



**HAL**  
open science

# Analyse de données spatialisées issues de la production pour améliorer le diagnostic agronomique en bananeraie - Prise en compte de l'asynchronisme de la culture

Julien Lamour

► **To cite this version:**

Julien Lamour. Analyse de données spatialisées issues de la production pour améliorer le diagnostic agronomique en bananeraie - Prise en compte de l'asynchronisme de la culture. Sciences agricoles. Montpellier SupAgro, 2019. Français. NNT : 2019NSAM0017 . tel-02879774

**HAL Id: tel-02879774**

**<https://theses.hal.science/tel-02879774>**

Submitted on 24 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Génie des procédés

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau  
Portée par l'Université de Montpellier

Unité de recherche ITAP – Information, Technologies,  
Analyse environnementale, Procédés agricoles

Analyse de données spatialisées issues de la production  
pour améliorer le diagnostic agronomique en bananeraie -  
Prise en compte de l'asynchronisme de la culture

Présentée par Julien LAMOUR

Le 12/07/2019

Sous la direction de Bruno TISSEYRE

Devant le jury composé de

Louis Longchamps, Professeur, Agriculture and Agri-Food Canada, Canada

Jean Roger-Estrade, AgroParisTech, France

Alexandra Jullien, Professeur, Inra, France

Philippe Tixier, Chercheur, Cirad, France

Bruno Tisseyre, Professeur, Montpellier Supagro, France

Olivier Naud, Ingénieur de recherche, Irstea, France

Mathieu Léchaudel, Chercheur, Cirad, France

Gilles Le Moguédec, Chercheur, Inra, France

Alain Normand, Directeur de la R&D et de l'agronomie, Compagnie Fruitière, France

Rapporteur

Rapporteur & Président du jury

Examinatrice

Examineur

Directeur de thèse

Examineur, Encadrant

Invité, Encadrant

Invité, Encadrant

Invité, Encadrant



UNIVERSITÉ  
DE MONTPELLIER

Montpellier  
SupAgro



# Introduction générale

L'objectif de cette thèse est de proposer des approches nouvelles visant à faciliter le diagnostic agronomique des bananeraies à partir de l'étude des données produites par les producteurs. Cet objectif est celui de l'Agriculture de Précision, discipline scientifique qui a principalement été développée sur les grandes cultures. La banane est une culture atypique dans cette communauté avec un cycle de développement différent qui remet en cause certains concepts et outils.

En agriculture de précision, une importance toute particulière a été donnée à la dimension spatiale des observations. La production d'une parcelle est étudiée comme le résultat d'une expérience scientifique d'homogénéité (Uniformity Trial). Les conditions de croissance, dont on cherche à voir si elles sont variables dans la parcelle, sont mises en évidence par la comparaison spatialisée des caractéristiques de la production (quantité, qualité des récoltes ou encore propriétés agronomiques du couvert : biomasse, maturité, etc.). Le résultat de cette expérience permet de générer des connaissances locales sur le système de production que les agriculteurs peuvent exploiter pour améliorer leurs pratiques.

La production bananière n'est pas synchrone comme la plupart des grandes cultures. Le cycle de développement des bananiers n'est pas saisonnier et n'est pas synchronisé par le mode de culture industriel. En conséquence, les parcelles sont constituées de plants à tous les stades phénologiques et le stade phénologique moyen des parcelles d'une plantation est variable. Il en résulte que les propriétés végétatives que l'on peut mesurer sur le couvert d'une parcelle ne dépendent pas seulement des conditions de croissance mais aussi de ces stades.

La question à laquelle je cherche à répondre dans cette thèse est donc : « Comment valoriser les données spatialisées des producteurs afin d'extraire des connaissances sur l'agro-système bananier malgré l'asynchronisme » ?

Nous avons fait l'hypothèse que les données produites lors de l'exploitation des bananeraies, qui couvrent de grandes surfaces (plusieurs milliers d'hectares) sur de longues périodes de temps (environ 20 ans) et qui sont obtenues dans des conditions pédo-climatiques variées peuvent être utilisées pour générer des connaissances sur le fonctionnement du bananier et en améliorer la gestion. Outre les difficultés liées à l'asynchronisme, l'exploitation de données de production met en évidence des problématiques propres, relatives à la qualité des données générées dont la fiabilité est difficile à estimer, à des informations manquantes, à l'abondance de données et à l'absence de plan d'expérience.

Cette thèse a été réalisée en contrat CIFRE (Convention Industrielle de Formation par la Recherche) avec la Compagnie Fruitière et les données que j'ai utilisées proviennent principalement d'une de ses plantations appelée « Plantation du Haut Penja » au Cameroun.

Le premier chapitre de ce mémoire décrit synthétiquement le fonctionnement du système de production bananier, ses particularités et ses enjeux. Le contexte qui a poussé la Compagnie Fruitière à engager des travaux de recherche est aussi présenté ainsi que les enjeux opérationnels de la thèse. Le second chapitre décrit l'Agriculture de Précision en tant que discipline scientifique et en tant qu'approche agricole. Les objectifs et concepts généraux de cette discipline sont présentés ainsi que les quelques articles qui concernent spécifiquement la production bananière. L'énoncé de la problématique scientifique conclut ce chapitre avec l'ensemble des questions auxquelles j'ai voulu répondre dans ce mémoire.

