



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

Bastien Reyné

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



Année universitaire 2018/2019

Résumé

Le manguier est un arbre fruitier qui a une floraison étalée dans le temps. Cela produit des conditions favorables au développement des ravageurs, dont la cécidomyie des fleurs qui s'attaque aux inflorescences et fait donc beaucoup de dégâts sur la production. Une expérimentation a été menée sur des vergers en 2017 pour tester différentes modalités de couverture du sol (car la cécidomyie se développe en partie dans le sol). L'acquisition de données est cependant difficile, la cécidomyie étant un insecte volant ; les données disponibles quantifient donc le nombre de larves s'éjectant des inflorescences. À partir de ces données et des connaissances présentes dans la littérature, un modèle décrivant la dynamique de population de cécidomyie des fleurs à différents stades (larves et adultes) en fonction de la dynamique de population d'inflorescences a été établi sous la forme d'équations. Il prend en compte les différentes étapes du cycle de développement de la cécidomyie des fleurs : ponte des œufs, développement larvaire, enfouissement des larves dans le sol, phase de pupaison et émergence des adultes dans le verger. Le modèle a été calibré avec un algorithme génétique multicritères (NSGA-II) et les solutions optimales ont été analysées. Ce modèle a permis de recréer les dynamiques de larves capturées de manière convaincante. Le modèle semble aussi indiquer que les cécidomyies sont soumises à un phénomène de saisonnalité, autre que l'absence de ressources ou un changement de température, qui met un terme à leur prolifération en fin de saison.

Remerciements

Après six mois de stage, je me dois de remercier quelques personnes. Et en premier lieu mes tuteurs, Isabelle et Fred. Pour beaucoup de choses. Notamment les nombreuses relectures du présent mémoire. Plus généralement, merci pour ce stage qui fut très intéressant, autant d'un point de vue professionnel que personnel. Si c'était à refaire, je le referais sans aucune hésitation.

Je souhaite aussi remercier Mohammed pour ton apport sur ce stage. Tes conseils me furent utiles, et nos échanges profitables. Alain, merci pour la sortie terrain, merci aussi pour toutes tes connaissances entomologiques qui me furent utiles à une meilleure compréhension du sujet. Frédéric, merci pour tes connaissances agronomiques et entomologiques, ainsi que pour quelques de tes photos qui parsèment ce rapport.

À Montpellier, j'aimerais remercier l'équipe M2P2 pour leur accueil, pour ces trois mois qui se sont déroulés dans un climat très agréable.

À la Réunion, des remerciements s'imposent aussi. Merci Sylvie, pour ta gentillesse d'abord, puis ensuite pour avoir rendu toutes ces formalités administratives, d'ordinaires si compliquées, si faciles. Merci André pour ton accueil et ton indéfectible disponibilité. Merci Antoine pour tes scripts fonctionnels. Merci Gwenvael pour m'avoir tenu la branche.

Difficile de ne pas mentionner ici toutes les personnes que j'ai pu côtoyer à la Réunion, qu'elles furent stagiaires, doctorants ou dans une autre catégorie. Merci Coco, Gwen, Maël et les autres. Pour cette bonne ambiance qui régna partout : à la station, dans les cases, en rando ou ailleurs. Pour ces instants *chill* qui surgirent n'importe quand, lors d'une pause au travail, lors d'une journée à la plage, entre la *pera e il formaggio*, entre les lasagnes et le gâteau. Le séjour lété bon !

Merci Robin, pour tes quelques relectures, et aussi pour ces discussions autour d'une bière qui ont pu provoquer quelques conseils et commentaires pertinents.

Enfin, j'aimerais remercier aussi l'ensemble des professeurs de l'IMAG pour leurs enseignements, de la L1 au M2, qui furent bien souvent intéressants.

Table des matières

Remerciements	iii
Table des matières	iv
1 Introduction	1
2 Connaissances biologiques	3
2.1 Le mangueier	3
2.2 La cécidomyie des fleurs	5
3 Données expérimentales disponibles	7
3.1 Dispositif expérimental et suivis	7
3.2 Données	8
4 Le modèle	13
4.1 Approche retenue et hypothèses	13
4.2 Formalisme	14
5 Calibration du modèle	19
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
6 Mise en œuvre et résultats	25
6.1 Solutions pour le modèle initial	25
6.2 Solutions avec prise en compte de la température	31
6.3 Solutions avec prise en compte de contraintes liées aux ressources	34
6.4 Solutions avec l'introduction d'un paramètre de saisonnalité	37
7 Discussion	39
8 Conclusion	41
Bibliographie	43
A Verger n°2	45
B Simulation d'inflorescences attractives	47

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 mangue¹, 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.

Pour satisfaire ces objectifs on s'intéressera dans le chapitre 2 aux connaissances biologiques nécessaires à la compréhension du sujet avant de revenir dans le chapitre 3 sur l'expérimentation réalisée et les données que nous possédons. On formalisera dans le chapitre 4 notre modèle, et on expliquera dans le chapitre 5 comment en calibrer les paramètres. Enfin, on explorera les résultats dans le chapitre 6 avant de les discuter dans le chapitre 7.

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.

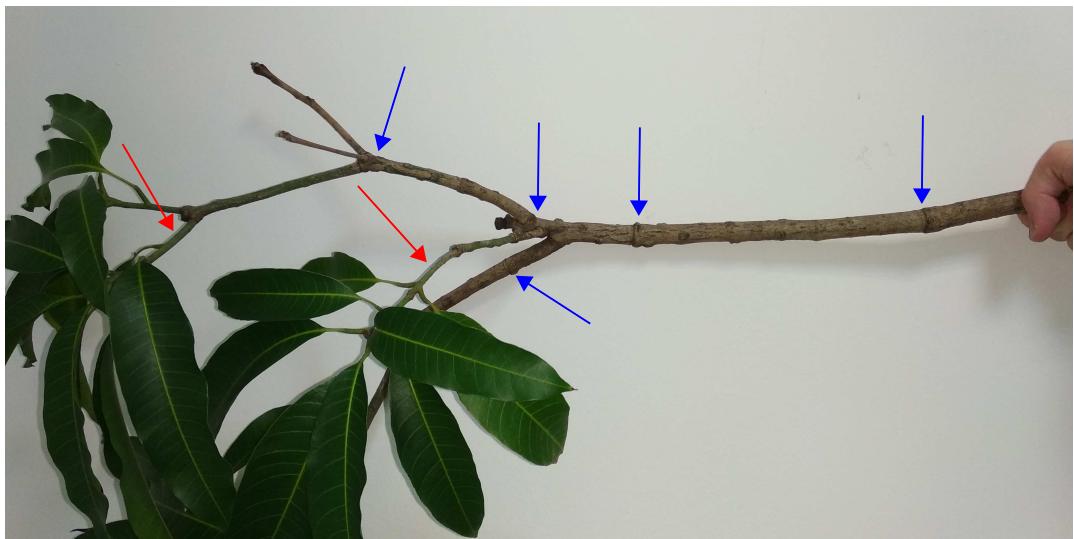


FIGURE 2.1 – Photographie d'une branche portant des unités de croissance. Les flèches rouges montrent des unités de croissance feuillés. Les flèches bleues montrent les délimitations des unités de croissance antérieures; la portion de la branche entre deux flèches bleues correspond à une ancienne unité 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 inflorescences. 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 manguier (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.

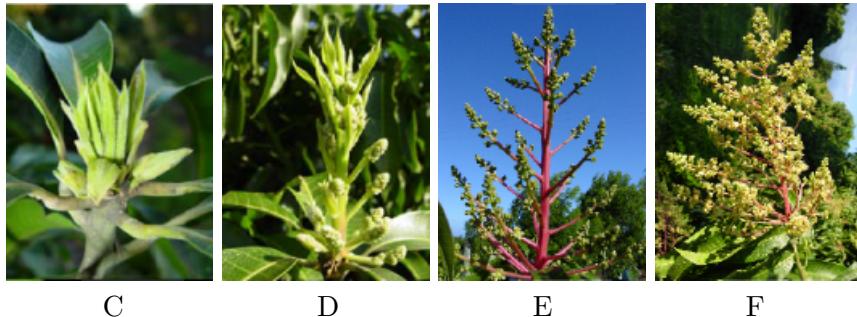


FIGURE 2.3 – Les stades phénologiques C à F d'une inflorescence de manguier (photos : F. Normand)

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.

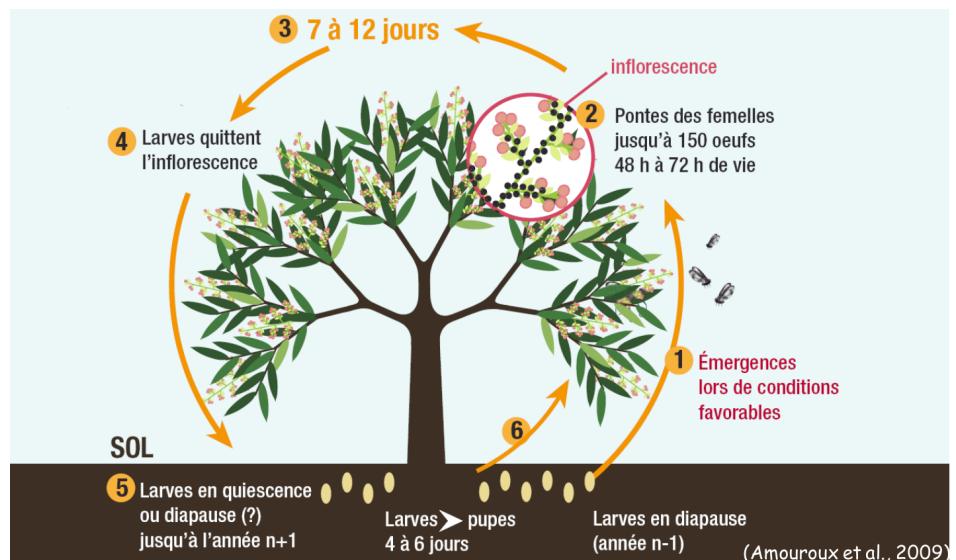


FIGURE 2.4 – Représentation du cycle de développement de la cécidomyie des fleurs du manguier par [Amouroux et al. \[2009\]](#).

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 due 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

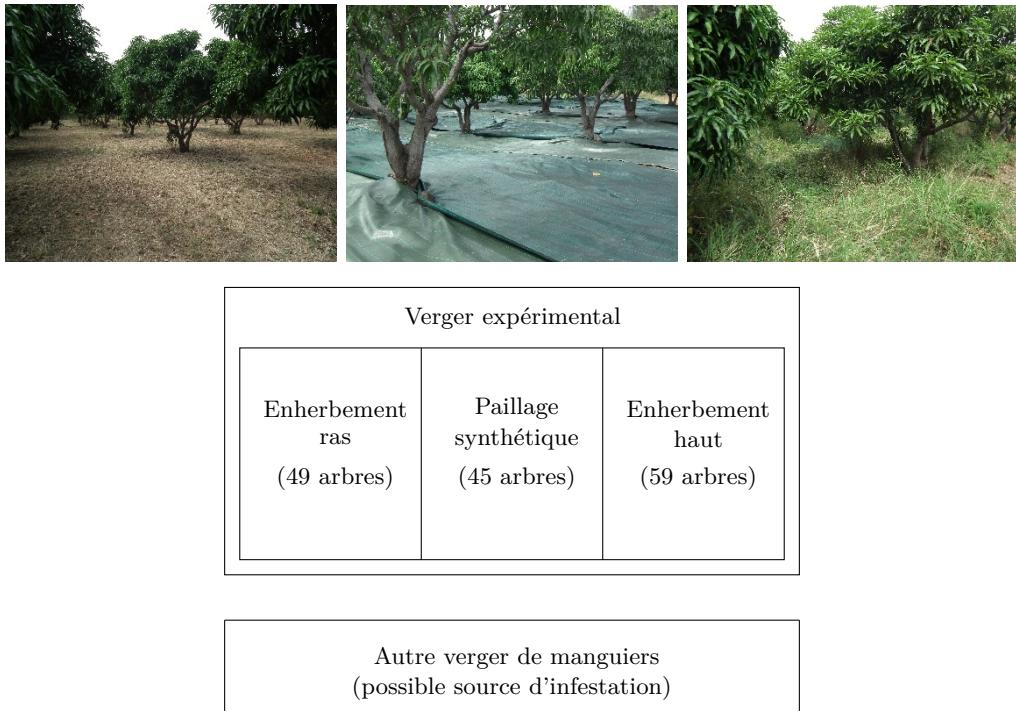


FIGURE 3.1 – Description du dispositif expérimental de la parcelle n°1. En haut, photographie des trois modalités de couverture du sol : enherbement ras, paillage synthétique et enherbement haut. En bas, schéma du dispositif expérimental et de la configuration des sous-parcelles. Le verger sur lequel 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 données¹, 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 α 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 α . 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 α 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.

