commencent à apparaître en juillet, les premiers fruits apparaissent à la mi-septembre. La récolte a lieu de décembre à janvier. Ces dates varient en fonction de la zone géographique.

## 2.2 La cécidomyie des fleurs

Les cécidomyies des fleurs sont des diptères (des sortes de moucherons) dont le manguier est la seule plante-hôte. À la Réunion, les cécidomyies sont présentes toute l'année et se reproduisent sur les inflorescences et les jeunes feuilles [Amouroux, 2013]. Leur présence est importante pendant la saison des inflorescences en hiver et plutôt faible en été où elles survivent en se reproduisant sur des jeunes feuilles.

Le cycle de développement de la cécidomyie des fleurs est schématisé sur la figure 2.4. Les femelles présentes dans le verger au jour  $t$  peuvent pondre jusqu'à 150 œufs dans les inflorescences. Les œufs laissent place à des larves. Au bout de sept à douze jours après la ponte, une fois le troisième stade de développement larvaire atteint, les larves s'éjectent des inflorescences en direction du sol. Et c'est ce développement larvaire qui provoque des dégâts sur les inflorescences car les larves consomment le tissu végétal et creusent des galeries, fragilisant ainsi les inflorescences. Une fois au sol, les larves s'enfouissent dans le sol. Elles peuvent alors entrer en phase de *pupaison*, qui correspond à la transformation de la larve en pupe. Cette phase dure quatre à six jours, et une fois finie les cécidomyies adultes émergent du sol et infestent le verger — perpétuant ainsi le cycle. Une fois adultes, leur durée de vie n'excède pas 72 heures.

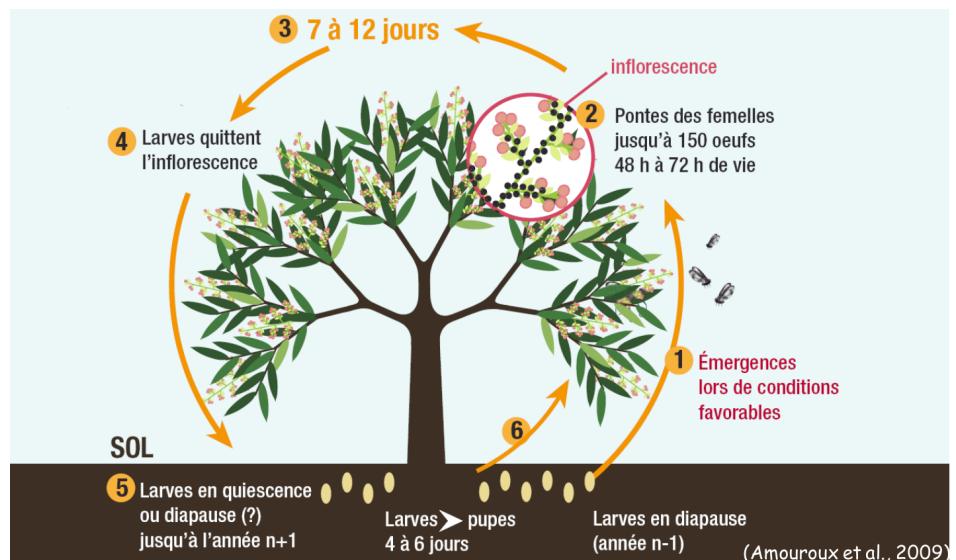


FIGURE 2.4 – Représentation du cycle de développement de la cécidomyie des fleurs du manguier par [Amouroux et al. \[2009\]](#).

Alternativement à la pupaison, les larves peuvent aussi rentrer en *diapause*. La diapause peut être vue comme une sorte d'*hibernation*, où le cycle est mis en pause un an ou deux avant l'émergence de la cécidomyie adulte.

Les individus qui sortent de diapause permettent notamment le lancement de la dynamique d'*infestation* au début des périodes de floraison des cycles suivants. La sortie de diapause est dûe à une baisse de température qui coïncide avec le début de la floraison [[Amouroux et al., 2014](#)].

# 3

## Données expérimentales disponibles

Ce chapitre explique l'expérimentation qui a été menée en 2017 sur deux vergers dans la commune de Saint-Paul (à la Réunion). On décrira le dispositif de l'expérimentation et les données qui en ont été acquises. Le but de cette expérimentation était de déterminer quel est l'impact de la modalité de couverture du sol dans le degré d'infestation d'un verger.

### 3.1 Dispositif expérimental et suivis

Le verger expérimental (que l'on appellera aussi parcelle par la suite) n°1 a une superficie de 0.33 hectares et est séparé en trois sous-parcelles. Sur chacune des trois sous-parcelles, une modalité de couverture du sol différente est mise en place. Sur un côté il y avait un enherbement entretenu de sorte qu'il reste ras, la modalité *enherbement ras (ER)* fera référence à ce traitement dans la suite du document. La sous-parcelle du milieu fut bâché, afin que les cécidomyies ne puissent ni entrer dans le sol ni en sortir ; cette modalité correspond au *paillage synthétique (PS)*. La dernière partie fut laissée telle quelle, sans entretien particulier, donnant ainsi un *enherbement haut (EH)*. À noter qu'à côté de ce verger, il y en avait un autre qui a probablement pu servir de source d'infestation exogène au verger expérimental. Tout cela est schématisé sur la figure 3.1. Le dispositif sur la parcelle n°2, comprenant également ces trois modalités de couverture de sol, est décrit dans l'annexe A.

Sur ces deux vergers expérimentaux, deux types d'observations furent effectués entre juin et octobre 2017. Le premier porte sur les inflorescences. Huit unités de croissance furent sélectionnées sur chacun des vingt-cinq arbres échantillonés aléatoirement dans chacune des trois sous-parcelles. Et c'est ainsi qu'entre le 26 juin et le 3 octobre 2017 furent notés les dates de débourrement des inflorescences présentes sur les deux cents unités de croissance suivies. À partir du 6 septembre furent aussi notées les dates de mort des inflorescences. Les relevés ont été effectués deux fois par semaine. Les données de ce relevé seront rassemblées en un jeu de données que l'on nommera *dataset 1*.

La seconde catégorie d'observations porte sur la capture des larves de cécidomyies des fleurs. Dans chacune des trois sous-parcelles, dix arbres furent sélectionnés. Sous chacun

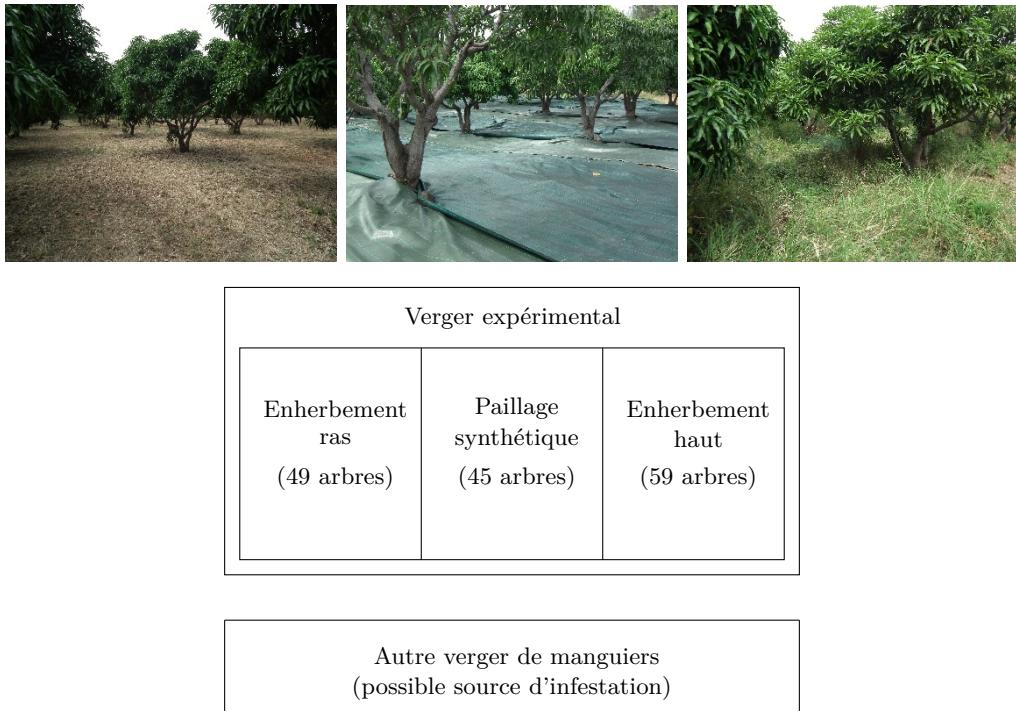


FIGURE 3.1 – Description du dispositif expérimental de la parcelle n°1. En haut, photographie des trois modalités de couverture du sol : enherbement ras, paillage synthétique et enherbement haut. En bas, schéma du dispositif expérimental et de la configuration des sous-parcelles. Le verger sur lequel ont été testées les trois modalités de couvertures du sol était situé à côté d'un autre verger.