Le chapitre 3 décrit l'asynchronisme et propose des méthodes objectives pour le mesurer dans les

---

plantations, à l'échelle inter-parcellaire et intra-parcellaire. Ce chapitre s'appuie en grande partie sur la modélisation du nombre de floraisons qui est observé par les ouvriers dans les parcelles. Ce modèle a été utilisé pour étudier la variabilité de la durée des cycles de floraison qui est à l'origine de l'asynchronisme, à partir des bases de données de la Compagnie Fruitière. Des indices qui permettent de mesurer l'asynchronisme et ses conséquences dans les parcelles ont aussi été formulés à partir des équations du modèle.

Le chapitre 4 décrit une méthode qui permet de diminuer les biais liés à l'asynchronisme lors de l'étude des données de production. L'intérêt de cette méthode est évalué sur un indicateur particulier mesuré en production : le temps qui sépare la floraison et le moment où le régime doit être récolté. Le chapitre 5 évalue la faisabilité d'utiliser la télédétection pour faire un diagnostic de la vigueur des parcelles malgré la diversité des stades phénologiques. Enfin le chapitre 6 présente le bilan des travaux effectués. Nous présentons aussi différentes perspectives de recherches qui permettraient de poursuivre ces travaux.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Le bananier et sa culture . . . . .	8
1.1.1	Caractéristiques biologiques du bananier . . . . .	8
	Classification biologique . . . . .	8
	Description de la souche de bananier ou tige souterraine . . . . .	9
	Description du système racinaire . . . . .	9
	Description du système foliaire . . . . .	9
	Description de l'inflorescence . . . . .	9
1.1.2	Description du système de production industrielle des bananes d'exportation . . .	11
	Caractéristiques générales du système de production . . . . .	11
	Modes de multiplication . . . . .	11
	Conduite des parcelles . . . . .	12
	Gestion des maladies du bananier . . . . .	12
	Gestion de la nutrition hydrique et minérale . . . . .	13
	L'asynchronisme du développement des bananiers au sein des parcelles . . . . .	14
	Gestion des récoltes . . . . .	14
	Conditionnement et transport . . . . .	15
	Système d'information de la production . . . . .	16
1.1.3	La production bananière, une production importante au niveau mondial et soumise à de nouveaux enjeux . . . . .	17
	Caractéristiques de la production mondiale . . . . .	17
	Organisation et mutation du marché mondial . . . . .	17
	Des risques importants liés à l'apparition et à la propagation de nouvelles maladies	18
1.2	Contexte de la thèse et objectifs opérationnels . . . . .	19
1.2.1	Les objectifs de R&D de la Compagnie Fruitière . . . . .	19
	Présentation de la Compagnie Fruitière . . . . .	19
	Objectifs de R&D de la Compagnie Fruitière . . . . .	19
	Moyens mis en œuvre . . . . .	20
1.2.2	Objectifs opérationnels de la thèse . . . . .	20
<b>2</b>	<b>Les difficultés liées à l'asynchronisme pour adopter une démarche d'agriculture de précision</b>	<b>21</b>
2.1	État de l'art général sur l'agriculture de précision . . . . .	22
2.1.1	Principes et définition de l'agriculture de précision . . . . .	22
2.1.2	Démarche et exemples en agriculture de précision . . . . .	22
2.1.3	Méthodes et disciplines impliquées dans l'agriculture de précision . . . . .	23
2.1.4	Conditions d'application d'une agriculture de précision . . . . .	24
2.2	Les structures spatiales en agriculture de précision . . . . .	24
2.2.1	Origine de la variabilité spatiale des cultures . . . . .	24

2.2.2	Stabilité temporelle de la variabilité spatiale . . . . .	26
2.2.3	Étude et modélisation des structures spatiales . . . . .	26
	Approche générale . . . . .	26
	Théorie des variables aléatoires régionalisées . . . . .	27
2.3	État de l'art de l'agriculture de précision sur la banane . . . . .	29
2.3.1	Bibliographie . . . . .	29
2.3.2	Particularités de la culture bananière par rapport aux cultures sur lesquelles s'est développée l'agriculture de précision . . . . .	30
2.3.3	Les difficultés liées à l'asynchronisme pour l'étude de la variabilité spatiale . . . . .	31
2.4	Objectifs scientifiques de la thèse . . . . .	32
<b>3</b>	<b>Caractérisation de l'asynchronisme avec les données de production</b>	<b>35</b>
3.1	Variabilité de la durée moyenne des cycles de développement des parcelles à l'échelle d'une plantation . . . . .	37
3.1.1	Résumé . . . . .	37
3.1.2	Introduction . . . . .	37
3.1.3	Méthodes . . . . .	39
	Présentation du modèle . . . . .	39
	Description du site d'étude et des données utilisées . . . . .	42
3.1.4	Paramétrisation et qualité d'ajustement du modèle . . . . .	43
	Analyse des sources de variabilité du . . . . .	46
	Logiciel utilisé . . . . .	47
3.1.5	Résultats . . . . .	47
	Description et qualification de la qualité des données de floraison . . . . .	47
	Présentation du modèle et de son ajustement sur les données de floraison . . . . .	48
	Analyse de la structure spatiale de . . . . .	48
	Analyse de la variabilité de . . . . .	53
3.1.6	Discussion . . . . .	53
	Représentation probabiliste d'une courbe de floraison . . . . .	53
	Sources de variabilité du . . . . .	55
3.1.7	Conclusion . . . . .	56
3.2	Variabilité de la durée moyenne des cycles de développement à l'échelle intra-parcellaire . . . . .	57
3.2.1	Introduction . . . . .	57
3.2.2	Matériels et Méthodes . . . . .	58
	Parcelle d'étude . . . . .	58
	Enregistrement des dates de floraison . . . . .	58
	Prétraitement des données . . . . .	58
	Analyses statistiques . . . . .	58
	Cartographie de la variabilité de CD . . . . .	60
	Logiciel utilisé . . . . .	60
3.2.3	Résultats et discussion . . . . .	60
	Présentation de la courbe de floraison de la parcelle . . . . .	60
	Estimation de dans chaque maille de la grille . . . . .	63
	Analyse variographique . . . . .	63
	Cartographie de . . . . .	63
	Validité de l'estimation des . . . . .	63
3.2.4	Conclusion . . . . .	66
3.3	Formalisation de l'asynchronisme . . . . .	66
3.3.1	Introduction . . . . .	66
3.3.2	Définition des indices . . . . .	66
	Indice d'hétérogénéité des conditions de croissance . . . . .	66