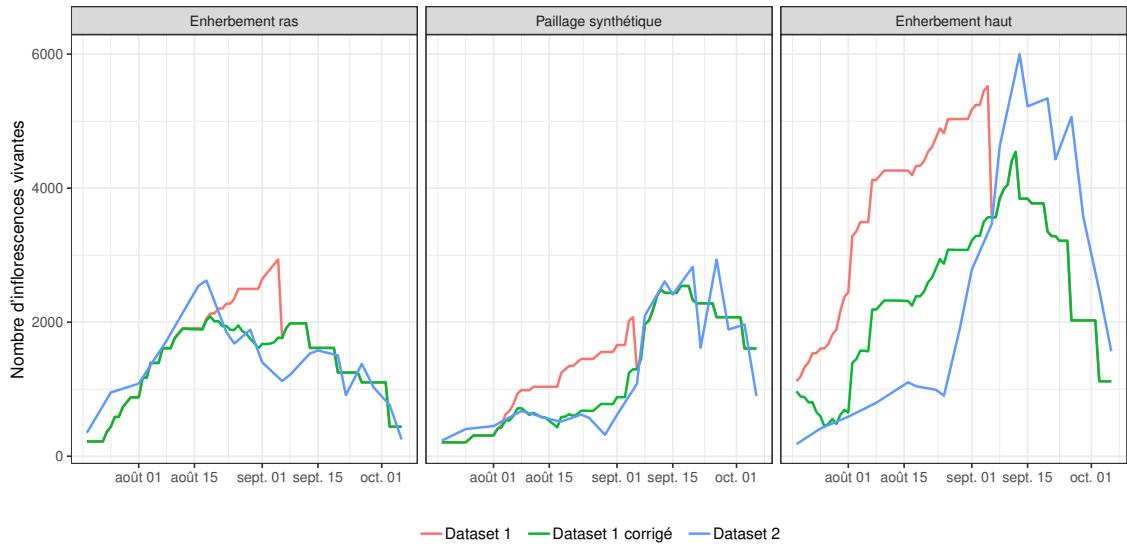


FIGURE 3.2 – Comparaison des différentes dynamiques d’inflorescences vivantes du verger n°1 en fonction du dataset utilisé.

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. On notera qu’une hypothèse est faite : toutes les larves qui s’éjectent des inflorescences au-dessus des pièges sont piégées. Il y a également 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 dynamique de larves peut donc s’obtenir en utilisant la formule

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

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 , $I_{k,t}^{2,p}$ le nombre d’inflorescences au-dessus des pièges dans l’arbre k à la date t et on définit $L_{k,t}$, le nombre de larves dans les pièges de l’arbre k à la date t , par

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

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.

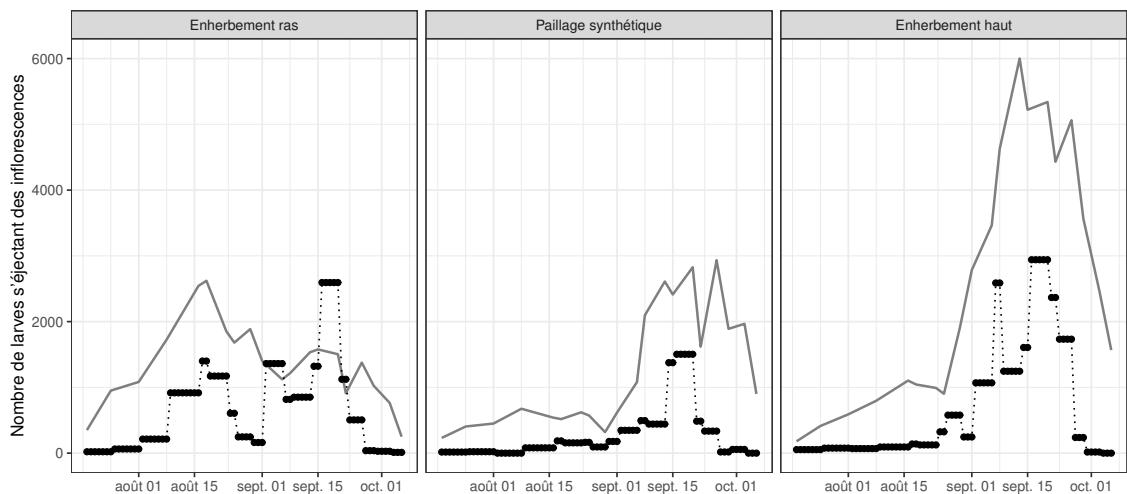


FIGURE 3.3 – Dynamiques de larves s'éjectant des inflorescences des manguiers chaque jour dans le verger n°1 pour chacune des trois sous-parcelles. En gris sont visibles les dynamiques d'inflorescences vivantes (issues du dataset 2, I_t^2).

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'oeufs 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 une rupture de cycle de développement pour les cécidomyies après l'éjection des larves 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 ;
- il y a un taux de mortalité pour les larves lorsqu'elles rentrent et pour les adultes lorsqu'ils sortent du sol. Et ce taux est différent entre l'enherbement ras et l'enherbement haut. Pour le paillage synthétique il est égal à 1 (*i.e.*, aucune cécidomyie ne rentre dans le sol ou ne sort du sol) ;
- les larves qui n'entrent pas en pupaison (soit qui meurent, soit qui rentrent en diapause) sont exclues du système ;
- 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 les sous-parcelles (qui exercent un effet d'attraction) ;
- la reproduction des femelles (ponte et/ou le développement des larves) est limitée par la ressource disponible, c'est à dire le nombre d'inflorescences par femelle.

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.

Ce schéma traduit le cycle de développement de la cécidomyie détaillé dans la partie 2.2. Les femelles présentes dans la sous-parcelle i sont représentées par F_i . Les femelles sont soit exogènes, soit endogènes. On distinguera les femelles endogènes en fonction de la sous-parcelle dans laquelle elles émergent. Il en résulte que dans la sous-parcelle i , les femelles ayant pondus les larves sont celles qui viennent de l'extérieur du verger, celles qui ont émergées dans la sous-parcelle considérée et qui y restent et celles qui ont émergées dans la sous-parcelle j ou k et qui viennent dans la sous-parcelle i .

Les œufs, pondus par les femelles F_i , permettent l'émergence des larves L_i , après la durée de développement nécessaire, notée d_ℓ . Les larves s'éjectent ensuite en direction du sol. Si elles n'entrent pas en pupaison, alors elles sont exclues de notre système. Si elles entrent en pupaison, alors des femelles émergeront dans la sous-parcelle après une durée de développement d_p .

Les femelles qui émergent ainsi constituent, avec les femelles qui émergent de diapause, les femelles endogènes de la sous-parcelle i , notées sur le schéma F_i^{endo} . Et elles se répartissent entre les trois sous-parcelles i , j et k .

Ceci est le schéma global de notre modèle. Il faut maintenant la formaliser. On commence par les larves. Le nombre de larves à une certaine date résulte du nombre d'œufs pondus par les femelles, qui ont survécus et se sont développés. Ce développement dure quelques jours et le nombre de femelles qui pond est conditionné à la disponibilité en ressources. Formellement, cela peut se traduire par

$$L_{t,i} = F_{t-d_\ell,i} \times R \times E_0 \times \mu_\ell. \quad (4.1)$$

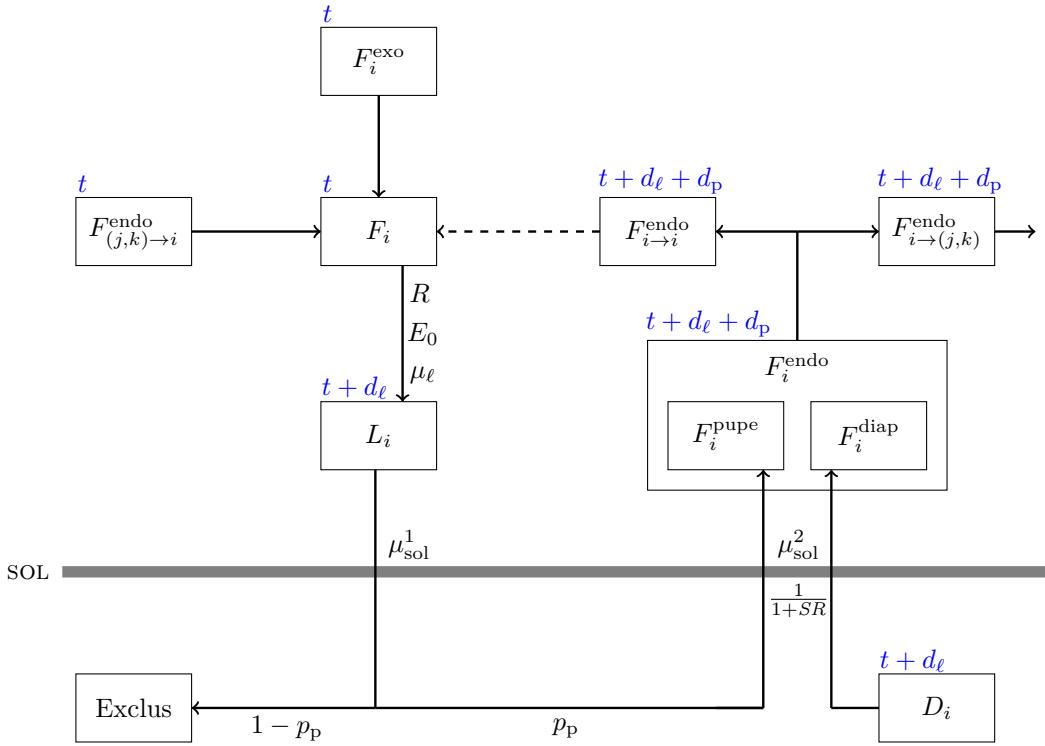


FIGURE 4.1 – Schéma conceptuel du modèle pour la sous-parcelle i . En bleu est visible la date, la flèche en pointillés marque une rupture du temps.

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 μ_ℓ la probabilité de survie des œufs et larves.

Le coefficient R représentant la disponibilité en ressources traduit le fait que lorsqu'il n'y a pas suffisamment d'inflorescences dans le verger par rapport au nombre de femelles, alors il y a une compétition entre cécidomyies pour ces ressources. Seulement une proportion des femelles pourra alors accéder aux ressources et pondre et/ou il y aura une plus forte mortalité des larves. Cette idée sera prise en compte comme suit :

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

R vaut 1 lorsque toutes les cécidomyies femelles peuvent pondre ; il est compris entre 0 et 1 lorsque le ratio entre nombre d'inflorescences et nombre de cécidomyie est insuffisant. Ce coefficient R dépend du paramètre k qui indique combien de femelles peut accueillir une inflorescence chaque jour sans générer de compétition entre les cécidomyies.

Dans l'équation 4.1, on peut noter que le nombre de larves au jour $t + d_\ell$ dépend du nombre de femelles présentes dans la sous-parcelle i au jour t . Ces femelles sont composées de femelles exogènes et de femelles endogènes. Parmi les femelles endogènes, il y a celles qui proviennent de la sous-parcelle i et qui y restent et celles qui proviennent des sous-parcelles j et k et qui viennent dans la sous-parcelle i . Le nombre de femelles dans la

sous-parcelle i au jour t peut s'écrire

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

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 . On rappelle également que le nombre de femelles exogènes est, d'après notre hypothèse, proportionnel au nombre d'inflorescences. Chose qui peut se formaliser par

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

Les échanges de femelles entre les différentes sous-parcelles est basé sur l'hypothèse qu'elles préfèrent rester dans la sous-parcelle de laquelle elles émergent plutôt que d'aller ailleurs. Ainsi, le nombre de femelles qui ont émergées dans la sous-parcelle i et qui vont dans la sous-parcelle j est donné par

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

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 δ traduit «l'effet distance» (autrement dit, les cécidomyies préfèrent se déplacer le moins possible) et $F_{t,i}^{\text{endo}}$ désigne le nombre de femelles qui ont émergées dans la sous-parcelle i le jour t . Concrètement si $p_m = 0$, il n'y a aucun échange. Et si $p_m = 1$, alors les femelles se répartissent proportionnellement aux nombres d'inflorescences présentes dans chaque sous-parcelle.