de ces arbres furent placés deux pièges en dessous des inflorescences présentes. Les pièges sont des bidons plastiques carrés de douze centimètres de côté remplis d'eau. À noter que les pièges furent déplacés au cours du temps pour qu'ils soient toujours en-dessous d'inflorescences. Et c'est ainsi qu'entre le 18 juillet et le 6 octobre 2017 furent notés le nombre de larves piégées, le nombre d'inflorescences vivantes au-dessus du piège et le nombre d'inflorescences vivantes dans l'arbre. Les relevés ont été effectués deux fois par semaine. Les données de ce relevé seront rassemblées en un jeu de données que l'on nommera *dataset 2*.

## 3.2 Données

Après mise en forme des données<sup>1</sup>, on peut extraire les dynamiques qui nous intéressent. Il faut cependant noter que les deux jeux de données n'ont pas la même échelle (200 unités de croissance contre 10 arbres) et ont été acquis sur des sous-échantillons d'arbres différents. On choisira de tout mettre à l'échelle de la sous-parcelle.

### Inflorescences vivantes

On peut extraire les dynamiques d'inflorescences vivantes grâce aux deux jeux de données : elles seront notées  $I_t^1$  pour le *dataset 1* et  $I_t^2$  pour le *dataset 2*. Pour le *dataset 2*,

1. Les scripts et les données utilisés sont disponibles à l'adresse <https://github.com/bastienreyne/cecidomyie>

on possède le nombre d’inflorescences vivantes dans les arbres suivis aux différentes dates ; il suffit alors de mettre à l’échelle de la sous-parcelle comme suit :

$$I_t^2 = \frac{N}{n} \sum_{k=1}^n I_{k,t}^2,$$

avec  $N$  représentant le nombre d’arbre dans la sous-parcelle,  $n$  le nombre d’arbre suivis et  $I_{k,t}^2$  le nombre d’inflorescences sur l’arbre  $k$  à la date  $t$ . À noter que l’on choisira un pas de temps journalier, le nombre d’inflorescences vivantes entre deux relevés effectif est alors le résultat d’une interpolation linéaire entre lesdits relevés.

Pour le *dataset 1*, on possède le nombre de débourrements journalier  $B_t$  et le nombre de morts journalier  $D_t$ . Le nombre d’inflorescences vivantes au jour  $t$  s’écrit alors

$$I_t^1 = \alpha \left( \sum_{j=1}^t B_j - \sum_{j=1}^t D_j \right),$$

où  $\alpha$  représente le coefficient de mise à l’échelle pour passer de deux cents unités de croissance à la sous-parcelle. Il faut cependant apporter une correction à cette dynamique. En effet, l’observation des inflorescences mortes n’a été faite qu’à partir du 6 septembre. De ce fait, sur ce jeu de données la distinction entre inflorescences vivantes et mortes n’est possible qu’à partir du 6 septembre. Il en résulte une forte diminution du nombre d’inflorescences entre le 5 et le 6 septembre (voir figure 3.2). Cet écart correspond au nombre de mort cumulé jusqu’au 6 septembre, et qu’il faut donc répartir sur la période concernée. N’ayant aucune indication de comment la répartir, on utilisera la dynamique d’inflorescences vivantes du *dataset 2* afin que la dynamique du *dataset 1* y ressemble le plus possible — et on en profitera au passage pour estimer le coefficient de mise à l’échelle  $\alpha$ . Plus précisément, on attribuera un poids (à calibrer numériquement) pour tous les jours entre le jour 1 et le 5 septembre, et le nombre de morts chaque jour sera donné par la formule

$$D_t^c = \frac{p_t \times m}{\sum_j p_j},$$

où  $D_t^c$  désigne le nombre de mort à la date  $t$ ,  $m$  le nombre de mort observés au 6 septembre et  $p_t$  le poids assigné au jour  $t$ . Les poids  $p_t$  et le coefficient de mise à l’échelle  $\alpha$  seront déterminés numériquement afin de résoudre le problème

$$\arg \min_{\alpha, p_t} \sum_t |I_t^2 - I_t^{1,c}|,$$

où  $I_t^{1,c}$  représente les inflorescences vivantes du *dataset 1* mises à l’échelle et corrigées ; cette dynamique est déterminée par la formule

$$I_t^{1,c} = \begin{cases} I_t^1 - D_t^c & \text{si } t \leq 6 \text{ septembre,} \\ I_t^1 & \text{sinon.} \end{cases}$$

Les différentes dynamiques d’inflorescences sont visibles sur la figure 3.2. On remarque des dynamiques très différentes pour la modalité «enherbement haut» selon le *dataset* considéré, et ce même après correction de la dynamique issue du *dataset 1*. On peut expliquer ce phénomène par la grande variabilité de la phénologie chez le mangouier ; ainsi des échantillonages différents peuvent produire des dynamiques très différentes. Les deux autres modalités ont en revanche des dynamiques similaires (après correction).

On fera le choix de privilégier, à chaque fois que cela s’avérera possible, les dynamiques issues du *dataset 2*. Ce choix découle du fait que ce sont les dynamiques associées aux dynamiques de larves.

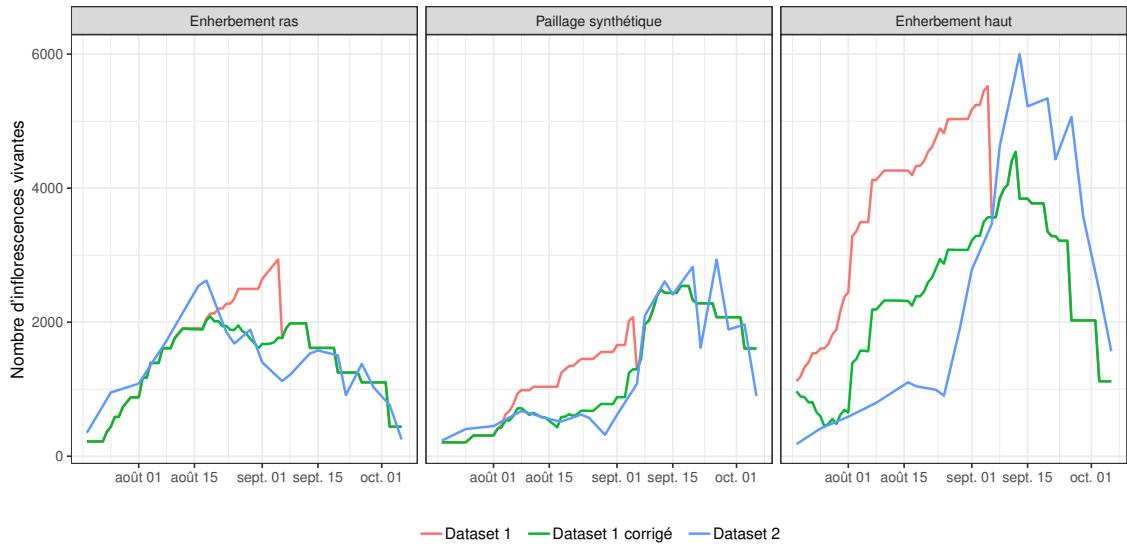


FIGURE 3.2 – Comparaison des différentes dynamiques d’inflorescences vivantes du verger n°1 en fonction du dataset utilisé.