	Indice de stade phénologique moyen et de diversité des stades phénologiques . . . .	67
3.3.3	Matériel et Méthodes . . . . .	67
	Données . . . . .	67
	Choix des CF et calcul des indices . . . . .	69
3.3.4	Résultats et discussion . . . . .	69
	Description des CF choisies . . . . .	69
	Représentation du comportement des indices d'asynchronicité . . . . .	70
	Conclusion . . . . .	71
3.4	Évolution de la structure spatiale des stades phénologiques au cours du temps . . . . .	71
3.4.1	Introduction . . . . .	71
3.4.2	Analyse de la structure spatiale de données circulaires . . . . .	74
3.4.3	Matériel et méthodes . . . . .	74
3.4.4	Résultats . . . . .	75
3.4.5	Discussion et conclusion . . . . .	75
<b>4</b>	<b>Proposition d'une méthode de correction des biais liés à l'asynchronisme sur les données de récolte</b>	<b>77</b>
4.1	Abstract . . . . .	79
4.2	Introduction . . . . .	79
4.3	Methodological aspects . . . . .	81
4.3.1	Theory . . . . .	81
4.3.2	Description of the studied plant response : Time to reach Fruit Maturity ( <i>t<sub>fm</sub></i> ) . .	82
4.3.3	Modeling of <i>t<sub>fm</sub></i> variability and discretization of the temporal component using cohorts . . . . .	82
4.3.4	Decomposition of the <i>t<sub>fm</sub></i> variability . . . . .	83
4.4	Material and Methods . . . . .	84
4.4.1	Site study and data collection process . . . . .	84
4.4.2	Pre-processing of the data . . . . .	86
4.4.3	Decomposition of the variance . . . . .	86
4.4.4	Spatial analysis . . . . .	87
4.4.5	Data mapping . . . . .	87
4.4.6	Comparison of maps to assess the variance decomposition approach . . . . .	88
4.4.7	Software used . . . . .	88
4.5	Results and Discussion . . . . .	88
4.5.1	General characteristics of the data . . . . .	88
4.5.2	Evaluation of the cohort effect . . . . .	89
4.5.3	Decomposition of the spatial structure . . . . .	91
4.5.4	Effect of the cohorts and of environmental factors on the maps of <i>t<sub>fm</sub></i> . . . . .	93
4.5.5	Assessment of the variance decomposition approach . . . . .	93
4.5.6	Interpretation of . . . . .	95
4.5.7	Interpretation of maps . . . . .	96
4.6	Conclusion . . . . .	97
4.7	Acknowledgements . . . . .	98
<b>5</b>	<b>Étude de faisabilité du diagnostic agronomique des bananeraies par proxydétection et télédétection</b>	<b>99</b>
5.1	Mise en place du plan d'expérience . . . . .	101
5.1.1	Abstract . . . . .	101
5.1.2	Value of the data . . . . .	101
5.1.3	Data . . . . .	101
	Background . . . . .	101



Dataset description . . . . .	102
Experimental design . . . . .	102
Leaf data acquisition . . . . .	105
Supplementary data of the validation dataset . . . . .	107
5.1.4 Acknowledgements . . . . .	108
5.2 Analyse de plusieurs sources de variabilité du taux de chlorophylle à l'échelle du bananier	108
5.2.1 Abstract . . . . .	108
5.2.2 Introduction . . . . .	108
5.2.3 Material and methods . . . . .	109
Definitions and notations . . . . .	109
Study site and experiments . . . . .	109
Data analysis . . . . .	111
5.2.4 Resultats and discussion . . . . .	112
Analysis of model 1 . . . . .	112
Analysis of model 2 . . . . .	112
5.2.5 Conclusion . . . . .	114
5.2.6 Acknowledgments . . . . .	115
5.3 Conception d'une chaîne de traitement pour la prédiction par drone du taux de chlorophylle dans les bananeraies . . . . .	115
5.3.1 Abstract . . . . .	115
5.3.2 Introduction . . . . .	115
5.3.3 Material and methods . . . . .	116
Calibration of a prediction model . . . . .	116
In-field assessment of the model . . . . .	118
5.3.4 Results and discussion . . . . .	119
PLS-R model for LCC estimation . . . . .	119
In-field LCC estimation . . . . .	120
5.3.5 Conclusion . . . . .	120
5.4 Perspectives d'utilisation de ces deux capteurs dans les bananeraies . . . . .	121
5.4.1 Utilisation du Dualex . . . . .	121
5.4.2 Utilisation de la télédétection et du capteur Sequoia . . . . .	122
Télédétection . . . . .	122
<b>6 Conclusion</b>	<b>123</b>
6.1 Méthodes permettant de caractériser l'asynchronisme d'un peuplement de bananiers . . . . .	124
6.2 Prise en compte de l'asynchronisme pour analyser les conditions de croissance dans les bananeraies . . . . .	125
6.3 Exploitation de la variabilité des conditions de croissance en bananeraies . . . . .	126
6.4 Intérêt et limites de l'agriculture de précision dans les bananeraies . . . . .	126
<b>Table des figures</b>	<b>142</b>
<b>Liste des tableaux</b>	<b>143</b>
<b>Publications et participations à des conférences</b>	<b>145</b>
<b>Remerciements</b>	<b>147</b>
<b>Résumé</b>	<b>149</b>
<b>Abstract</b>	<b>151</b>