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

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

Le nombre de femelles issues de la phase de pupaison qui émergent à une certaine date correspond au nombre de larves qui ont réussies à pénétrer dans le sol, qui sont entrées en pupaison, se sont transformées en adultes, ont survécues puis ont émergées du sol — en prenant en compte la durée de développement et la proportion de femelles chez les cécidomyies. On le retranscrit comme ceci :

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

On a ici $F_{t,i}^{\text{pupée}}$ 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 rentrées dans le sol de la sous-parcelle i une durée de pupaison auparavant, μ_{sol} la probabilité pour une larve d'entrer dans le sol et pour une femelle d'en sortir (probabilité 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 pondant pas, ils ne sont pas considérés dans le modèle). Notons qu'ici

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

où μ_{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 μ_{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). On posera néanmoins, par souci de simplicité, $\mu_{\text{sol}}^1 = \mu_{\text{sol}}^2$.

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 . Le stock total de larves en diapause est donné par

$$\text{stock} = \sum_t D_t.$$

Le tableau 4.1 recense les paramètres du modèle, et indique ceux qui seront calibrés et ceux dont les valeurs seront prises dans la littérature.

TABLE 4.1 – Les différents paramètres du modèle.

Paramètre	Définition	Valeur
γ	Paramètre régulant l'arrivée des individus exogènes au verger	Calibration
p_m	Paramètre régulant l'intensité des échanges entre sous-parcelles	Calibration
μ_{sol}^1	Probabilité de survie des larves lorsqu'elles s'enfouissent dans le sol à la modalité de couverture du sol	Calibration
μ_{sol}^2	Probabilité de survie des femelles qui émergent du sol (de pupaison ou de diapause) selon la modalité de couverture du sol	Calibration
k	Paramètre quantifiant le nombre de femelles que peut accueillir une inflorescence chaque jour sans générer de compétition entre les cécidomyies	Calibration
stock	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]
μ_ℓ	Probabilité de survie des œufs et des larves	~0.04 [Amouroux, 2013]
<i>SR</i>	<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_ℓ	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

ON détaille dans ce chapitre la méthodologie utilisée pour 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.

Il faut cependant noter qu'il n'y a que 20 relevés effectifs, et qu'ils ne furent pas fait à intervalles très réguliers. Si l'on appliquait notre fonction de coût à chacun des jours de la période considérée, on attribuerait plus d'importance aux relevés qui ont eu un écart relativement important avec le relevé précédent. Pour pallier ce problème, on comparera uniquement la moyenne des estimations correspondant à un relevé avec l'observation correspondante. Notre propos est illustré sur la figure 5.1.

Pour définir la fonction de coût, 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 $t^n = m$.)

Si l'on note les observations y et les estimations \hat{y} , notre fonction de coût peut s'écrire

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

où

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

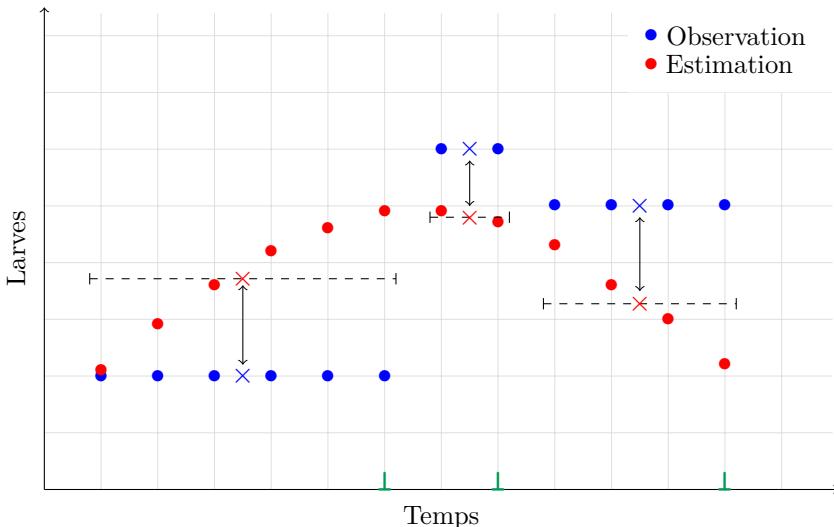


FIGURE 5.1 – Schéma illustrant le fonctionnement de la fonction objectif. À chaque relevé effectif (marqueurs verts), on fait correspondre la période correspondant à ce relevé (segments en pointillés). Et pour chacune de ces périodes, on calcule la moyenne des valeurs estimées (les croix rouges). On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

Concrètement \hat{y}_j^* donne la moyenne des estimations correspondant à un relevé et y_j^* donne la valeur observée associée. 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) permet de ne pas 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.

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.»¹

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 valeurs choisies pour les paramètres calibrés, 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

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.»

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 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 (ici les entrées X du modèle seront nos paramètres), c'est-à-dire $Y = f(X)$. Comme $X = (X_1, \dots, X_p)$ 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 $X_{\sim i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)$. L'indice S_i^T ainsi défini permet lui de prendre en compte l'impact du paramètre X_i et de ses interactions avec les autres paramètres 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, un algorithme d'optimisation est nécessaire.

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 qu'un test de exhaustif de toutes les valeurs de paramètres est trop coûteux. Une approche 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.

Une alternative est d'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 l'algorithme génétique NSGA-II (Nondominated Sorting Genetic Algorithm II) [Deb et al., 2002]. C'est celui que nous avons utilisé.

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. Si l'on note $P \subset \mathbf{R}^p$ ($p > 1$) l'ensemble des solutions, alors 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 «est préféré à» et «est strictement préféré à». 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 x 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, et les itérations seront appelées ici *générations*. C'est un algorithme convergent, plus il y a de générations, plus les solutions proposées sont proches du front de Pareto. En pratique, il ne renvoie donc pas un sous-ensemble du front de Pareto mais un ensemble de solutions se rapprochant d'un sous-ensemble du front de Pareto.

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 par la suite comment).

Avec la population P_t et les solutions filles associées O_t (*offspring* en anglais), cela forme un ensemble de solutions potentielles S_t de taille $2N$. On va alors sélectionner parmi ces solutions les N solutions les moins dominées possibles.

3. Pour ce faire, pour chaque solution potentielle $s \in S_t$:
 - on établit l'ensemble D_s d'éléments de S_t qui dominent s . On note n_s son cardinal ;
 - on établit l'ensemble E_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 $\bigcup_{s \in F^1} E_s$.) On continue de la même manière pour F^3 qui contient les solutions uniquement dominées par celles de F^2 (et *a fortiori* par F^1), 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 O_t effectué à chaque génération. Bon nombres d'algorithmes génétiques (dont NSGA-II) utilisent la *sélection*, le *crossover* et la *mutation* pour générer une solution fille. La sélection consiste à choisir certaines solutions mères qui seront les mieux à même de produire une bonne solution fille. Dans notre cas, les solutions du front F^1 seront choisies avec une plus grande probabilité que les autres. Le crossover consiste à créer une solution fille en mélangeant des coordonnées de deux solutions mères sélectionnées. La mutation effectue un changement aléatoire sur une des coordonnées, ce qui favorise la diversité des solutions.

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.

Il faut cependant noter que l'algorithme est stochastique, il ne renvoie jamais exactement deux fois les mêmes résultats. Pour pallier cet aspect, on exécutera trente fois la fonction `nsga2` [Mersmann, 2014] (avec une taille de population de 200, et 200 générations). Cela nous donnera ainsi 6000 solutions. Mais parmi ces 6000 solutions certaines seront peut-être dominées par d'autres solutions provenant d'une exécution de `nsga2` différente. On ne récupérera alors que les solutions non-dominées (et les jeux de paramètres correspondants) grâce à la fonction `is_dominated` [Mersmann, 2012].

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 jeux de paramètres possibles, les tester puis sélectionner celui qui semble le plus pertinent. On peut aussi sélectionner le jeu de paramètre qui minimise une norme sur nos trois critères, en estimant qu'il 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 certains de nos jeux de paramètres auront des valeurs proches. 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 jeux de paramètres centrés – réduits dans un espace métrique. On choisira \mathbf{R}^p muni de la distance euclidienne. On cherche à rassembler les différents jeux de paramètres en différentes classes. Les classes se doivent de contenir des individus aussi proches que possible (*i.e.* la distance 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.

On effectuera la CAH en utilisant la fonction `hclust` [R Core Team, 2018].

La pertinence biologique des classes de paramètres trouvées sera finalement validé par des experts.

6

Mise en œuvre et résultats

Maintenant que le modèle est posé et que l'on sait quelles méthodes utiliser pour la calibration et le choix des solutions, on peut calibrer le modèle et présenter et explorer les solutions optimales.

6.1 Solutions pour le modèle initial

On présente dans cette section des solutions pour le modèle initial, avant de tester d'autres hypothèses.

Analyse de sensibilité

La première étape à effectuer est l'analyse de sensibilité. Il faut cependant au préalable définir les intervalles dans lesquels évoluent nos paramètres. Si pour certains paramètres, les intervalles semblent naturels à choisir comme l'intervalle $[0; 1]$ pour les probabilités, le choix d'intervalles pour certains paramètres comme γ s'avère plus délicat. Procédons paramètre par paramètre.

En premier lieu, le paramètre γ qui, rappelons-le, correspond au coefficient de proportionnalité sur les inflorescences vivantes qui permet de déterminer le nombre de femelles exogènes arrivant dans chaque sous-parcelle à chaque date. Ce paramètre est propre à notre modèle, on ne peut donc pas trouver d'estimation dans la littérature. On peut fixer la borne inférieure de l'intervalle à 0, en partant du principe qu'une cécidomyie préfère rester dans le verger duquel elle émerge plutôt que de migrer dans un autre verger. On fixera la borne supérieure à 1. Cela signifie qu'il ne peut pas y avoir plus de femelles exogènes qui arrivent dans le verger que ce qu'il n'y a d'inflorescences.

Ensuite, par définition du paramètre p_m (qui régule les échanges de femelles entre les trois sous-parcelles), on sait qu'il évolue dans l'intervalle $[0; 1]$. De la même manière, μ_{ER} et μ_{EH} désignent les probabilités de survie aux modalités de couverture du sol ER et EH. Leurs intervalles seront donc aussi $[0; 1]$.

Le paramètre k relatif à la disponibilité en ressources gère le nombre d'attaques de cécidomyies que peut subir une inflorescence chaque jour sans qu'il n'y ait de compétition pour la ressource. En l'absence d'estimation ou de référence sur la valeur de ce paramètre, il sera fixé de manière assez large en 0.01 et 10.

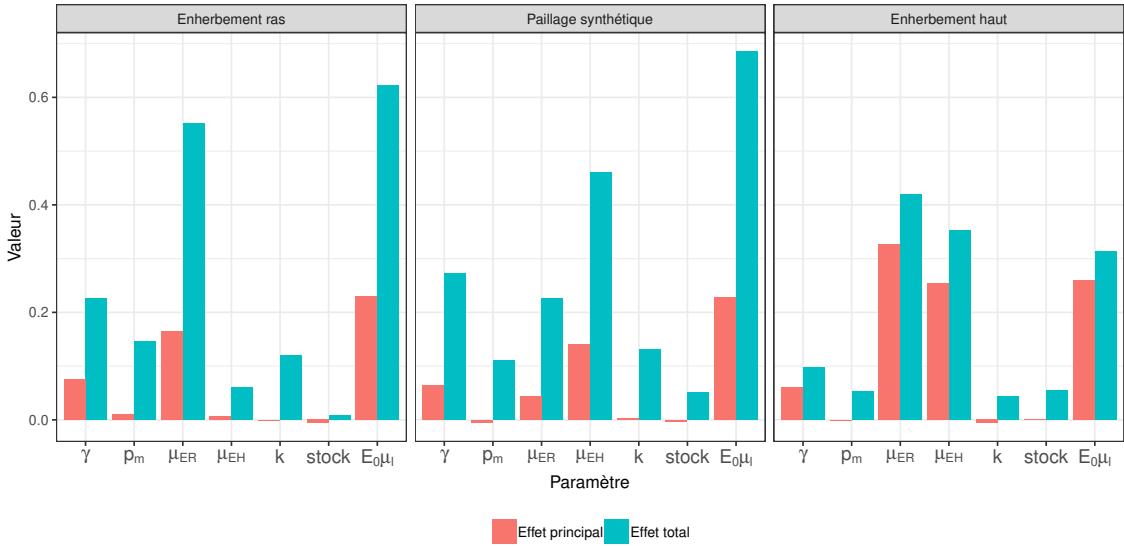


FIGURE 6.1 – Analyse de sensibilité de notre modèle avec la méthode Sobol.