## Larves

Le *dataset 2* permet aussi d’extraire la dynamique de larves, à partir des larves piégées. En effet, on connaît le nombre de larves par piège, le nombre d’inflorescences vivantes situées au-dessus des pièges et le nombre d’inflorescences vivantes dans les arbres suivis. De là, on peut estimer le nombre de larves qui s’éjecte des inflorescences à l’échelle d’un arbre, puis à l’échelle de la sous-parcelle. Il y a cependant ici une subtilité : le relevé des pièges n’est pas quotidien, il faut donc répartir le nombre de larves piégées entre deux relevés effectifs sur la période entre lesdits relevés. La dynamiques de larves peut donc s’obtenir en utilisant la formule

$$L_t = \frac{N}{n} \left( \sum_{k=1}^n L_t^k \times \frac{I_{k,t}^2}{I_{k,p}^2} \right),$$

où  $N$  représente le nombre d’arbre dans la sous-parcelle,  $n$  le nombre d’arbre suivis,  $I_{k,t}^2$  le nombre d’inflorescences sur l’arbre  $k$  à la date  $t$ ,  $I_{k,p}^2$  le nombre d’inflorescences au-dessus des pièges dans l’arbre  $k$  à la date  $t$  et on définit  $L_t^k$ , le nombre de larves dans les pièges de l’arbre  $k$  à la date  $t$ , par

$$L_t^k = \frac{L_{k,t}^j}{t^j - t^{j-1}},$$

avec  $t^j$  le nombre de jours entre la première observation et le  $j^{\text{ème}}$  relevé et  $L_{k,t}^j$  le nombre de larves dans les pièges de l’arbre  $k$  au  $j^{\text{ème}}$  relevé. Les différentes dynamiques de larves sont visibles sur la figure 3.3.

Les dynamiques pour le verger n°2 sont visibles dans l’annexe A.

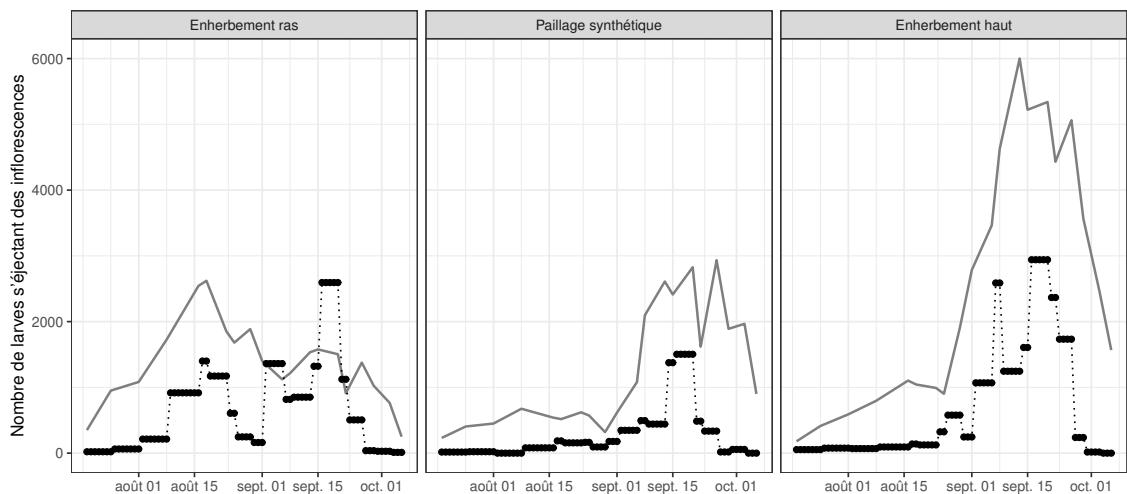


FIGURE 3.3 – Dynamiques de larves s'éjectant des inflorescences des manguiers chaque jour dans le verger n°1 pour chacune des trois sous-parcelles. En gris sont visibles les dynamiques d'inflorescences vivantes (issues du dataset 2,  $I_t^2$ ).



# 4

## Le modèle

ON s'intéresse ici aux différentes approches possibles pour modéliser le système manguier – cécidomyies des fleurs. Il y a dans la littérature quelques modèles relatifs aux manguiers et aux cécidomyies.

Concernant la modélisation en rapport avec le manguier, on peut citer le modèle *Virtual Mango* développé par [Boudon et al. \[2016\]](#). C'est un modèle qui simule le développement architectural du manguier, ainsi que la croissance et les différents stades phénologiques des unités de croissance et des inflorescences et la croissance des fruits. Le modèle ne prend cependant pas en compte l'impact des ravageurs sur le développement de l'arbre et sa production.

Concernant la modélisation en rapport avec les cécidomyies des fleurs, on peut noter l'existence d'un modèle de colonisation d'un verger par les cécidomyies [\[Amouroux, 2013\]](#). Les données utilisées proviennent d'une expérimentation menée sur un verger entièrement bâché. C'est un modèle stochastique et spatialisé qui prend en compte l'arrivée des femelles sur le verger, leurs pontes d'œufs dans les inflorescences, le développement de ces derniers jusqu'au troisième stade de développement larvaire et enfin l'éjection des larves des inflorescences. Cependant, le fait que le verger modélisé est entièrement bâché implique l'absence de cycle de développement complet pour les cécidomyies à l'intérieur du verger ainsi que l'absence d'émergence de cécidomyies issues de la diapause.

On choisira ici l'approche amorcée par [Saint-Criq \[2018\]](#), qui propose une modélisation non spatialisée du verger expérimental considérant l'infestation du verger par des femelles exogènes et la reproduction endogène, tout en prenant en compte les différentes modalités de couverture du sol. L'objectif sera de simuler les dynamiques de larves à l'échelle de la sous-parcelle en fonction du nombre d'inflorescences présentes.

### 4.1 Approche retenue et hypothèses

L'approche retenue se veut simple. On considérera les populations (de cécidomyies et d'inflorescences vivantes) dans chaque sous-parcelle et non les individus. Concrètement, cela veut dire que l'on ne considérera que les nombres totaux d'inflorescences vivantes et de cécidomyies qu'il y a chaque jour dans chaque sous-parcelle. Par ailleurs, la période considérée est la saison de floraison et le modèle n'est pas pluriannuel.

Les hypothèses sur les processus biologiques qui régissent le système, que nous faisons, sont :

- la modalité de couverture du sol n'influe pas sur le développement et la phénologie du manguier, elle n'influe pas non plus sur le déplacement des cécidomyies ;
- la durée de vie des femelles n'est que d'un jour, et il correspond au jour de la ponte ;
- aucune cécidomyie (larve ou adulte) ne rentre dans le sol ou ne sort du sol pour la sous-parcelle avec un paillage synthétique ;
- il y a un taux de mortalité pour larves lorsqu'elles rentrent ou et les adultes lorsqu'ils sortent du sol, et ce taux est différent entre l'enherbement ras et l'enherbement haut ;
- les larves qui n'entrent pas en pupaison (soit qui meurent, soit qui rentrent en diapause) sont exclues du système ;
- la probabilité d'entrer en pupaison dépend de la température ;
- la quantité de cécidomyies en diapause est la même pour les trois sous-parcelles ;
- un individu qui émerge dans une sous-parcelle peut se déplacer dans les autres sous-parcelles, il y a cependant un coût relatif à la distance (autrement dit, les cécidomyies préfèrent rester dans la sous-parcelle dans laquelle elles émergent, puis aller dans la sous-parcelle limitrophe plutôt que dans la troisième) ;
- les femelles exogènes arrivent proportionnellement aux inflorescences vivantes dans la parcelles (qui exercent un effet d'attraction) ;
- s'il y a significativement plus de femelles que d'inflorescences, alors elles ne peuvent pas toutes pondre ;
- il y a un taux de mortalité sur les œufs pondus par les femelles, ils ne donneront pas tous des larves.

On reprend le cycle de développement de la cécidomyie présenté dans la partie 2.2, et on essaye de le traduire par des équations simples.