# Chapitre 1

## Introduction

Dans ce chapitre, je présente les principales caractéristiques du bananier et de sa culture industrielle. Ces descriptions sont importantes pour comprendre les différences avec les grandes cultures sur lesquelles les concepts de l'agriculture de précision ont été initialement définis dans les années 1990. En particulier, les éléments qui expliquent pourquoi cette culture est essentiellement manuelle et pourquoi elle est asynchrone seront présentés. Un accent est aussi mis sur les (nouvelles) contraintes qui pèsent sur la culture bananière et qui motivent une amélioration des pratiques culturales et de la gestion. Tous ces éléments permettent de comprendre le contexte de R&D de la Compagnie Fruitière et le contexte particulier de la thèse.

### Sommaire

---

<b>1.1</b>	<b>Le bananier et sa culture . . . . .</b>	<b>8</b>
1.1.1	Caractéristiques biologiques du bananier . . . . .	8
1.1.2	Description du système de production industrielle des bananes d'exportation . . . . .	11
1.1.3	La production bananière, une production importante au niveau mondial et soumise à de nouveaux enjeux . . . . .	17
<b>1.2</b>	<b>Contexte de la thèse et objectifs opérationnels . . . . .</b>	<b>19</b>
1.2.1	Les objectifs de R&D de la Compagnie Fruitière . . . . .	19
1.2.2	Objectifs opérationnels de la thèse . . . . .	20

---



FIGURE 1.1 – Diversité des bananes vendues sur un marché en Inde. (Source : Heslop-Harrison and T. Schwarzacher, 2007)

## 1.1 Le bananier et sa culture

Les éléments généraux sur le bananier et sa production qui sont présentés dans ce chapitre proviennent essentiellement des articles ou livres suivants : Turner et al. (2007) ; Lassoudière (2007) ; Heslop-Harrison and Schwarzacher (2007) ; Ploetz and Evans (2015).

### 1.1.1 Caractéristiques biologiques du bananier

#### Classification biologique

Les bananiers sont des monocotylédones qui font partis de l'ordre des Zingibérales (comme l'arbre du voyageur, les oiseaux du paradis ou encore le curcuma et le gingembre). Les bananiers sont très proches génétiquement des graminées dont ils forment un groupe frère. Cette ressemblance leur vaut le nom d'herbe géante. L'ensemble des bananiers fait parti de la famille des musacées qui est organisée en plusieurs genres. Les bananiers cultivés pour leurs fruits sont le plus souvent parthénocarpique (sans graine) et sont essentiellement issus de l'espèce *Musa acuminata* du genre *Musa*. Ces bananiers cultivés peuvent aussi être issus de croisement entre les espèces *Musa acuminata* et *Musa balbisiana*.

Plus de 1200 variétés (aussi appelées cultivars) existent parmi les fruits sans graine. Elles sont utilisées en banane dessert, banane à cuire ou encore pour faire de la bière (Figure 1.1). Le plus souvent, les bananes sucrées et consommées non cuites sont appelées bananes ou bananes desserts, et les bananes plus riches en amidon et consommées cuites sont appelées plantains, bien qu'il n'y ait pas de distinction botanique claire entre ces deux classes. La diversité de ces variétés provient de la sélection par l'homme depuis des millénaires. L'origine géographique de ce groupe a été localisée en Asie du Sud-Est. Les migrations humaines et l'échange de matériel végétal ont favorisé sa dispersion et son utilisation dans les zones tropicales et subtropicales humides des Amériques, d'Afrique et d'Asie jusqu'en Europe (îles Canaries) et en Australie (Queensland). Parmi les variétés cultivées, les bananes desserts du groupe Cavendish sont les plus connues par les consommateurs Européens car ce sont celles qui sont les plus commercialisées. Ce sont des bananes jaunes, courbes, relativement grandes. Quelques bananes telles que la Frayssinette ou la Figue peuvent aussi être trouvées, elles sont plus petites et pas nécessairement courbes. Les différences entre les variétés sont aussi visibles sur les caractéristiques de la plante, mais dans les parties suivantes nous nous intéressons à la description morphologique de la Cavendish, qui est la plus étudiée à cause de son intérêt commercial.

### Description de la souche de bananier ou tige souterraine

Le plant de bananier est constitué d'une tige souterraine également appelée rhizome ou souche qui est à l'origine du système foliaire et du système racinaire (Figure 1.2). Le rhizome est formé d'un ensemble d'unités appelées les méristèmes. Chaque méristème ressemble à un bulbe d'environ 20 cm qui possède un bourgeon terminal au niveau de sa partie apicale et des racines au niveau de sa partie basale. Le bourgeon terminal de chaque méristème émet les feuilles et produit une pousse que l'on appelle couramment un bananier. Chaque bananier ne produit qu'une seule inflorescence avant de mourir (plante monocarpique).