Le **stock** d’individus en diapause sera fixé de manière assez large, dû à une absence d’estimation, entre 500 et 20000.

Enfin, le nombre d’œufs pondus qui survivent $E_0\mu_\ell$ est présent dans la littérature [Amouroux, 2013]. Sa valeur est de 6, mais nous paraît peu fiable. On le calibrera autour de cette valeur, entre 1 et 11.

Les intervalles étant définis, on peut maintenant faire l’analyse de sensibilité. Les résultats, pour un échantillon de taille $N = 50000$, sont visibles sur la figure 6.1.

On obtient ainsi les effets principaux et les effets totaux de Sobol pour chaque sous-parcelle.

Pour l’enherbement ras, le paramètre qui induit le plus de variance (effet principal) est $E_0\mu_\ell$ qui donne le nombre d’œufs arrivant au troisième stade de développement larvaire. Sans surprise le probabilité de survie à la modalité de couverture du sol ER apporte aussi beaucoup de variance à cette sous-parcelle. Ensuite, l’apport en variance du paramètre γ n’est pas négligeable non plus. En revanche, les quatre autres paramètres n’apportent en eux-même que peu de variance relativement aux trois autres nommés ci-dessus. On remarque également que les paramètres qui apportent le plus de variance sont aussi ceux dont les interactions induisent le plus de variance. (La variance apportée par les interactions d’un paramètre correspond à la différence entre l’effet total et l’effet principal).

Pour le paillage synthétique, c’est aussi $E_0\mu_\ell$ qui apporte le plus de variance. Vient ensuite μ_{EH} , γ et μ_{ER} . Les paramètres p_m , k et **stock** n’apportent pas de variance. Il est intéressant de noter que la probabilité de survie à la modalité de couverture du sol de la sous-parcelle EH a plus d’impact que celle de la sous-parcelle ER. Ici aussi, ce sont les paramètres qui apportent le plus de variance qui ont le plus d’interactions.

Pour l’enherbement haut, c’est la probabilité de survie à la modalité de couverture du sol ER μ_{ER} qui apporte le plus de variance, ce qui est assez peu intuitif. Il est suivi de près par deux autres paramètres : μ_{EH} et $E_0\mu_\ell$. La part de variance apportée par le paramètre γ n’est pas négligeable. Celle apportée par p_m , k et **stock** est négligeable. Et une fois encore, ce sont les paramètres qui apportent le plus de variance qui ont le plus d’interactions.

Certains de ces résultats ne sont pas surprenants. Il faut notamment se rappeler que

le modèle est évalué sur l'estimation du nombre de larves, et que $E_0\mu_\ell$ intervient dans l'équation du modèle donnant le nombre de larves. Ainsi, ce paramètre seul — à valeur entre 1 et 11 — peut facilement doubler le nombre de larves en fonction de sa valeur. Et induit donc naturellement une forte variance dans le modèle. Et ce paramètre n'est interprétable que si l'on considère la valeur des autres paramètres. Car un nombre élevé d'œufs pondus qui survivent peut être compensé par une faible probabilité de survie aux modalités de couverture de sol et une faible arrivée d'individus exogènes — et vice-versa. Ce qui peut expliquer ses fortes interactions. Cette analyse de sensibilité montre surtout l'impact des probabilités de survie aux modalités de couverture du sol en fonction de chaque sous-parcelles, qui ne respecte pas ce que l'on pourrait s'imaginer *a priori*. On retiendra également que trois de nos paramètres (p_m , k et **stock**) n'ont qu'un impact très limité — comparativement aux autres — sur la sortie du modèle.

Calibration et choix des solutions

En utilisant les mêmes intervalles pour les paramètres que ceux utilisés par l'analyse de sensibilité, on peut utiliser NSGA-II. Parmi les 6000 solutions générées par la calibration du modèle avec l'algorithme d'optimisation NSGA-II, on a obtenu 842 jeux de paramètres produisant des solutions non-dominées. Pour repérer par CAH les différentes solutions types parmi ces 842 solutions non-dominées, il faut dans un premier temps choisir un nombre de classes. À cette fin, on regarde l'inertie intra-classes en fonction du nombre de classes (voir figure 6.2). Dans une optique de classification «classique», choisir 2, 3, 6 ou 8 classes pourrait s'avérer pertinent. (Ces nombres de classes produisant une minimisation relativement importante de l'inertie intra-classes.) Nous préférons cependant ne pas risquer de passer à côté d'une catégorie de solution potentiellement intéressante. Nous choisirons 16 classes.

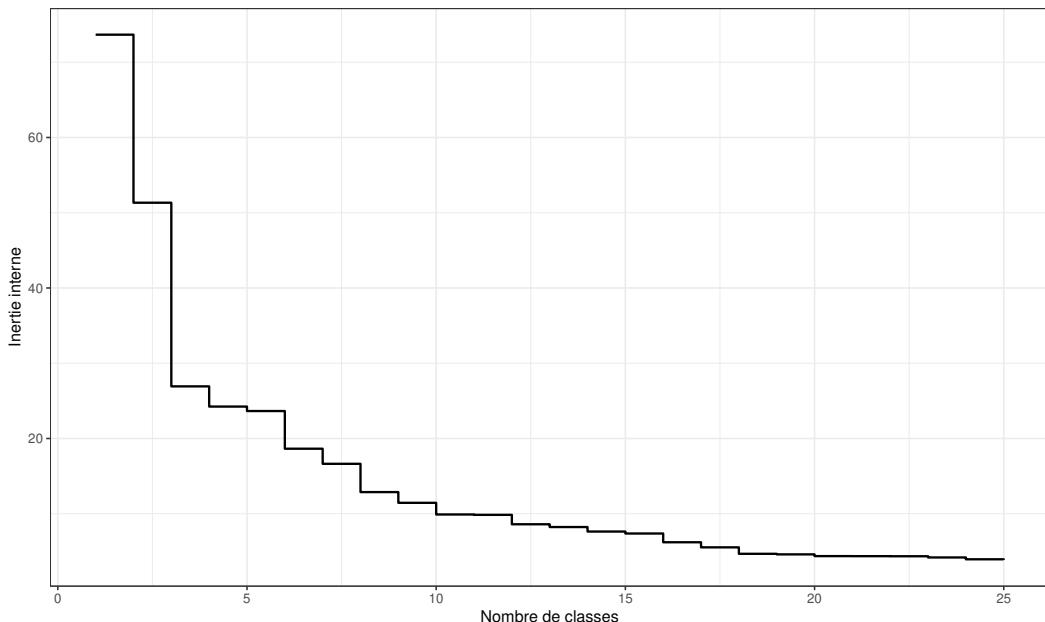


FIGURE 6.2 – Inertie intra-classes en fonction du nombres de classe obtenue par une CAH sur les jeux de paramètres renvoyés par NSGA-II.

Parmi les 16 classes, trois solutions-types se distinguent. Les solutions sont visibles

sur la figure 6.3. On peut voir sur la figure la décomposition du nombre de larves en fonction de la provenance des femelles qui ont pondus les œufs. Par exemple, si l'aire sous la courbe est essentiellement rouge, alors cela voudrait dire qu'il y a une forte proportion de femelles exogènes. Et si elle est principalement verte, cela veut dire que la majorité des femelles qui pondent sont des femelles qui ont émergées dans une autre sous-parcelle. Cela peut questionner la pertinence biologique de la solution proposée, indépendamment de la qualité d'ajustement de la dynamique.

On détaille les trois solutions-types :

- **Solution-type 1 :** Les dynamiques pour les trois sous-parcelles sont grossièrement captées. Le modèle permet de simuler la présence plus précoce de cécidomyies sur la sous-parcelle ER et plus tardive sur les sous-parcelles PS et EH. Il simule également un plus faible nombre de cécidomyies sur la sous-parcelle PS. Cependant le double pic caractérisant la dynamique des cécidomyies sur ER n'est pas capté, et la forte décroissance du nombre de cécidomyies en fin de saison est mal simulé.

Les paramètres associées à ces trois dynamiques sont :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0.08	0.494	0.976	0.046	1.928	5600	3.084

Pour cette solution-type, on observe une quantité non négligeable de larves pondus par des femelles exogène (aire rouge), ce qui d'un point de vue biologique paraît discutable. L'absence d'individus qui émergent de la sous-parcelle EH s'explique par μ_{EH} qui est fixé à 0.046, ce qui veut dire que 95% des larves meurent en essayant de s'enfouir dans la sous-parcelle ainsi que 95% des femelles issues de pupaison ou de diapause qui essayent d'en émerger. À la différence de la sous-parcelle ER où il n'y a pratiquement aucune mortalité induite par l'enherbement ras ($\mu_{ER} = 0.976$). La valeur 0.8 attribuée à γ est élevée (chose qui se confirmera empiriquement par les simulations qui suivront et qui peut déjà se voir sur la figure 6.3).

- **Solution-type 2 :** La qualité d'ajustement des dynamiques est assez similaire à celle de la première solution-type. Seule la décroissance en fin de saison est un peu mieux captée. Cependant, contrairement à la solution précédente qui avait trop d'individus exogènes, cette solution n'en a pas du tout. On remarque également que les dynamiques des sous-parcelles PS et EH se composent uniquement de larves provenant d'œufs pondus par des femelles issues de la sous-parcelle ER.

Les paramètres sont :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0	0.968	1	0.025	0.1	14483	6.26

Ici aussi, il y a une absence totale d'individus qui émergent de la sous-parcelle avec un enherbement haut ($\mu_{EH} = 0.025$). L'absence d'individus exogènes se compensent par un nombre d'œufs pondus plus élevé que précédemment (6.26 contre 3.08) et un stock de larves diapausantes significatif (stock = 14483).

- **Solution-type 3 :** Contrairement aux deux premières solutions, il y a ici des femelles qui émergent de la sous-parcelle EH. Il faut cependant reconnaître que les dynamiques sont très mal captées : les problèmes d'ajustement des dynamiques identifiés pour les deux précédentes solutions sont toujours présents, et la décroissance en fin de saison est encore moins bien captée.

Les paramètres sont :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0.026	0.93	0.565	0.647	0.174	13823	4.818

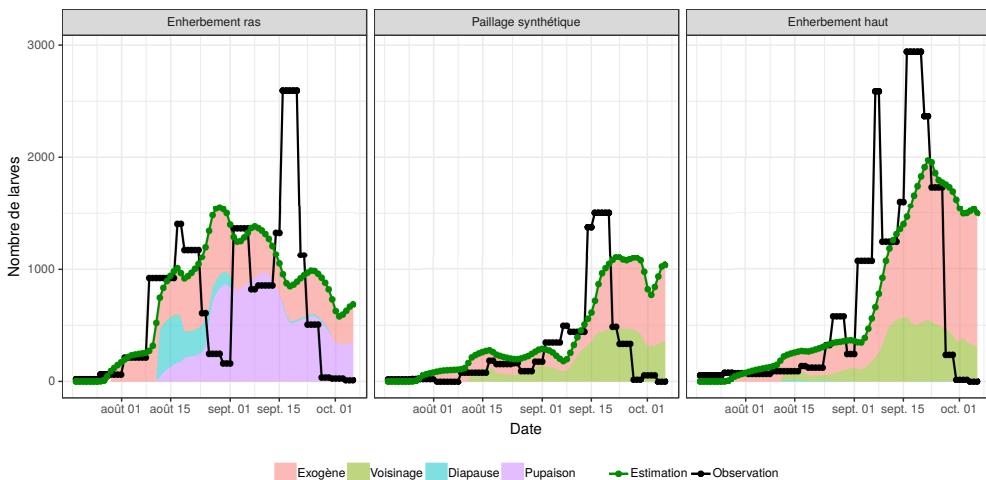
On observe que l'émergence d'individus dans la sous-parcelle EH ($\mu_{EH} = 0.647$ contre 0 précédemment) a entraîné une baisse d'émergence des individus dans la sous-parcelle ER ($\mu_{ER} = 0.565$ contre 0.97 et 1 pour les deux solutions précédentes)

Ces trois scénarios semblent tous montrer que le modèle actuel n'est pas suffisamment complet pour capter de façon satisfaisante le phénomène observé. Pour les deux premières solutions, il y a quasiment aucun individus qui émergent de la sous-parcelle EH, et les individus se composent surtout d'individus exogènes (première solution-type) ou d'individus endogènes venant de la sous-parcelle ER (deuxième solution-type). Et lorsqu'il y a des individus qui émergent de la sous-parcelle EH (troisième solution-type), le modèle ne parvient pas à recréer les dynamiques observées.