## 4.2 Formalisme

Le schéma conceptuel du modèle est visible sur la figure 4.1. On le détaille ci-dessous.  
Le nombre de larves est donné par

$$L_{t,i} = F_{t-d_\ell, i} \times R \times E_0 \times \mu_\ell.$$

Ici,  $L_{t,i}$  désigne le nombre de larves qui s'éjectent des inflorescences à la date  $t$  dans la sous-parcelle  $i$ ,  $F_{t-d_\ell, i}$  le nombre de femelles qu'il y avait dans la sous-parcelle  $i$  une durée de développement larvaire auparavant (durée qu'il y a entre la ponte des œufs et l'apparition du dernier stade larvaire où les larves s'éjectent des inflorescences pour s'enfoncer dans le sol),  $R$  représente la disponibilité d'inflorescences pour les cécidomyies,  $E_0$  le nombre maximal d'œufs pondus par une femelle et  $\mu_\ell$  la probabilité de survie des œufs et larves. En d'autres termes, cela signifie simplement que le nombre de larves à une certaine date résulte du nombre d'œufs pondus par les femelles, qui ont survécus et se sont développés — sans oublier de prendre en compte que le développement dure quelques jours et que le nombre de femelles qui pond est conditionné à la disponibilité en ressources.

Dans l'équation ci-dessus, le coefficient  $R$  est donné par

$$R = \min \left\{ 1, \frac{k I_{t,i}}{F_{t,i}} \right\}.$$

Ce coefficient traduit la disponibilité en ressources pour les cécidomyies, c'est-à-dire que lorsqu'il n'y a pas suffisamment d'inflorescences dans le verger par rapport au nombre de femelles, alors toutes les femelles ne peuvent pas pondre. Il vaut 1 lorsque toutes les cécidomyies femelles peuvent pondre ; il est compris entre 0 et 1 lorsque certaines cécidomyies femelles ne peuvent pas pondre. Ce coefficient  $R$  dépend du paramètre  $k$  qui indique combien de femelles peut accueillir une inflorescence chaque jour.

Toujours dans l'équation donnant le nombre de larves, on peut noter que le nombre de femelles présente dans la sous-parcelle  $i$  au jour  $t$  est donné par

$$F_{t,i} = F_{t,i \rightarrow i}^{\text{endo}} + F_{t,(j,k) \rightarrow i}^{\text{endo}} + F_{t,i}^{\text{exo}},$$

où  $F_{t,i \rightarrow i}^{\text{endo}}$  sont les femelles qui émergent dans la sous-parcelle  $i$  et qui y restent,  $F_{t,(j,k) \rightarrow i}^{\text{endo}}$  sont les femelles qui ont émergées dans les sous-parcelles  $j$  et  $k$  et qui viennent dans la sous-parcelle  $i$ ,  $F_{t,i}^{\text{exo}}$  sont les femelles exogènes à la parcelle qui viennent dans la sous-parcelle  $i$ . Et le nombre de femelles exogènes est déterminé par

$$F_{t,i}^{\text{exo}} = \gamma \times I_{t,i},$$

avec  $\gamma$  à déterminer. Il y a aussi des échanges, entre les différentes sous-parcelles, de femelles qui ont émergées dans la parcelle. Ainsi, le nombre de femelles qui ont émergées dans la sous-parcelle  $i$  et qui vont dans la sous-parcelle  $j$  est donné par

$$F_{t,i \rightarrow j}^{\text{endo}} = \frac{I_{t,j} \times p_m^{\delta(i,j)}}{\sum_{n \in \{i,j,k\}} I_{t,n} \times p_m^{\delta(i,n)}},$$

où  $p_m \in [0; 1]$  et  $\delta(i, n) = \begin{cases} 0 & \text{si } i = n \text{ (même sous-parcelle),} \\ 1 & \text{si } i \text{ et } n \text{ sont des sous-parcelles limitrophes,} \\ 2 & \text{si } i \text{ et } n \text{ sous-parcelles non-limitrophes.} \end{cases}$

Le paramètre  $p_m$  traduit l'intensité de la migration et le paramètre  $\delta$  traduit «l'effet distance» (autrement dit, les cécidomyies préfèrent se déplacer le moins possible).

On peut noter que les femelles qui émergent comprennent les femelles issues de la phase de pupaison et celles issues de la sortie de diapause. Ainsi, on a

$$F_{t,i}^{\text{endo}} = F_{t,i}^{\text{pupe}} + F_{t,i}^{\text{diap}},$$

avec

$$F_{t,i}^{\text{pupe}} = L_{t-d_p,i} \times \mu_{\text{sol}} \times p_p \times \frac{1}{1 + SR}.$$

On a ici  $F_{t,i}^{\text{pupe}}$  qui désigne le nombre de femelles issues du cycle de développement et qui émergent des pupes présentent dans le sol de la sous-parcelle  $i$  à la date  $t$ ,  $L_{t-d_p,i}$  le nombre de larves qui sont rentrés dans le sol de la sous-parcelle  $i$  une durée de pupaison auparavant,  $\mu_{\text{sol}}$  la probabilité pour une larve d'entrer dans le sol (qui dépend de la modalité de couverture du sol),  $p_p$  la probabilité pour une larve d'entrer en phase de pupaison et d'y survivre et  $SR$  le ratio du nombre de mâles sur le nombre de femelles. Ainsi,  $\frac{1}{1+SR}$  donne la proportion de larves qui se transforment en femelles (les mâles ne pondent pas, ils ne nous intéressent pas). Cela signifie les femelles issues de la phase de pupaison qui émergent à une certaine date correspondent au nombre de larves qui ont réussies à pénétrer dans le sol, qui sont entrées en pupaison, se sont transformées en cécidomyies, ont survécues puis ont émergées — en prenant en compte la durée de développement et la proportion de femelles chez les cécidomyies. Il faut cependant noter que l'on a

$$\mu_{\text{sol}} = \mu_{\text{sol}}^1 \times \mu_{\text{sol}}^2,$$

où  $\mu_{\text{sol}}^1$  correspond à la probabilité de survie d'une larve à la modalité de couverture du sol lorsqu'elle s'enfouit dans le sol et  $\mu_{\text{sol}}^2$  désigne la probabilité de survie à la modalité de couverture du sol d'une femelle qui émerge du sol. Cette distinction est nécessaire car les femelles qui sortent de diapause ne sont impactées par la modalité de couverture du sol qu'à la sortie des femelles (les larves étant entrées dans le sol avant la mise en place du dispositif). Le nombre de femelles qui sortent de diapause est donné par

$$F_{t,i}^{\text{diap}} = D_t \times \frac{1}{1 + SR} \times \mu_{\text{sol}}^2,$$

où  $D_t$  est le nombre de larves en diapause des années précédentes qui sort au jour  $t$ .