Le bananier est considéré comme une plante semi-pérenne : « semi » parce que chaque bananier ne produit qu'un seul régime de bananes avant de mourir, « pérenne » parce que la souche de bananier peut vivre plusieurs dizaines d'années et produit de nouveaux bananiers indéfiniment.

### Description du système racinaire

Le système racinaire du bananier est fasciculé sans pivot. Les racines du bananier sont de trois classes, les racines primaires, secondaires et tertiaires. Les racines primaires sont plus grandes et peuvent atteindre plusieurs mètres. Elles poussent de plusieurs centimètres par jour. Les racines secondaires font généralement moins d'un mètre et les racines tertiaires mesurent quelques centimètres.

### Description du système foliaire

Chaque pousse est constituée d'un ensemble de feuilles produites par le bourgeon au rythme d'environ une feuille par semaine. Les feuilles sont composées d'un pétiole, d'une nervure principale, de nervures secondaires et d'un limbe. Au fur et à mesure de leur émission par le bourgeon, leur dimension maximale change, elles s'allongent pour faire au final près de 2 m 50. L'empilement des pétioles forme le pseudo-tronc qui va croître en hauteur et en largeur. Quand environ 30 feuilles sont apparues, le bourgeon se différencie en bourgeon floral qui va produire une inflorescence. A ce moment, le potentiel de feuilles maximum que la pousse peut produire est atteint. Environ 12 feuilles vont encore être émises puis l'inflorescence va émerger et être visible au moment de la floraison.

### Description de l'inflorescence

L'inflorescence (Figures 1.3 et 1.4) est constituée de fleurs mâles et femelles qui sont insérées sur la hampe en groupes appelées les mains. Chaque main est constituée de deux rangées de fleurs. Les mains femelles sont initiées en premier, et sont constituées d'environ 10 à 40 fleurs qui vont grossir pour donner des bananes aussi appelées doigts. Chaque main est produite par le bourgeon terminal à un rythme d'environ une main par jour. Finalement, entre 5 et 15 mains femelles sont produites puis ce sont des fleurs mâles. Ces dernières sont la plupart du temps caduques. L'initiation de la floraison est considérée comme indépendante de la photopériode (day-neutral). Elle peut avoir lieu toute l'année et dépend principalement de l'état de développement du pied. La croissance du fruit se fait en plusieurs phases : la première phase de division cellulaire dure environ quatre semaines, la seconde phase, qui commence en même temps que la phase de division mais qui dure plus longtemps, est une phase d'accroissement cellulaire (remplissage d'amidon), et enfin la troisième phase est la phase de maturation.

Une particularité de ce fruit est d'être climactérique, c'est-à-dire que son mûrissement est initié par l'éthylène qui est produit par le fruit lui-même ou par une source extérieure et qui déclenche une augmentation de l'activité respiratoire du fruit et une succession de transformations. Ces transformations caractéristiques du mûrissement incluent la dégradation de la chlorophylle de la peau qui va passer du vert au jaune, la dégradation de l'amidon en sucre, la diminution de l'épaisseur de la peau, etc.