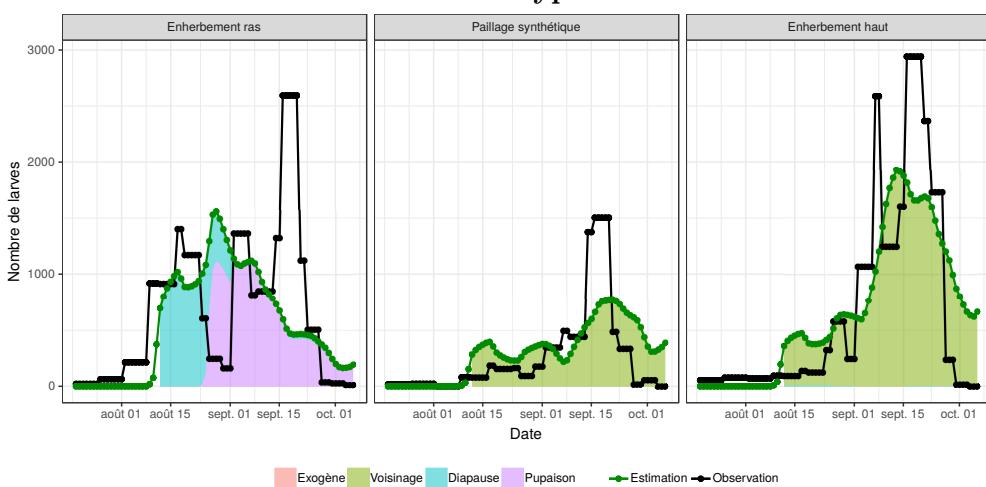
Cela peut s'expliquer par l'incapacité pour le modèle d'arrêter la reproduction des femelles et donc de stopper la dynamique en fin de saison induisant la baisse du nombre de larves observé. C'est peut-être l'une des raisons pour laquelle le modèle fait appel à beaucoup de femelles exogènes : elles sont proportionnelles aux inflorescences (dont la population baisse en fin de saison), ce qui entraîne donc une baisse du nombre de femelles et *a fortiori* une baisse du nombre de larves. Il y a dans le modèle un problème de décroissance en fin de saison.

Le modèle peut néanmoins être utilisé pour tester des hypothèses qui viendraient compléter ou remplacer les hypothèses initiales. Et si en testant une de ces hypothèses, on trouve une solution qui permet d'ajuster convenablement les dynamiques et qui ne soit pas aberrante d'un point de vue biologique, on pourra alors essayer de valider la solution sur le verger n°2.

Solution-type 1



Solution-type 2



Solution-type 3

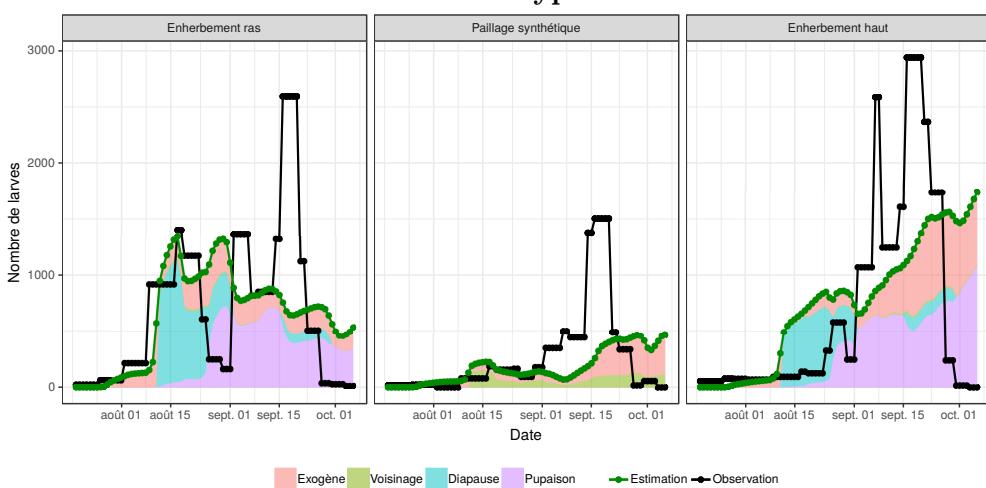


FIGURE 6.3 – Dynamiques observées et simulées pour chacune des trois solutions-types. La décomposition indiquant la provenance des femelles qui ont pondus les œufs est disponible pour les dynamiques simulées.

6.2 Solutions avec prise en compte de la température

La première hypothèse à laquelle on s'intéresse est le lien entre la température et la probabilité d'entrer en pupaison.

On sait, d'après [Amouroux et al. \[2014\]](#), que la probabilité d'entrer en diapause et d'entrer en pupaison varie tout au long de l'année, et dépend de la température. On espère ici qu'il y ait au cours de notre saison considérée (de juillet à octobre 2017) des variations de température pouvant expliquer une chute de la probabilité d'entrer en pupaison en fin de saison ce qui permettrait potentiellement au modèle de mieux reproduire la diminution du nombre de larves observées en fin de saison.

On choisira cependant la simplicité en essayant de trouver une relation linéaire permettant d'exprimer la probabilité d'entrer en pupaison en fonction de la température. Pour réaliser ceci, on récupère les données de l'article de [Amouroux et al. \[2014\]](#). Et à chaque taux de pupaison observé sur des larves prélevées en verger, on fait correspondre la température moyenne sur une quinzaine de jours (de 7 jours avant à 7 jours après la date du prélèvement). On fait cela pour prendre en compte l'effet la température sur le cycle de développement complet de la cécidomyie — depuis la ponte des œufs jusqu'à l'émergence des femelles issues de pupaison.

On effectue une régression linéaire simple de la probabilité d'entrer en pupaison en fonction de la température sur la quinzaine. On fixera le seuil du risque de première espèce à $\alpha = 5\%$ pour le test de non-nullité des coefficients. Les résultats sont visibles dans la table 6.1.

TABLE 6.1 – Régression linéaire simple de la proportion d'individus en pupaison par la température moyenne sur 15 jours

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.9555	0.3665	5.34	0.0002
temp15j	-0.0550	0.0160	-3.43	0.0050

On en conclut que les coefficients associés à l'ordonnée à l'origine et à la température ne sont pas nuls et que l'on peut écrire

$$p_p = 1.9555 - 0.055 \times t_{15j}.$$

La différence produite avec une probabilité constante fixée à 0.77, comme auparavant, est visible sur la figure 6.4. La différence en fin de saison est faible (environ 2%), cela ne permettra pas *a priori* d'améliorer le problème de la fin de saison. En revanche, il y a des différences plus importantes au début de la saison de floraison qui pourraient peut-être améliorer l'ajustement général des dynamiques.

Les données utilisées proviennent d'observations effectuées en condition de laboratoire. On laissera alors au modèle la possibilité d'amplifier la variation de la probabilité de pupaison afin de prendre en compte au mieux les conditions climatiques qu'il y aurait pu avoir sur le verger. On peut faire ceci en utilisant la relation linéaire de la probabilité de pupaison en fonction de la température (donné ci-dessus par p_p) et un coefficient $\varpi \in [1; 4]$ qui intervient comme suit :

$$\tilde{p}_p = (p_p - \bar{p}_p) \times \varpi + \bar{p}_p,$$

où \tilde{p}_p sera la probabilité d'entrer en pupaison et d'y survivre utilisée par le modèle et \bar{p}_p la moyenne de la probabilité de pupaison.

Après calibration du modèle et classification des solutions, on trouve trois solutions-types, qui sont très similaires à celles trouvées avec le modèle initial. Les voici :

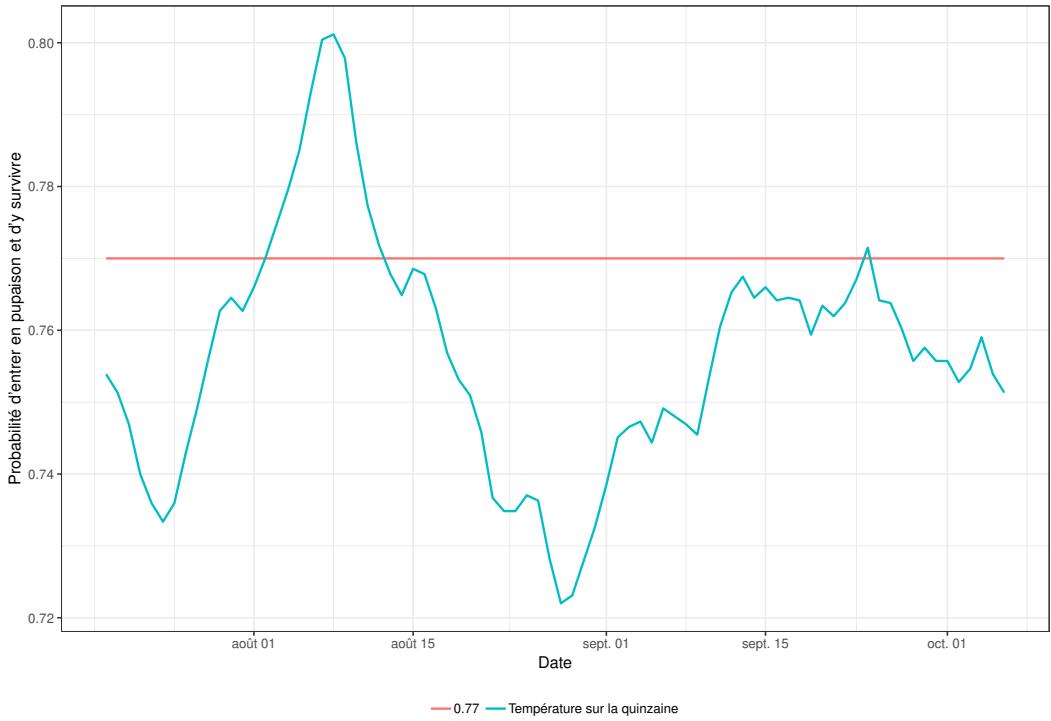


FIGURE 6.4 – Différence entre la probabilité d'entrer en pupaison et d'y survivre constante (égale à 0.77) et celle qui est fonction de la température moyenne sur la quinzaine.

- **Solution-type 1 :** Une solution faisant appel à beaucoup d'individus exogènes et sans individus qui émergent de la sous-parcelle EH.

Les paramètres sont :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$	ϖ
0.044	0.715	0.998	0.002	0.143	1548	4.202	1.167

- **Solution-type 2 :** Une solution sans individus exogènes ni sans individus qui émergent de la sous-parcelle EH et sans individus exogènes.