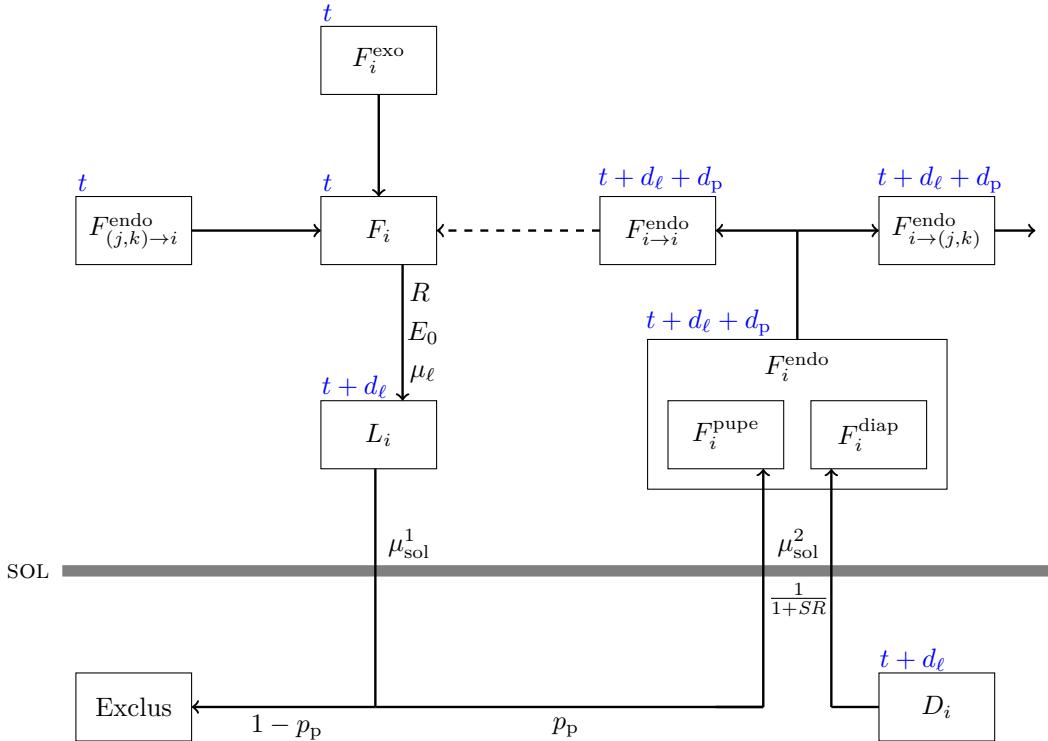


FIGURE 4.1 – Schéma conceptuel du modèle pour la sous-parcelle  $i$ . En bleu est visible la date, la flèche en pointillés marque une rupture du temps.

Le tableau 4.1 recense les paramètres du modèle.

TABLE 4.1 – Les différents paramètres du modèle.

<b>Paramètre</b>	<b>Définition</b>	<b>Valeur</b>
$\gamma$	Paramètre régulant l'arrivée des individus exogènes au verger	Calibration
$p_m$	Paramètre régulant l'intensité des échanges entre sous-parcelles	Calibration
$\mu_{\text{sol}}^1$	Probabilité de survie des larves lorsqu'elles s'enfouissent dans le sol à la modalité de couverture du sol	Calibration
$\mu_{\text{sol}}^2$	Probabilité de survie des femelles qui émergent (de pupaison ou de diapause) la modalité de couverture du sol	Calibration
$k$	Paramètre quantifiant le nombre de femelles que peut accueillir une inflorescence chaque jour	Calibration
$D_t$	Nombre d'individus entrés en diapause les années précédentes qui émergent l'année considérée	Calibration
$E_0$	Nombre maximal d'œufs pondus par une femelle	~150 [Amouroux, 2013]
$\mu_\ell$	Probabilité de survie des œufs et des larves	~0.04 [Amouroux, 2013]
$SR$	<i>Sex-ratio</i>	0.5 [Amouroux, 2013]
$p_p$	Probabilité pour une larve d'entrer en phase de pupaison et d'y survivre	~0.77 [Amouroux <i>et al.</i> , 2014]
$d_\ell$	Durée (en jours) de la période entre la ponte et l'apparition du troisième stade de développement larvaire	7 à 12 [Amouroux, 2013]
$d_p$	Durée (en jours) de la phase de pupaison	4 à 6 [Amouroux, 2013]



# 5

## Calibration du modèle

Il est question ici de définir la méthodologie afin de calibrer les paramètres libres du modèle. On s'intéressera notamment à définir une fonction de coût, pour évaluer la qualité de la calibration. Dans cette optique, on essayera de calibrer les paramètres en utilisant uniquement les dynamiques du verger n°1 ; le deuxième servira à la validation.

### 5.1 Fonction de coût

Une étape importante pour la calibration du modèle est de définir une fonction de coût qui permet de mesurer la qualité de nos estimations. Pour ce faire, on utilisera une fonction permettant de comparer le nombre de larves estimées avec le nombre de larves observées. Pour définir cette fonction, on pose :

- $m$ , le nombre de jours entre la première observation et la dernière ;
- $n$ , le nombre de relevés effectif ;
- $t$ , le nombre de jours passés depuis la première observation ;
- $t^j$ , le nombre de jours entre la première observation et le  $j^{\text{ème}}$  relevé.

(On a donc  $t^1 = 0$  et  $t^n = m$ .)

La fonction est alors définie par

$$f(y, \hat{y}) = \sqrt{\frac{\frac{1}{n-1} \sum_{j=2}^n (y_j^* - \hat{y}_j^*)^2}{\max(y^*) - \min(y^*)}},$$

où

$$y_j^* = y_{t^j} \quad \text{et} \quad \hat{y}_j^* = \frac{1}{t^j - t^{j-1}} \sum_{k=t^{j-1}}^{t^j} \hat{y}_k.$$

La figure 5.1 illustre le fonctionnement de la fonction. Appliquer la fonction à  $n - 1$  valeurs (correspondant aux relevés sur le terrain) plutôt qu'à chacun des  $m$  jours (correspondant à l'étendue des relevés) est motivé par le fait que les relevés ne furent pas fait à des intervalles de temps régulier. Ce faisant, on évite d'attribuer plus d'importance aux relevés qui ont eu un écart relativement important avec le relevé précédent.

Par la suite, l'objectif sera de minimiser cette fonction pour chacune des trois sous-parcelles.

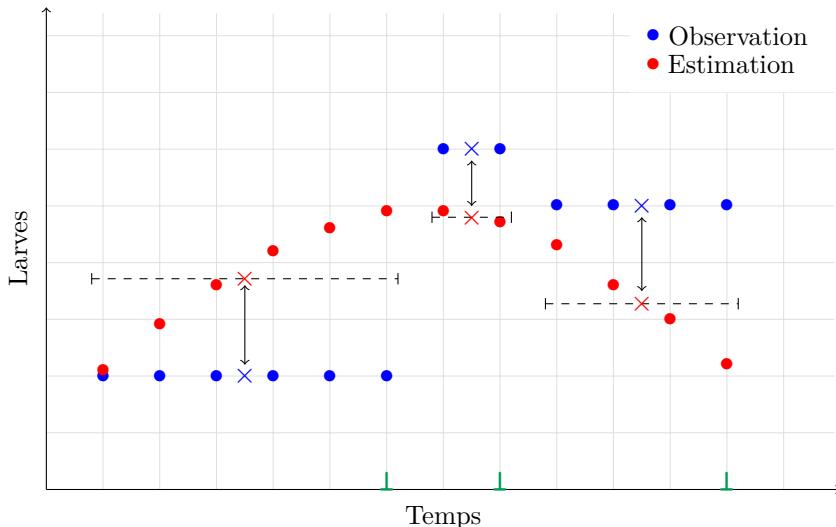


FIGURE 5.1 – Schéma illustrant le fonctionnement de la fonction objectif. À chaque relevé effectif (marqueurs verts), on fait correspondre la période correspondant à ce relevé (segments en pointillés). Et pour chacune de ces périodes, on calcule la moyenne des valeurs estimées (les croix rouges). On compare ensuite les moyennes ainsi calculées avec les valeurs observées associées (les croix bleues).

## 5.2 Analyse de sensibilité

Avant de calibrer le modèle, il est pertinent d'effectuer une analyse de sensibilité. L'analyse de sensibilité est définie par [Saltelli et al. \[2004\]](#) comme

«l'étude de comment l'incertitude de la sortie d'un modèle — qu'elle soit numérique ou non — peut être répartie entre les différentes sources d'incertitudes présentes dans les entrées du modèle.»<sup>1</sup>

Autrement dit, on cherche à connaître les paramètres les plus influants sur la sortie du modèle. La calibration comportant toujours une part d'arbitraire, cette analyse permet de prendre du recul sur les choix de paramètres, relativement à leur impact sur les sorties du modèle.

Il existe deux grandes catégories d'analyses de sensibilité, celles qui ont une approche globale et celles qui ont une approche locale (parfois appelées *one-at-the-time*). L'approche locale consiste à étudier la sensibilité des paramètres les uns après les autres, les uns indépendamment des autres. Cette approche est valable si et seulement si le modèle est linéaire par rapport à chacune de ses entrées  $x_i$  et qu'il n'y a aucune interaction entre les différentes entrées du modèle [\[Saltelli et al., 2019\]](#). Dès lors qu'il y a la moindre incertitude sur la linéarité du modèle ou sur la non-interaction entre les paramètres, il faut privilégier une approche globale. Notre modèle n'est pas linéaire et il n'y a dans notre cas aucune raison de supposer la non-interaction entre nos paramètres, bien au contraire. On utilisera donc une approche globale.