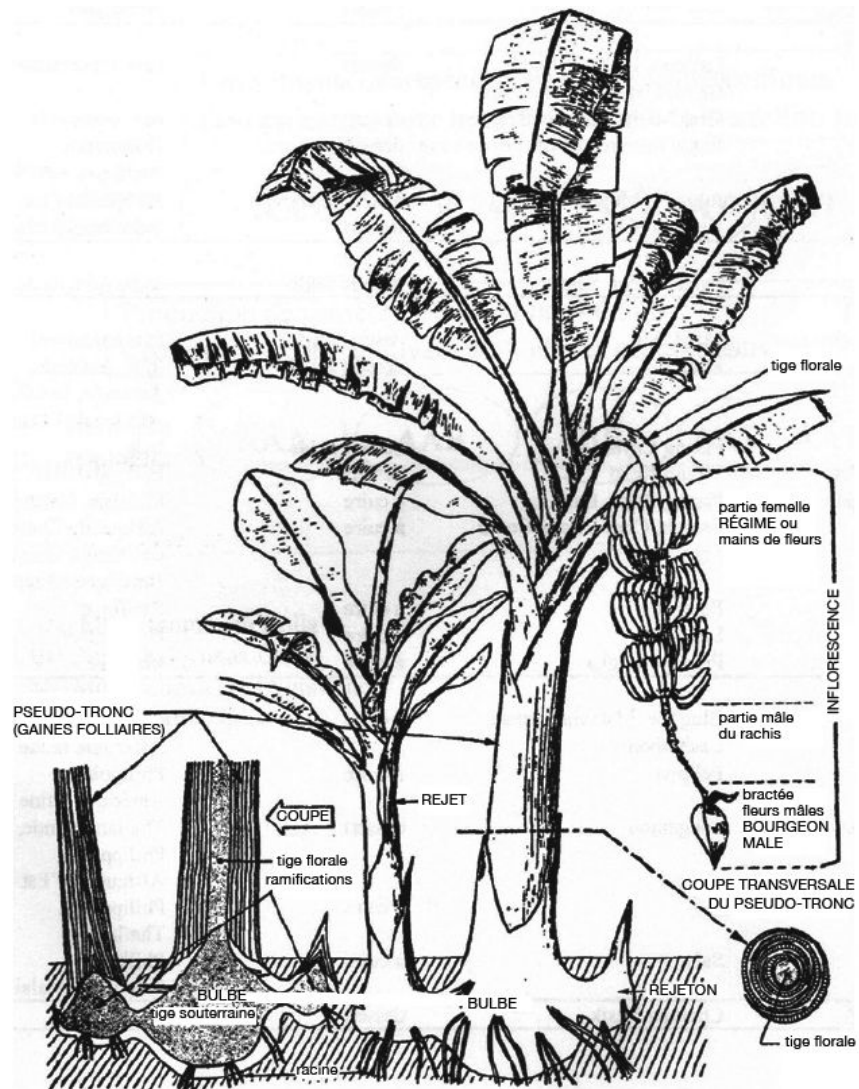


FIGURE 1.2 – Schéma d'un bananier et de son rejet. (Source : Champion 1963)

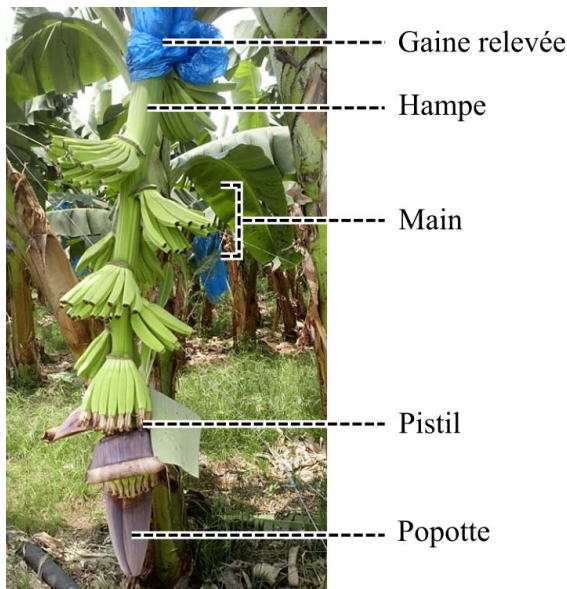


FIGURE 1.3 – Schéma d'une inflorescence de bananier (Photo : Paul Esnault)



FIGURE 1.4 – Inflorescence au stade « Tête de cheval » (Photo : Paul Esnault)

### 1.1.2 Description du système de production industrielle des bananes d'exportation

#### Caractéristiques générales du système de production

Les systèmes de cultures bananiers sont très diversifiés. Les systèmes de cultures vivriers sont très différents de ceux industriels avec une vocation pour l'export auxquels nous nous intéressons dans la thèse. Ces derniers sont mis en place dans des plantations de très grandes surfaces, avec des rendements d'environ 45 à 75 tonnes par hectare et une variabilité génétique très faible car les bananiers sont des clones. C'est une production intensive en intrants (engrais et produits phytosanitaires, avec par exemple environ 400 kg d'azote et 700 kg de potassium apportés par hectare et par an) dont la charge représente environ 40 % du coût de production total (Ploetz and Evans 2015). C'est aussi une culture intensive en main d'oeuvre avec environ 1 à 1.5 employés par hectare. En effet de nombreuses opérations culturales sont réalisées sur les plants et les fruits (plantation, sélection des rejets, éclaircissage des fruits, suppression des pistils, protection du régime par une gaine, pose de mousse entre les mains pour éviter les frottements, tri des fruits récoltés). Elles sont souvent complexes et difficilement mécanisables.

#### Modes de multiplication

Les bananiers ne peuvent pas être multipliés par reproduction sexuée car, comme nous l'avons vu, les bananes d'export sont parthénocarpiques. En conséquence, deux modes de production des plants sont utilisés : la reproduction *in vitro* (par embryogénèse somatique ou par bourgeonnement) et la multiplication végétative. La multiplication végétative consiste à utiliser des rejets produits naturellement par le rhizome. La multiplication *in vitro* consiste à générer de des plants à partir de petits fragments tissulaires, les explants, dans un milieu artificiel. L'avantage de la multiplication *in vitro* est de pouvoir créer de nombreux plants et surtout, au contraire de la reproduction végétative, de produire des nouveaux plants sains et exempts de virus. Les micro-plants sont produits dans des laboratoires et sont transférés en pépinière pour subir une phase d'acclimatation avant d'être plantés dans les parcelles.