Les paramètres sont :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$	ϖ
0	1	1	0.001	0.1	17188	6.3	1.036

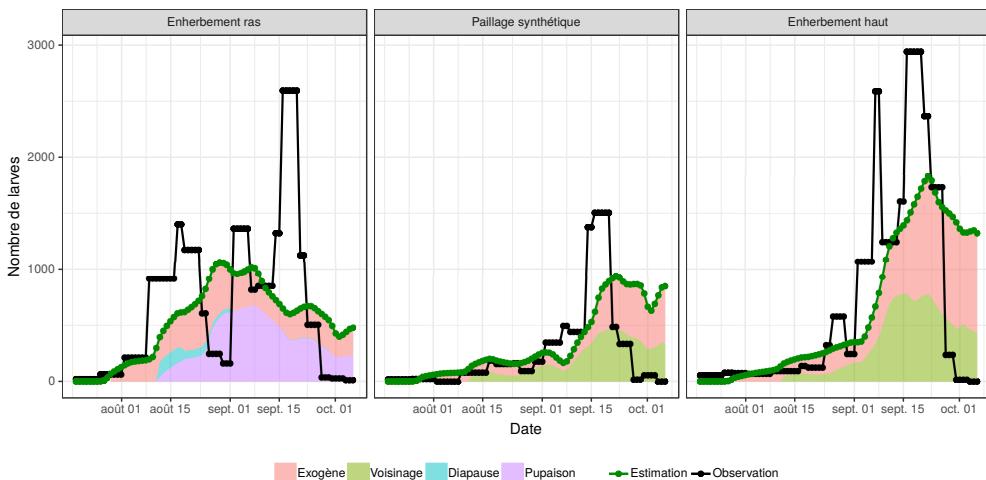
- **Solution-type 3 :** Une solution avec des femelles qui émergent de la sous-parcelle EH.

Les paramètres sont :

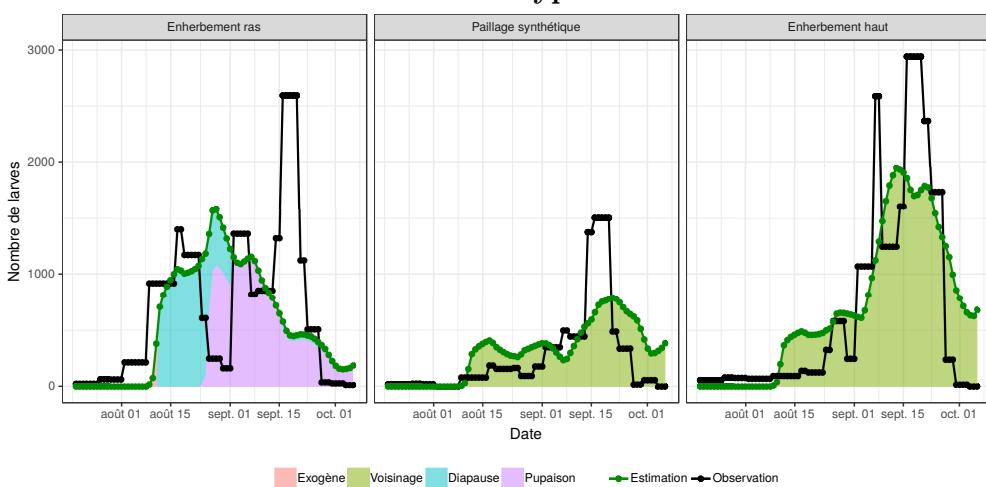
γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$	ϖ
0.002	0.280	0.923	0.986	0.177	20066	3.288	1.110

Il n'est probablement pas très pertinent de s'étendre longuement sur ces trois nouvelles solutions, dans la mesure où elles sont très similaires aux solutions trouvées avec le modèle initial. On peut remarquer que le paramètre ϖ ne dépasse jamais 1.5, ce qui semble indiquer que la prise en compte de l'effet de la température sur la probabilité d'entrer en pupaison ne semble pas améliorer le modèle.

Solution-type 1



Solution-type 2



Solution-type 3

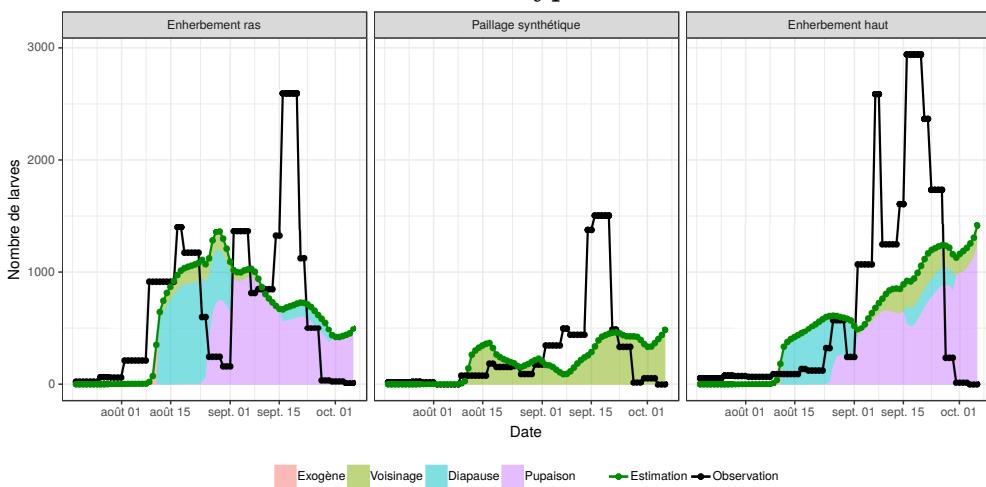


FIGURE 6.5 – Dynamiques observées et simulées pour chacune des trois solutions-types. La décomposition indiquant la provenance des femelles qui ont pondus les œufs est disponible pour les dynamiques simulées.

6.3 Solutions avec prise en compte de contraintes liées aux ressources

Puisque la température n'explique pas le phénomène observé en fin de saison, on teste une autre hypothèse. On s'intéresse aux ressources. L'intérêt pour les ressources est motivé par le fait que, sur les figures 3.3 et A.3, les dynamiques de larves semblent corrélées avec les dynamiques d'inflorescences vivantes. Les coefficients de corrélations de Spearman (du verger n°1) confirment plus ou moins cette intuition :

$$\rho^{\text{ER}}(L_t, I_t) = 0.46, \quad \rho^{\text{PS}}(L_t, I_t) = 0.68, \quad \rho^{\text{EH}}(L_t, I_t) = 0.76.$$

On pourrait s'attendre à un décalage entre les dynamiques de larves et d'inflorescences, correspondant à la durée de développement qu'il faut entre la ponte et l'apparition du troisième stade larvaire. Or, sur les figures 3.3 et A.3, cela ne s'observe pas.

Pour expliquer cette absence de décalage, on émet une hypothèse. Une inflorescence n'est pas attractive pour une cécidomyie de son débourrement jusqu'à la fin de sa vie mais uniquement lors des premiers stades phénologiques de l'inflorescence. Lorsqu'elle atteint le stade phénologique F, la tige se durcit progressivement, rendant peut-être plus difficile la ponte pour les cécidomyies. Notre hypothèse est que seules les inflorescences aux stades phénologiques C, D et E sont considérées comme attractives par les cécidomyies.

On détaille dans l'annexe B comment nous sommes capables de simuler des dynamiques d'inflorescences aux stades C, D et E.

Nous ferons ici l'hypothèse que la durée d'attractivité est de seize jours, ce qui correspond à la durée théorique des stades phénologiques C, D et E. La différence entre les inflorescences vivantes et les inflorescences attractives est visible sur la figure 6.6.

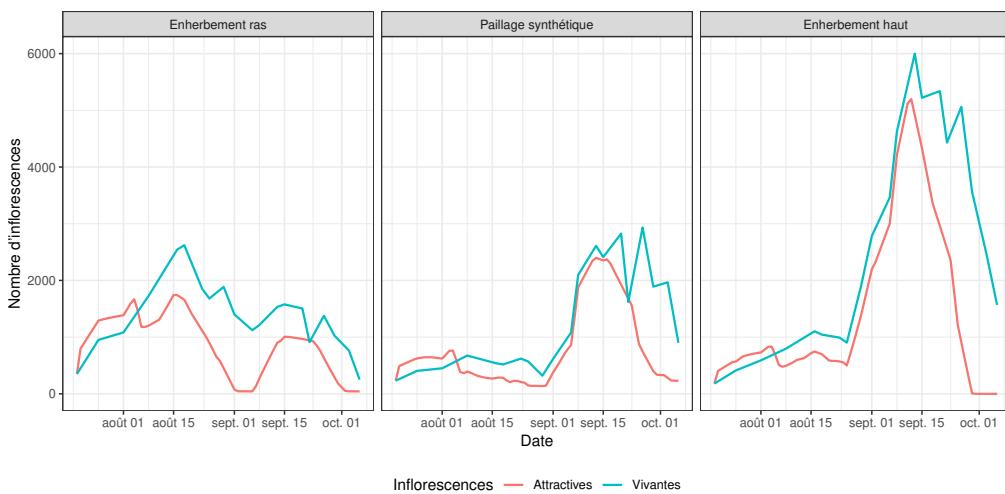


FIGURE 6.6 – Différence entre les inflorescences vivantes et les inflorescences attractives.

Sur les trois solutions-types trouvées après la calibration du modèle, deux correspondent à celles trouvées précédemment. Les dynamiques sont visibles sur la figure 6.7. On observe une amélioration sur les sous-parcelles PS et EH, notamment en ce qui concerne la fin de dynamique en fin de saison. La parcelle ER est en revanche toujours mal captée. Et la pertinence biologique des solutions (concernant la forte présence ou l'absence d'individus exogènes et l'absence d'individus qui émergent de la sous-parcelle EH) reste toujours discutable.

Les paramètres pour les trois solutions-types sont les suivants :

Solution-type 1

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0.115	0.189	0.898	0.012	1.877	515	3.097

Solution-type 2

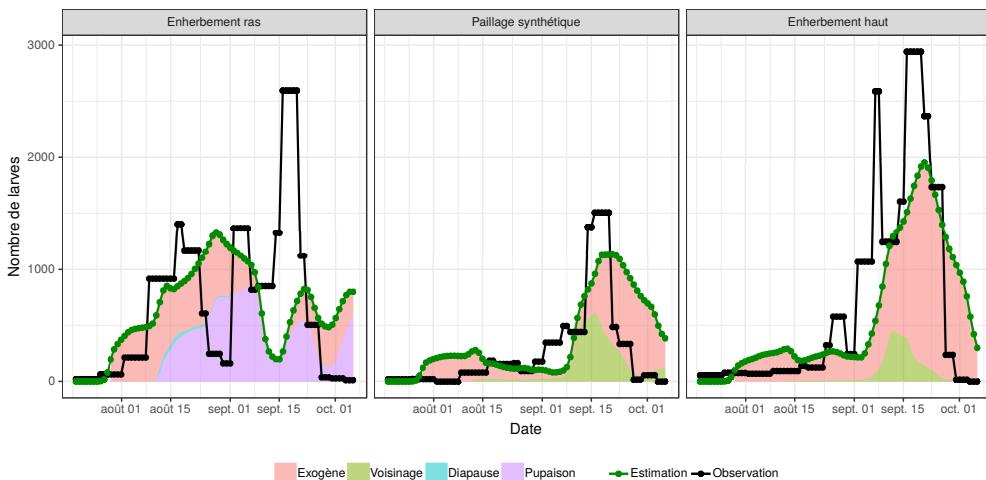
γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0.004	0.490	0.990	0.018	0.242	9970	5.638

Solution-type 3

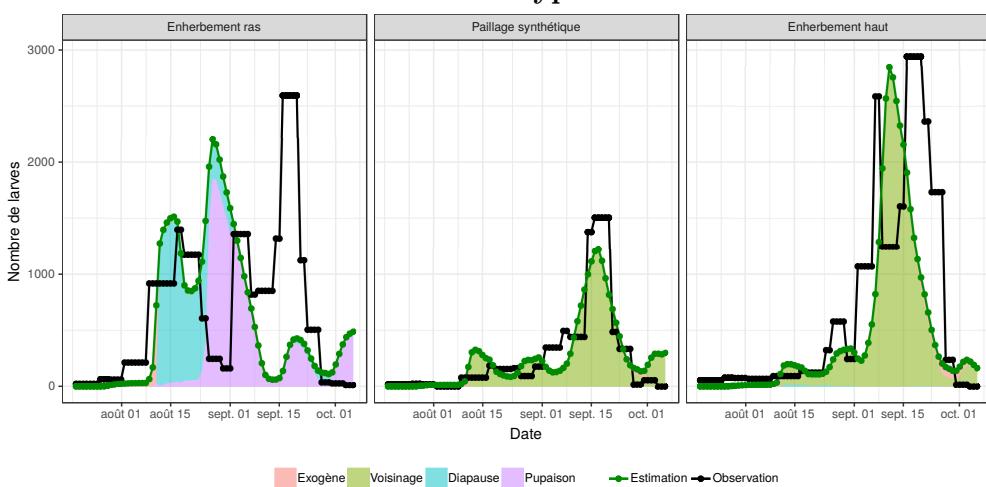
γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$
0	0.094	0.999	0.514	0.181	18214	8.259

On remarquera que la prédiction sur la troisième solution-type est toujours aussi mauvaise. La prise en compte de l'attractivité des inflorescences ne permet toujours pas la baisse des dynamiques observées en fin de saison lorsqu'il y a des individus endogènes dans la sous-parcelle EH.