Bien qu'il existe plusieurs méthodes ayant une approche globale, elles ont toutes en commun de fonctionner dans un cadre non-linéaire et de prendre en compte les différentes interactions entre les différents paramètres. Parmi les méthodes les plus connues et les plus

1. «The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.»

utilisées, on peut en citer qui fonctionnent par décomposition de la variance comme Sobol ou FAST (Fourier Amplitude Sensitivity Test) ou d'autres qui fonctionnent en effectuant des perturbations élémentaires des entrées du modèle comme la méthode Morris. Notre modèle possède moins de 20 paramètres et s'exécute en moins d'une minute, nous utiliserons alors la méthode Sobol conformément aux recommandations de [Saltelli et al. \[2008, chap. 6\]](#).

Le fonctionnement de cette méthode est relativement intuitif. On considère que la sortie de notre modèle  $Y$  peut s'exprimer comme une fonction des entrées de notre modèle  $X$ , c'est-à-dire  $Y = f(X)$ . Comme  $X = (X_1, \dots, X_n)$  peut prendre un nombre important de valeurs possibles, il en résulte qu'il y a *a priori* un nombre important de résultats possibles pour  $Y$ . L'objectif est alors de déterminer quelles entrées du modèle  $X_i$  induisent le plus de changements dans ces résultats possibles — dans cette variance qu'admet  $Y$ . À cette fin, on utilise les indices principaux de Sobol définis par

$$S_i = \frac{\mathbf{Var}(\mathbf{E}[Y|X_i])}{\mathbf{Var}(Y)}.$$

Ici,  $\mathbf{Var}(\mathbf{E}[Y|X_i])$  permet de voir l'effet du seul paramètre  $X_i$  sur la variance de  $Y$ . Diviser par la variance totale de  $Y$  permet de le faire relativement aux autres paramètres. Cet indice ne prend cependant pas en compte les interactions entre  $X_i$  et les autres entrées du modèle. Il faut pour ça utiliser les indices totaux de Sobol définis par

$$S_i^T = 1 - \frac{\mathbf{Var}(\mathbf{E}[Y|X_{\sim i}])}{\mathbf{Var}(Y)},$$

où  $X_{\sim i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ . L'indice  $S_i^T$  ainsi défini permet lui de prendre en compte l'impact du paramètre  $X_i$  et de ses interactions sur la variance de  $Y$ . À noter que l'interaction entre  $X_i$  et  $X_j$  est à la fois prise en compte par  $S_i^T$  et  $S_j^T$ . C'est pour cette raison qu'il est souvent pertinent d'interpréter les indices principaux et les indices totaux conjointement.

Une fois que l'on connaît la sensibilité du modèle aux différents paramètres, on peut alors passer à la calibration desdits paramètres.

### 5.3 Algorithme d'optimisation

L'objectif de l'optimisation est de trouver les jeux de paramètres qui minimisent notre fonction de coût, ceux qui permettent d'ajuster au mieux les dynamiques de larves simulées aux dynamiques observées. Pour ce faire, on a besoin d'un algorithme d'optimisation. Et il y a du choix.

On peut déjà noter que nous avons sept paramètres à calibrer, et qu'ils évoluent tous dans des intervalles (que l'on définira plus tard). Il apparaît évident que toutes les solutions ne pourront pas être testées. Une solution peut être d'utiliser un algorithme basé sur une méthode MCMC comme l'algorithme du recuit simulé. Nous avons cependant trois dynamiques à ajuster (une pour chaque sous-parcelle), il faudrait alors minimiser la somme (ou la moyenne) des trois fonctions objectifs. Ou alors on peut directement utiliser un algorithme d'optimisation multicritères.

Ces algorithmes ont l'avantage de pouvoir optimiser simultanément des objectifs qui ne sont pas toujours comparables — à des échelles différentes, par exemple. Un des plus connus est sans conteste NSGA-II (Nondominated Sorting Genetic Algorithm II) [[Deb et al., 2002](#)]. C'est celui que l'on utilisera.

C'est un algorithme qui ne renvoie pas une unique solution mais un ensemble de solutions. Cet ensemble de solutions converge vers un sous-ensemble du front de Pareto. Le front de Pareto désigne l'ensemble des solutions non-dominées pour un problème donné. Dans  $P \subset \mathbf{R}^p$  ( $p > 1$ ), une solution  $x^* = (x_1, \dots, x_p) \in P$  est dite non-dominée lorsque

$$\{x \in P \mid \forall i \in \{1, \dots, p\}, x_i \succcurlyeq x_i^* \text{ et } \exists i \text{ tel que } x_i \succ x_i^*\} = \emptyset,$$

où  $\succcurlyeq$  et  $\succ$  veulent respectivement dire «au moins aussi bien que» et «meilleur que». Dans notre cas, vu que l'on veut minimiser notre fonction de coût,  $x^* \succcurlyeq x$  se traduira par  $f(x^*) \leq f(x)$  et  $x^* \succ x$  se traduira par  $f(x^*) < f(x)$ , où la fonction  $f$  représente la composée de notre modèle suivi de notre fonction de coût. En d'autres termes,  $x$  est non-dominée signifie qu'il n'existe pas de solution  $y$  qui soit strictement meilleure sur un des critères et au moins aussi bonne sur tous les autres critères.

Le front de Pareto étant un sous-ensemble de  $\mathbf{R}^p$ , il n'est (*a priori*) pas dénombrable. De ce fait, NSGA-II ne peut pas renvoyer le front tout entier mais seulement un sous-ensemble. C'est un algorithme itératif, les itérations seront appelées ici *générations*. Et il faut plusieurs générations pour se retrouver avec des solutions non-dominées. À la génération  $t$ , l'algorithme effectue plusieurs opérations.

1. On possède un ensemble de solutions potentielles que l'on nommera *population*. La taille de la population  $N$  est fixée arbitrairement.
2. À chaque solution de cette population  $P_t$ , on attribue une solution fille (on verra plus tard comment).

Avec la population  $P_t$  et les solutions filles associées  $F_t$ , cela forme un ensemble de solutions potentielles  $S_t$  de taille  $2N$ . On va alors sélectionner parmi ces solutions les  $N$  solutions qui nous intéressent le plus.

3. Pour ce faire, pour chaque solution potentielle  $s \in S_t$  :
  - on calcule le nombre  $n_s$  d'éléments de  $S_t$  qui dominent  $s$  ;
  - on établit l'ensemble  $D_s$  des autres solutions de  $S_t$  qui sont dominées par la solution  $s$ .
4. On répartit alors nos solutions en plusieurs groupes, en plusieurs *fronts*. Vont ainsi dans le premier front  $F_1$  les solutions  $s$  qui vérifient  $n_s = 0$ , qui ne sont donc pas dominées par d'autres solutions de  $S_t$ . Dans le deuxième front  $F_2$  se trouvent les solutions qui sont dominées uniquement par les solutions présentes dans  $F_1$ . (Autrement dit, les solutions qui se trouvent dans et uniquement dans  $\bigcup_{s \in F_1} D_s$ .) On continue de la même manière pour  $F_3$  (qui contient les solutions uniquement dominées par celles de  $F_1$  et  $F_2$ ),  $F_4$ ... jusqu'à ce que toutes les solutions de  $S_t$  soient assignées à un front.
5. On choisira alors de garder en priorité les solutions de  $F_1$  puis de  $F_2$  et ainsi de suite jusqu'à en obtenir  $N$ . Il faut noter qu'il existe probablement un front  $F_k$  où toutes les solutions ne pourront être garder pour ne pas dépasser la limite imposée de  $N$  solutions. Plutôt que de choisir aléatoirement le nombre nécessaires de solutions dans  $F_k$ , seront sélectionnées les solutions qui maximisent une certaine distance — the *crowding distance* — avec les autres solutions. (On n'entrera pas dans le détail ici.) C'est fait pour assurer une certaine diversité entre les différentes solutions. Les  $N$  solutions ainsi sélectionnées constitueront  $P_{t+1}$ , la population initiale de la génération suivante  $t + 1$ .