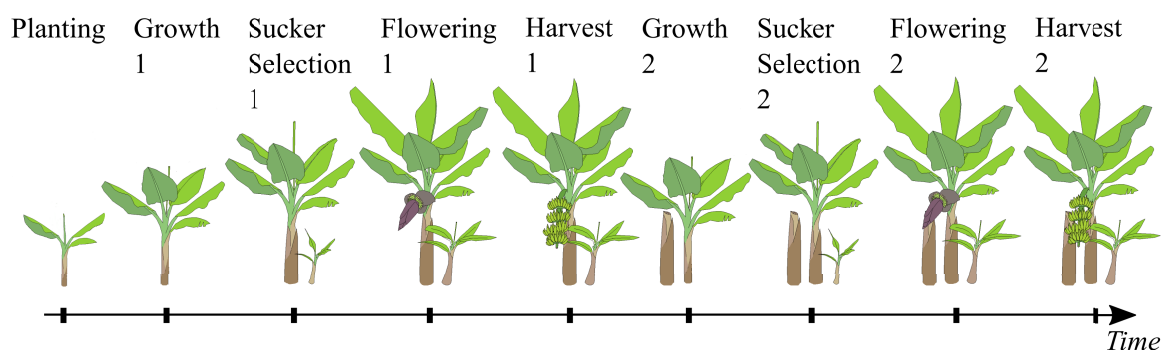


FIGURE 1.5 – Schéma de développement d'un bananier dans une plantation

### Conduite des parcelles

Les nouvelles parcelles sont plantées avec des micro-plants à une densité de 1300 à 2000 plants par hectare. Les micro-plants ont alors plusieurs semaines et environ six feuilles. Ils vont se développer selon plusieurs phases : une phase juvénile d'acclimatation, puis une phase de croissance végétative et enfin une phase de reproduction pendant laquelle l'inflorescence apparaît et croît pour donner un régime de bananes. Cette phase se termine à la récolte du régime qui condamne aussi le pied qui est coupé. La pérennité de la production est assurée par la sélection d'un seul rejet (multiplication végétative) parmi l'ensemble des rejets produits régulièrement par le pied parent. Cette sélection est faite plusieurs semaines après la mise en place de la parcelle. Le rejet, qui est appelé le « fils » va assurer le cycle de floraison suivant en se développant à son tour selon une phase de croissance végétative et une phase reproductive. Ce cycle est répété jusqu'à la mise en jachère de la parcelle. Le schéma du développement des bananiers dans une parcelle est représenté Figure 1.5.

Les parcelles assurent plusieurs cycles de floraison avant d'être mises en jachère. Cette décision est prise lorsque la vigueur de la parcelle chute. Plusieurs raisons peuvent en être à l'origine : une pression parasitaire forte, une baisse de fertilité du sol, une erreur de conduite de plantation, ou tout simplement le vieillissement des pieds. Cela peut être aussi un problème de compaction du sol, une baisse brutale de la densité productive (suite à une tornade sévère), etc.

Pour éviter la baisse de rendement d'une parcelle, le travail des producteurs est de veiller à maintenir la population productive des parcelles à leur maximum, en changeant ou replantant les pieds faibles mais aussi en sélectionnant les meilleurs rejets successeurs. Le remplacement des pieds faibles ne se fait pas nécessairement avec des micro-plants mais aussi par multiplication végétative. Dans ce cas, les nouveaux plants sont prélevés dans les parcelles en production qui présentent des rejets surnuméraires ou dans les parcelles qui vont être mises en jachère.

### Gestion des maladies du bananier

Le bananier est sensible à de nombreuses maladies et ravageurs qui vont affecter directement les rendements en perturbant la nutrition ou la photosynthèse ou alors la qualité des fruits en affectant leur maturité ou leur apparence cosmétique. Nous présentons ici quelques unes des maladies et ravageurs importants en termes de dommages et de coûts de production, tels qu'identifiés par Gonçalves and Kernaghan (2014).

Les cercosporioses noires et jaunes sont des maladies fongiques qui entraînent une nécrose des feuilles et une perte de rendement jusqu'à 50 %. Ces maladies diminuent aussi la qualité des fruits en induisant un mûrissement précoce pendant le transport (Castelan et al. 2012) ce qui peut empêcher l'export en cas d'infestation sévère. Les moyens de luttent incluent des mesures prophylactiques qui visent à réduire la

quantité d'inoculum en supprimant les feuilles très atteintes. La lutte chimique permet d'agir de manière préventive pour éviter la dissémination ou de manière curative. Excepté aux Antilles françaises, cette lutte se fait par voie aérienne (avion, ULM, hélicoptère) juste au dessus de la canopée plusieurs fois par an (jusqu'à un traitement par semaine). Depuis les années 1990, des protocoles ont été mis en place dans les plantations pour suivre l'état d'évolution de la maladie à partir d'observation des symptômes sur les feuilles. Cela permet de surveiller l'évolution de la maladie et d'adapter les traitements pour éviter de perdre le contrôle de la maladie sachant qu'elle se développe très rapidement à partir d'un seuil d'infestation.

Les bananiers sont aussi vulnérables aux nématodes, vers microscopiques, qui s'attaquent à leur racines. Ils limitent la capacité d'absorption des racines et fragilisent l'ancrage des bananiers dans le sol. Le cultivar cavendish est particulièrement sensible à ce ravageur du fait de la taille importante de son régime qui favorise sa chute et qui oblige les planteurs à arrimer les bananiers porteurs d'un régime avec des ficelles (haubanage). Les techniques culturales permettent de limiter la dissémination des nématodes et diminuer leur pression sur les cultures. Il s'agit d'utiliser des bananiers obtenus *in vitro* lors de la mise en place des nouvelles parcelles ce qui permet d'avoir du matériel végétal sain par rapport à l'utilisation de rejets sélectionnés sur des parcelles potentiellement contaminées. Cette pratique permet de limiter l'usage de nématicides jusqu'à environ 4 ans après la mise en plantation (Lassoudière 2007). Après quelques années les parcelles sont mises en jachères et les bananiers et adventices hôtes des nématodes sont supprimés (jachères améliorées) ce qui permet de diminuer fortement la quantité de nématodes dans les parcelles (Chabrier and Queneherve 2003).