Solution-type 1



Solution-type 2



Solution-type 3

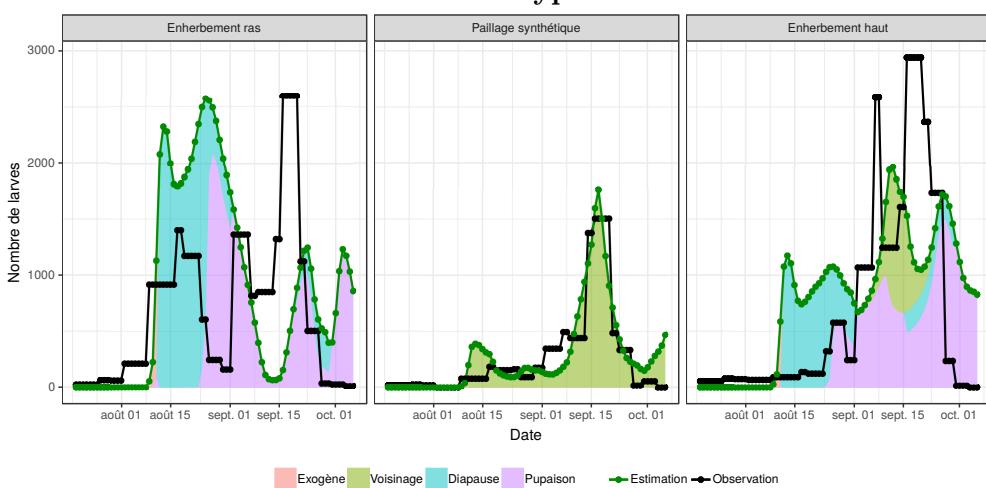


FIGURE 6.7 – Dynamiques observées et simulées pour chacune des trois solutions-types. La décomposition indiquant la provenance des femelles qui ont pondus les œufs est disponible pour les dynamiques simulées.

6.4 Solutions avec l'introduction d'un paramètre de saisonnalité

Les hypothèses émises jusqu'à présent n'ont pas permis d'expliquer ce qu'il se produit en fin de saison. Aux hypothèses déjà faites, on en rajoute une autre. Ce que l'on observe sur les dynamiques de larves, c'est qu'il semble y avoir une forte baisse de la population de larves qui s'amorce aux alentours du 15 septembre (sur le verger n°1). On émet alors l'hypothèse qu'un phénomène intervient à cette date. Ce phénomène non-observé caractériserait un changement de la conjoncture sur le verger et induirait une baisse du nombre de larves.

Et on modélise cette hypothèse de manière radicale en introduisant un paramètre de saisonnalité $\xi \in [0; 1]$ sur le nombre de femelles dans le verger à partir du 15 septembre. Cela donne en pratique

$$F_{t,i} := \begin{cases} F_{t,i} & \text{si } t < 15 \text{ septembre}, \\ \xi F_{t,i} & \text{si } t \geq 15 \text{ septembre}. \end{cases}$$

Concrètement, si le modèle décide de mettre une valeur éloignée de 1, cela signifie qu'à partir du 15 septembre que le cycle de développement de la cécidomyie subit des perturbations qui entraînent une baisse du nombre de larves produites.

Après calibration du modèle, un scénario se dégage des autres. Il est visible sur la figure 6.8. Les paramètres associés sont les suivants :

γ	p_m	μ_{ER}	μ_{EH}	k	stock	$E_0\mu_\ell$	ξ
0.021	0.105	0.938	0.916	1.992	516	6.018	0.004

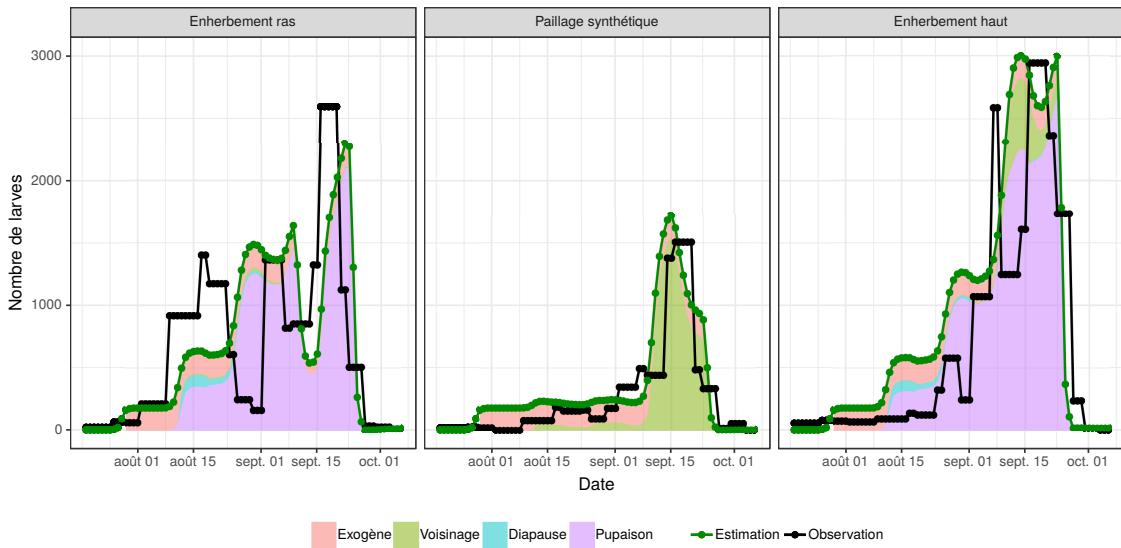


FIGURE 6.8 – Dynamiques observées et simulées. La décomposition indiquant la provenance des femelles qui ont pondus les œufs est disponible pour les dynamiques simulées.

Les dynamiques sont ici très bien captées pour les sous-parcelles PS et EH. La dynamique de la sous-parcelle ER est plutôt bien captée, surtout si l'on relativise vis à vis des estimations précédentes qui ont toujours été mauvaises. Les paramètres montrent

(et le graphique aussi) qu'il y a principalement des femelles endogènes, ce qui semble vraisemblable.

Cela semble être une bonne estimation. On procède à la validation sur le verger n°2. Il faut cependant noter que le phénomène de saisonnalité a lieu sur ce verger aux alentours du 23 août pour les modalités PS et EH et aux alentours du 7 septembre pour la modalité ER.

Les résultats produits sur le verger n°2 sont visibles sur la figure 6.9. On remarque que les dynamiques des modalités ER et EH sont plus ou moins captées. En revanche, si la temporalité de la dynamique de la sous-parcelle PS semble correcte, son intensité n'est pas restituée correctement.

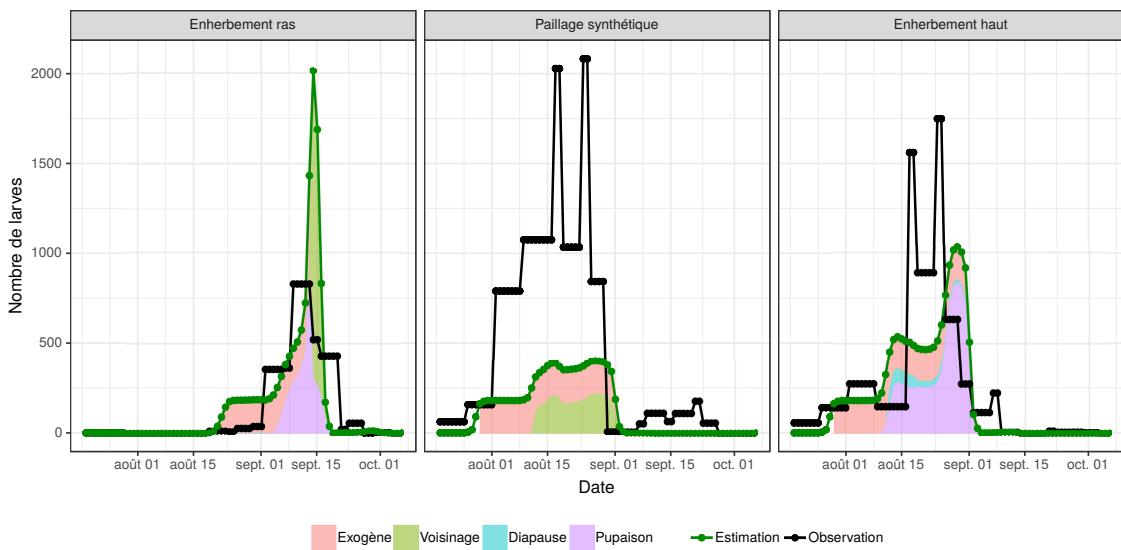


FIGURE 6.9 – Dynamiques observées et simulées sur le verger n°2. La décomposition indiquant la provenance des femelles qui ont pondus les œufs est disponible pour les dynamiques simulées. Les paramètres utilisés sont ceux calibrés par le modèle sur le verger n°1.

7

Discussion

ON revient dans ce chapitre sur les résultats obtenus, les hypothèses faites et les méthodes employées.

Commençons par l'aspect biologique. On peut faire quelques commentaires sur certaines hypothèses qui ont pu être faites. En premier lieu, le fait que la venue de femelles exogènes se fasse proportionnellement aux inflorescences est une hypothèse forte. Ce n'est pas très grave dans le cas des solutions-types qui privilégient les individus endogènes, où les femelles exogènes sont peu nombreuses. Lorsque les femelles exogènes contribuent fortement aux dynamiques de larves, il pourrait être pertinent de mener une expérimentation afin de confirmer ou d'infirmer cette hypothèse.

Concernant l'hypothèse d'attractivité des inflorescences, ne prendre en compte que les stades phénologiques C, D et E est peut-être un peu radical. En effet, si le début du stade F commence avec l'apparition de la première fleur cela ne veut pas dire que l'inflorescence devient subitement inattractive pour les cécidomyies. Si cette attractivité existe bien en réalité, alors ce serait probablement un phénomène plus progressif et le début de la période d'inattractivité interviendrait au milieu du stade F. Là encore, une expérimentation pourrait s'avérer utile pour déterminer à quel stade phénologique les inflorescences sont le plus attaquées.

La mise en place d'un paramètre de saisonnalité telle que nous l'avons faite revient, en pratique, à éradiquer toutes les femelles (ou les empêcher de pondre) du jour au lendemain. Cette prédiction du modèle est pour l'instant difficile à expliquer. On remarque toutefois sur nos observations (pour chacune des sous-parcelles des deux vergers) une chute brutale du nombre de larves en fin de saison. Ce qui laisse penser qu'il se passe effectivement quelque chose en fin de saison. Or, aucune connaissance présente dans la littérature n'explique l'absence de reproduction des cécidomyies autrement que par un changement conséquent de température ou une absence de ressources. Des expérimentations supplémentaires seront nécessaires pour identifier ce phénomène. On peut cependant émettre des hypothèses comme, par exemple, une compétition entre la cécidomyie et un autre ravageur qui s'attaquerait à des stades phénologiques plus avancés des inflorescences, et apparaîtraient par exemple qu'à partir du stade phénologique F.

Concernant le fait que les paramètres trouvés ne s'appliquent pas parfaitement aux dynamiques sur le verger n°2, cela peut s'expliquer par des conditions différentes entre les deux vergers. Le verger n°2 est en pente, et les modalités de couverture du sol n'ont pas la

même disposition. Notamment, la sous-parcelle PS se trouve sur le bord à proximité d'un autre verger, ce qui pourrait expliquer une plus forte présence d'individus exogènes.

On notera également qu'une expérimentation est en cours visant à quantifier le nombre d'individus qui sortent de diapause, ce qui permettra également de confirmer ou d'infirmer l'ordre de grandeur de la valeur trouvée pour le stock de larves en diapauses, et également sur l'effet de la couverture du sol sur le taux de survie des cécidomyies à l'entrée et la sortie du sol.