Un élément important pour assurer une convergence vers le front de Pareto est le choix des solutions filles  $F_t$  effectué à chaque génération. Pour générer une solution fille, plusieurs étapes sont là aussi nécessaires. Ces étapes (simplifiées) sont :

1. On sélectionne deux solutions  $s_1$  et  $s_2$  dans  $P_t$ .
2. On construit une solution fille  $f_1$  en prenant la moitié des coordonnées de  $s_1$  et l'autre moitié de  $s_2$ . On construit  $f_2$  en prenant les moitiés de  $s_1$  et  $s_2$  inutilisées dans la construction de  $f_1$ .
3. Sur chacune des solutions filles, une coordonnée choisie aléatoirement est remplacée par une valeur aléatoire choisie dans un intervalle centré sur la valeur initiale.
4. On réitère ces étapes jusqu'à avoir  $N$  solutions filles.

Et c'est ainsi qu'après un nombre suffisant d'itérations, et pour une taille de population suffisamment grande, l'algorithme NSGA-II renvoie un ensemble de points suffisamment proche du front de Pareto et qui en retranscrit sa diversité.

Dans notre cas, cela signifie que l'on possède différents jeux de paramètres et les valeurs de la fonction de coût pour les trois sous-blocs qu'ils produisent. Et qu'il ne reste plus qu'à faire un choix.

## 5.4 Exploration de l'ensemble des solutions

Choisir une solution n'est cependant pas trivial. En effet, parmi les solutions possibles, aucune n'est objectivement meilleure que les autres. Il en découle que le choix sera forcément arbitraire.

Plusieurs approches sont possibles. On peut par exemple choisir au hasard un nombre restreint de solutions possibles, les tester puis sélectionner celle qui semble la plus pertinente. On peut aussi sélectionner la solution qui minimise une norme sur nos trois critères, en estimant qu'elle représente un bon compromis entre les différents critères.

On préférera une autre approche. Puisque NSGA-II essaye de renvoyer des solutions couvrant au maximum le front de Pareto, cela implique que certaines solutions seront proches en termes de valeurs pour nos paramètres. Et qu'elles renverront donc des dynamiques similaires. Il peut être donc pertinent d'effectuer une classification non-supervisée de nos solutions afin de recenser les différentes solutions-types. Pour ce faire, on peut effectuer une classification ascendante hiérarchique (CAH).

Le principe est le suivant. On traduit nos solutions centrées – réduites dans un espace métrique. On choisira  $\mathbf{R}^p$  muni de la distance euclidienne. On cherche à rassembler les différentes solutions en différentes classes. Les classes se doivent de contenir des individus aussi proches que possible (*i.e.* la distance euclidienne entre un individu de la classe et l'individu moyen de la classe doit être petite), et être aussi différentes des autres classes que possible (*i.e.* la distance entre l'individu moyen de la classe et l'individu moyen de la classe la plus proche doit être grande). D'un point de vue formel, cela peut s'exprimer par la décomposition de l'inertie totale en inertie externe des classes et en inertie interne des classes. Cette décomposition est donnée par

$$\underbrace{\frac{1}{n} \sum_{i=1}^n \|x_i\|^2}_{\text{inertie totale}} = \underbrace{\sum_{k=1}^K w^k \|\bar{x}^k\|^2}_{\text{inertie externe}} + \underbrace{\sum_{k=1}^K \sum_{x_i \in C_k} \frac{1}{n} \|x_i - \bar{x}^k\|^2}_{\text{inertie interne}},$$

où  $K$  représente le nombre de classes,  $w^k$  le poids de la classe  $C_k$  et  $\bar{x}^k$  l'individu moyen de la classe  $C_k$ . L'objectif est alors de trouver les classes  $C_k$  qui minimisent l'inertie interne et qui maximise l'inertie externe. Cependant, les classes  $C_k$  dépendent du nombre de classes  $K$ , et ce nombre est choisi par l'utilisateur de la méthode. Il est évident que plus il y a de classes, plus l'inertie intra-classe sera faible ; néanmoins la classification a pour but de

rassembler les individus, avoir maintes classes n'est alors pas très pertinent. On reviendra sur le choix de  $K$  plus tard.

Initialement, la CAH commence avec  $n$  classes : chaque individu se trouve dans une classe ne contenant que lui. Et à chaque étape, on regroupe deux classes similaires jusqu'à n'en avoir plus qu'une unique contenant tous les individus. Les deux classes qui sont regroupées à chaque étape sont choisies car elles minimisent un *indice d'agrégation*. Il en existe un certain nombre, nous choisirons comme indice d'agrégation l'indice de Ward. Il est défini par

$$\mu(C_k, C_\ell) = \frac{w^k w^\ell}{w^k + w^\ell} \|\bar{x}^k - \bar{x}^\ell\|^2,$$

où  $w^k$  est le poids de la classe  $k$  et  $\bar{x}^k$  est l'individu moyen de la classe  $k$ . Cet indice a l'avantage d'être exprimable en fonction de l'inertie, et notamment de l'inertie intra-classes que l'on souhaite minimiser [Bry, 2017].

Ainsi, la CAH donne les classes  $C_k$  pour tous les nombres de classes possibles, *i.e.*  $K = 1, \dots, n$ . Il ne reste plus qu'à choisir  $K$ . Agréger deux classes a un certain coût qui se traduit par une augmentation de l'inertie intra-classes. Donc *a fortiori*, réduire le nombre de classes est une bonne chose si l'augmentation de l'inertie intra-classes est minime. En pratique, on choisira un nombre de classes  $K^*$  qui minimisera significativement l'inertie intra-classes par rapport à celle obtenue avec  $K^* - 1$  classes.

Il faut aussi noter que dans notre cas précis, cette classification est faite dans un but exploratoire. Il vaudra mieux avoir un nombre de classes délibérément grand, quitte à avoir des classes semblables, afin de ne pas passer à côté d'une catégorie de solutions intéressante.

# 6

## Mise en œuvre et résultats



# 7

## Conclusion

## Remerciements

Ce rapport de stage a été réalisé dans le cadre du projet ECOVERGER, action pilotée 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 et dans le cadre du programme de recherche agronomique du Cirad à la Réunion, DPP COSAQ, (activités 2015-2018) financé par la communauté européenne (fond structurel FEDER) et le Conseil Régional de La Réunion.



MINISTÈRE  
DE LA TRANSITION  
ÉCOLOGIQUE  
ET SOLIDAIRE



MINISTÈRE  
DE L'AGRICULTURE  
ET DE  
L'ALIMENTATION



AGENCE FRANÇAISE  
POUR LA BIODIVERSITÉ  
ÉTABLISSEMENT PUBLIC DE L'ÉTAT



# Bibliographie

- P. Amouroux. *Bio-écologie et dynamique des populations de cécidomyie des fleurs (*Procontarinia mangiferae*), un ravageur inféodé au manguier (*Mangifera indica*), en vue de développer une lutte intégrée.* PhD thesis, La Réunion, 2013.
- P. Amouroux, F. Normand, and D. Vincenot. Guide de production intégrée de mangues à la Réunion, chapitre VII, 2009.
- P. Amouroux, F. Normand, H. Delatte, A. Roques, and S. Nibouche. Diapause incidence and duration in the pest mango blossom gall midge, *Procontarinia mangiferae* (Felt), on Reunion Island. *Bulletin of entomological research*, 104(5) :661–670, 2014.
- F. Boudon, S. Persello, A. Jestin, A.-S. Briand, P. Fernique, Y. Guédon, M. Léchaudel, I. Grechi, and F. Normand. An FSPM approach for modeling fruit yield and quality in mango trees. FSPMA, 2016.
- X. Bry. Cours d'analyse de données multidimensionnelles, 2017.
- K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm : NSGA-II. *IEEE transactions on evolutionary computation*, 6(2) :182–197, 2002.
- Food and Agriculture Organization. <http://www.fao.org/faostat/fr/#data/QC>, 2017.  
ONU. Site internet visité le 25 juin 2019.
- C. Magne. Effet de la charge en fruits sur la croissance végétative de plusieurs variétés de manguier à l'île de la Réunion. Master's thesis, Université Blaise Pascal, 2004.
- F. Normand. Guide de production intégrée de mangues à la Réunion, chapitre IV, 2009.
- F. Normand, A. K. P. Bello, C. Trottier, and P.-E. Lauri. Is axis position within tree architecture a determinant of axis morphology, branching, flowering and fruiting? An essay in mango. *Annals of botany*, 103(8) :1325–1336, 2009.
- L. Saint-Criq. Modélisation du système manguier-cécidomyie des fleurs pour une évaluation de modes de gestion du ravageur et de ses dégâts. Master's thesis, Université Paul Sabatier, 2018.
- A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. Sensitivity analysis in practice : a guide to assessing scientific models. *Chichester, England*, 2004.
- A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis : the primer*. John Wiley & Sons, 2008.

A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu.  
Why so many published sensitivity analyses are false : A systematic review of sensitivity  
analysis practices. *Environmental modelling & software*, 114 :29–39, 2019.

# A

## Verger n°2

On présente ici le dispositif présent sur le verger expérimental n°2 ainsi que les dynamiques d'inflorescences et de larves obtenues grâce aux données des relevés effectués sur cette parcelle.

Le dispositif mis en place était similaire à celui du verger n°1, à la différence de l'ordre des modalités de couverture du sol (voir figure A.1).

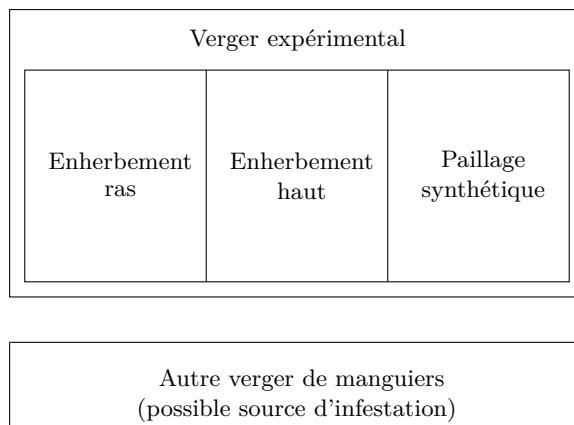


FIGURE A.1 – Schéma de l'expérimentation menée. Le verger sur lequel ont été testées les trois modalités de couvertures du sol était situé à côté d'un autre verger. Bien que cet autre verger n'avait pas de rôle particulier, il a probablement servi de source d'infestation exogène au verger expérimental.

Le verger n°2 présente des caractéristiques particulières, rajoutant des différences supplémentaires entre les différentes sous-parcelles, en plus des différentes modalités. Ce verger est disposée «en escalier», avec une modalité sur chaque «marche». L'ensoleillement n'était pas le même entre les différentes modalités, et cela se ressent sur la figure A.2. La sous-parcelle avec un paillage synthétique — la plus ensoleillée — présente ainsi plus d'inflorescences que les deux autres. On retrouve aussi des différences d'inflorescences entre les parcelles enherbées, celle avec l'enherbement haut avait plus de soleil que l'autre, et elle a eu plus d'inflorescences.

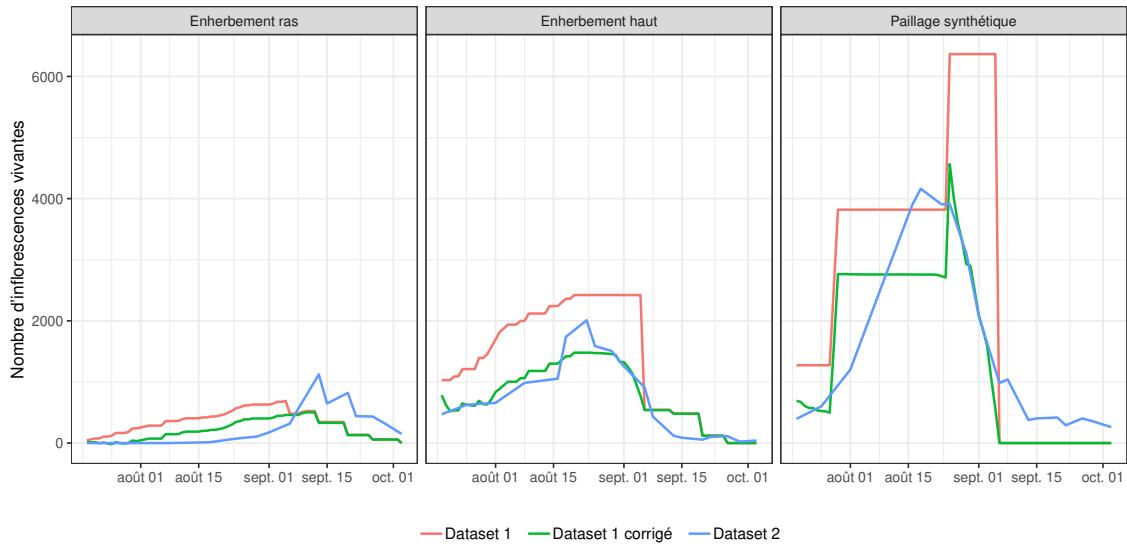


FIGURE A.2 – Comparaison des différentes dynamiques d’inflorescences vivantes du verger n°2 en fonction du *dataset* utilisé. Même après correction, on observe des différences entre les dynamiques issues des différents jeu de données, en particulier pour les deux premières modalités.

On peut noter aussi des différences entre les deux jeux de données. C'est flagrant pour la modalité «paillage synthétique», où il n'y a seulement que 5 observations dans le *dataset 1* produisant cette dynamique particulière. À cela se rajoute la variabilité du phénomène donnant des dynamiques très différentes pour des échantillonnages différents. Seule la sous-parcelle avec un enherbement haut présente des dynamiques similaires entre les deux jeux de données.

Les dynamiques de larves pour ce verger sont visibles sur la figure A.3.

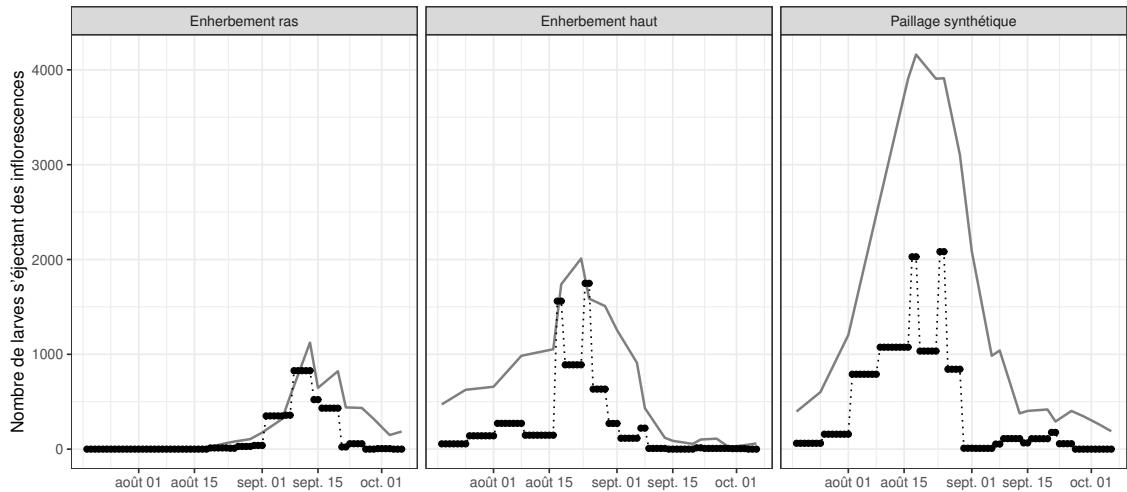


FIGURE A.3 – Dynamiques de larves s’éjectant des inflorescences des manguiers chaque jour dans le verger n°2 pour chacune des trois sous-parcelles. En gris sont visibles les dynamiques d’inflorescences vivantes (issues du *dataset* 2,  $I_t^2$ ).