On peut également émettre quelques remarques concernant la méthodologie employée. Concernant l'algorithme d'optimisation, seul NSGA-II a été utilisé. Son utilisation a été concluante pour déterminer des solutions optimales des modèles, c'est pourquoi l'on n'en a pas essayé d'autre. On aurait probablement pu trouver un autre algorithme multicritères qui converge plus vite ou plus près du front de Pareto. Cependant la diversité des solutions proposées par NSGA-II est déjà grande. Et avoir des paramètres plus précis n'est peut-être pas pertinent dans la mesure où le phénomène modélisé est très variable et que l'on calibre le modèle sur des données qui ne sont pas elles-mêmes extrêmement précises.

On peut aussi remettre en question la classification effectuée. Et surtout le choix d'effectuer la classification sur les valeurs de paramètres. On part ici du principe que deux jeux de valeurs de paramètres dont les valeurs sont proches produiront des dynamiques similaires. Ce principe ne s'applique pas tous le temps, notamment dans les systèmes chaotiques, où des valeurs initiales très proches peuvent aboutir à des résultats très différents. Peut-être que dans notre modèle, il se produit — dans une certaine mesure — un phénomène similaire, avec notamment la possibilité d'un «effet seuil» chez certains paramètres, qui pourrait produire des dynamiques très différentes malgré des valeurs de paramètres proches.

8

Conclusion

Si l'on considère que l'objectif principal du stage était d'améliorer la compréhension du système manguier – cécidomyies, alors le stage est une réussite. Plus précisément, il y avait trois objectifs.

Le premier était de modéliser le système manguier – cécidomyies des fleurs. Le modèle existe et il est fonctionnel. Cet objectif est atteint.

Le deuxième objectif était d'analyser le fonctionnement dudit système. Les grandes composantes du système ont été identifiées et modélisées. On n'a cependant validé que partiellement les paramètres sur le verger n°2. Il est important de noter que le modèle fait une prédiction et semble mettre en évidence l'existence d'un phénomène en fin de saison, jusqu'alors non identifié, responsable de la diminution du nombre de cécidomyies — ce qui permet d'envisager de nouvelles pistes pour mieux comprendre le système.

Le troisième objectif était de réaliser des tests *in silico* afin d'évaluer le meilleur mode de gestion des vergers. Cet objectif était conditionné aux deux autres. Bien que la finalisation du modèle a pris du temps, la dernière version du modèle doit permettre de faire une telle analyse dans le futur.

À l'issue de ce stage, on peut envisager plusieurs pistes pour de futures améliorations. Par exemple, réaliser de nouvelles expérimentations sur le terrain pour acquérir plus de données sur l'attractivité des différents stades phénologiques des inflorescences. Et aussi s'intéresser à la possible compétition entre la cécidomyie des fleurs et d'autres ravageurs. Il faudra *a priori* acquérir plus de connaissances et/ou de données avant d'entreprendre à nouveau une démarche de modélisation.

D'un point de vue personnel, ce stage aura été une réussite. D'abord d'un point de vue technique. J'ai pu appliquer des concepts mathématiques appris en cours à un cas concret. J'ai aussi été amené à apprendre par moi-même des concepts non vus en cours relatifs à l'optimisation et l'analyse de sensibilité d'un modèle, et à les appliquer. Cela fut très intéressant.

Ensuite d'un point de vue professionnel, j'ai pu avoir une première expérience dans le domaine de la recherche. Expérience fort plaisante, qui m'a permis de me faire une idée plus précise du métier de chercheur. Et de rappeler certaines évidences : les résultats obtenus ne sont pas toujours ceux que l'on cherche initialement, et que cette incertitude est au cœur même du métier. Et c'est ce qui rend ce métier passionnant.

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, financé par la communauté européenne (fond structurel FEDER) et le Conseil Régional de La Réunion.



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. Jestic, 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.
- O. Mersmann. *emoa : Evolutionary Multiobjective Optimization Algorithms*, 2012. URL <https://CRAN.R-project.org/package=emoa>. R package version 0.5-0.
- O. Mersmann. *mco : Multiple Criteria Optimization Algorithms and Related Functions*, 2014. URL <https://CRAN.R-project.org/package=mco>. R package version 1.0-15.1.
- 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.
- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL <https://www.R-project.org/>.
- 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).

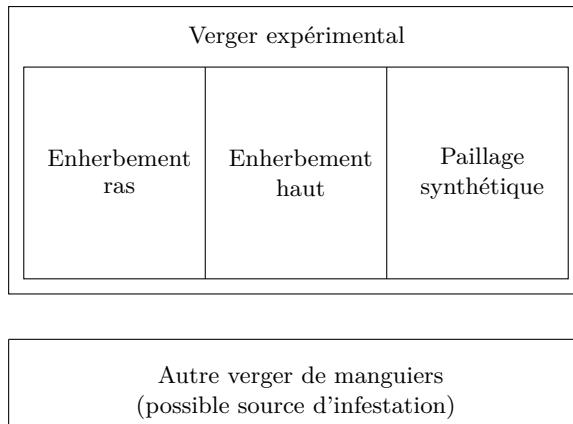


FIGURE A.1 – Schéma de l'expérimentation menée. Le verger sur lequel ont été testées les trois modalités de couvertures du sol était situé à côté d'un autre verger. Bien que cet autre verger n'avait pas de rôle particulier, il a probablement servi de source d'infestation exogène au verger expérimental.

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.

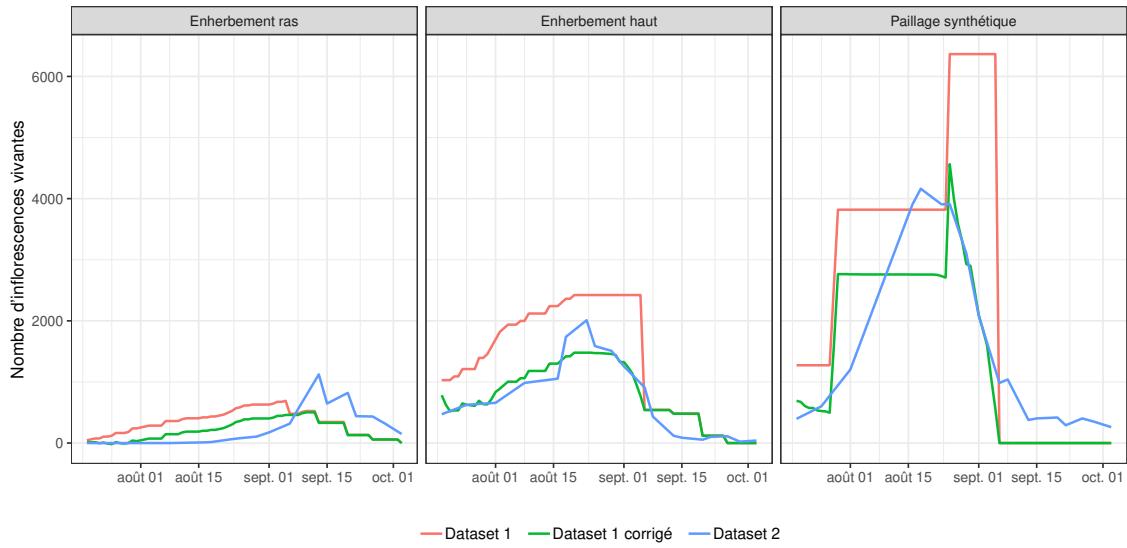


FIGURE A.2 – Comparaison des différentes dynamiques d’inflorescences vivantes du verger n°2 en fonction du *dataset* utilisé. Même après correction, on observe des différences entre les dynamiques issues des différents jeu de données, en particulier pour les deux premières modalités.

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.

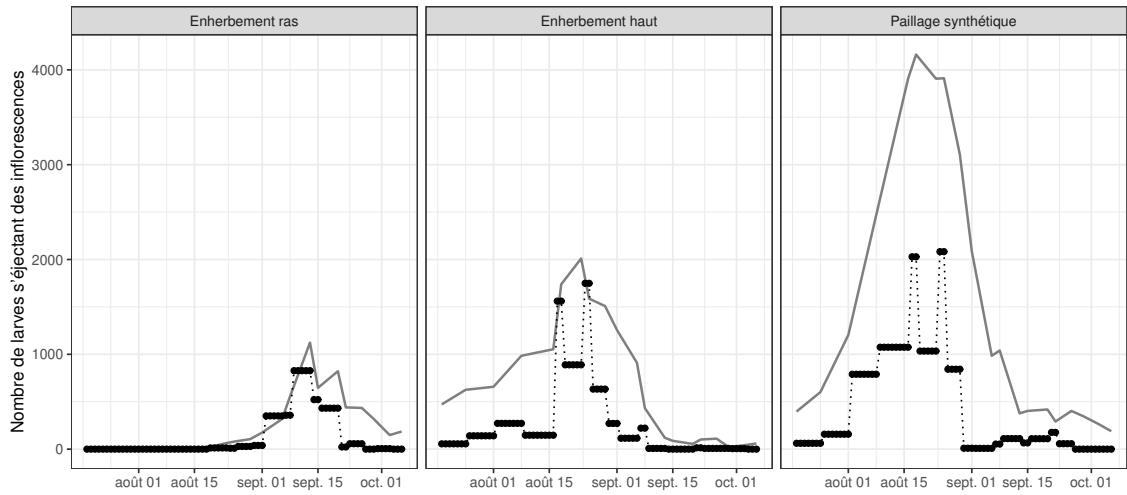


FIGURE A.3 – Dynamiques de larves s’éjectant des inflorescences des manguiers chaque jour dans le verger n°2 pour chacune des trois sous-parcelles. En gris sont visibles les dynamiques d’inflorescences vivantes (issues du *dataset* 2, I_t^2).

B

Simulation d’inflorescences attractives

On détaille dans cette annexe comment simuler des dynamiques d’inflorescences attractives. Pour ce faire, nous avons besoin des dates de débourrements des inflorescences ainsi que de la loi de mortalité des inflorescences.

Le *dataset 1* contient des dates de débourrements. On ne pourra cependant pas les utiliser. Dans le verger n°1, les dynamiques pour les modalités «enherbement ras» et «paillage synthétique» issues des deux jeux de données ont des dynamiques plutôt similaires ; on pourrait donc utiliser les débourrements du *dataset 1* mis à l’échelle. En revanche, pour la modalité «enherbement haut», on observe des dynamiques très différentes. Et comme l’on souhaite privilégier les dynamiques issues du *dataset 2* (qui sont associées aux dynamiques de larves), il faut procéder autrement. (Il en va de même pour les modalités «enherbement ras» et «paillage synthétique» du verger n°2.)

L’objectif est alors de simuler les dates de débourrements qui permettent de produire la dynamique d’inflorescences vivantes du *dataset 2*. Pour cela, on suppose que la durée de vie effective d’une inflorescence suit une loi normale. On possède les durées de vies effectives dans le *dataset 1*. Premièrement, en fixant le risque de première espèce à $\alpha = 5\%$, un test de normalité de Shapiro–Wilk nous confirme la normalité (*p*-valeur de 0.05447, on ne rejette donc pas l’hypothèse de normalité). Deuxièmement, on observe une durée de vie effective moyenne de 29 jours (avec un écart-type de 14). Ainsi, il faut donc simuler des débourrements, pour que des inflorescences ayant une durée de vie qui suit une $\mathcal{N}(29; 14)$, donne la dynamique d’inflorescences vivantes du *dataset 2*. Il faut donc trouver les B_t tels que

$$I_t^2 = B_t + \sum_{j=1}^{50} B_{t-j} \times (1 - F(j)), \quad \text{avec } B_t = 0 \text{ si } t \leq 0,$$

où F est la fonction de répartition d’une $\mathcal{N}(29; 14)$ et 50 la durée de vie théorique d’une inflorescence.

Par souci d’homogénéité, on utilisera les débourrements simulés pour les trois modalités, et ce pour les deux vergers.

Les inflorescences *attractives* peuvent alors se calculer en utilisant la formule

$$I_t^a = B_t + \sum_{t=1}^{d_A} B_{t-j} \times (1 - F(j)), \quad \text{avec } B_t = 0 \text{ si } t \leq 0.$$

Ici, B_t représente le nombre de débourvements à la date t , d_A la durée d'attractivité voulue et F la fonction de répartition d'une $\mathcal{N}(29; 14)$.

Dans notre cas, on choisira $d_A = 16$ jours, correspondant à la durée des stades C, D et E.