Le charançon du bananier est un coléoptère dont la larve creuse des galeries dans le rhizome. Ses effets sont similaires à ceux induits par les nématodes (diminution de la capacité d'absorption des racines, fragilisation de l'ancrage au sol). Les traitements chimiques contre le charançon ont été tristement célèbres (Chlordécone) à cause de leur rémanence dans les sols, la contamination de la chaîne alimentaire et de leurs effets sur la santé (perturbateur endocrinien : Multigner et al. 2010). Des méthodes alternatives aux traitements chimiques sont utilisées. De manière similaire à la gestion des nématodes, l'usage de matériel sain, les jachères améliorées permettent de réduire la population de charançons. Des pièges à phéromones ou des pièges à encoches (création d'une encoche dans des pseudo-troncs des pieds de bananiers qui viennent d'être récoltés, encoche dans laquelle vont se réfugier les charançons qui sont ensuite collectés manuellement) peuvent aussi être mis en place.

Enfin, on peut citer les maladies post-récoltes qui se développent sur les fruits récoltés et qui vont affecter la qualité cosmétique du fruits (décoloration) et leur maturité. Nous pouvons citer la pourriture de couronne, qui est traitée par dépôt de fongicide sur le pédoncule des bananes ou aussi par le dépôt de cire (en lutte biologique).

### Gestion de la nutrition hydrique et minérale

Le bananier est très sensible aux excès d'eau comme aux déficits. Le stress hydrique induit des modifications visibles du développement végétatif et reproductif. La distance entre deux pétioles peut être très réduite ce qui crée un engorgement des feuilles du bananier qui peut aller jusqu'à l'obstruction de la remontée de la fleur qui va se créer un passage par le pseudo-tronc au lieu d'être émise à l'extrémité du plant. Lors d'un déficit d'eau, l'apparence des fruits change : les bananes sont plus courbes et le régime est plus fermé (la distance entre les mains du régime est plus faible). Les parcelles de bananiers sont le plus souvent irriguées et les excès d'eau sont limités par des réseaux de drainage (fossés de 30 cm à 2 m de profondeur espacés tous les 10 m dans les bas fonds et les sols lourds). La consigne d'irrigation dépend de la météorologie et notamment de la pluviométrie et de l'évaporation. En saison des pluies, l'irrigation est arrêtée.

Les apports en engrais se font de manière fractionnée et très fréquente (mensuelle). Cela permet de limiter l'effet de la lixiviation dans les régions tropicales où la pluviométrie peut être très forte. Les besoins du bananier dépendent de son stade phénologique mais les apports sont raisonnés par rapport à l'âge de la parcelle plutôt qu'à l'état de développement des pieds pour des raisons de facilité des travaux. La



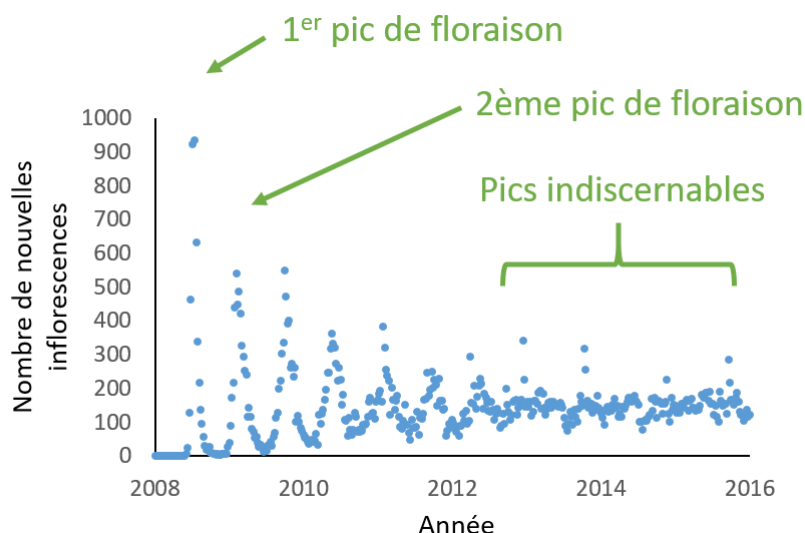


FIGURE 1.6 – Exemple de courbe de floraison d'une parcelle de bananiers étudiée au cours de la thèse. La première floraison est relativement synchronisée dans le temps, puis les floraisons suivantes sont de plus en plus étendues jusqu'à ce que les pics ne soient plus discernables

composition du sol ainsi que la composition foliaire servent à adapter les apports aux besoins de chaque parcelle. Les différentes carences (en N, K ou en micro-éléments) induisent des effets différents sur le bananier. L'observation des bananiers permet aussi de détecter des carences. Par exemple, la coloration de la bordure des feuilles en rouge est un symptôme d'une carence modérée en azote (Lassoudière 2007).

### L'asynchronisme du développement des bananiers au sein des parcelles

La production des bananiers d'une parcelle se désynchronise au fur et à mesure des cycles de floraisons. Initialement, tous les pieds plantés sont approximativement au même stade phénologique, mais au cours de leur développement et de la succession des cycles de floraison, un décalage se crée entre chacun des plants. Cela s'explique par les particularités biologiques du bananier détaillées précédemment (plante day-neutral, monocarpique et semi-pérenne) mais aussi par la conduite pluri-annuelle des parcelles et par le fait que la sélection des rejets ne se fasse pas de manière synchronisée ce qui fait qu'après chaque cycle le décalage s'accroît. L'asynchronisme a plusieurs conséquences : la première est que la production est continue dans le temps sur une parcelle âgée. En effet, si l'on représente le nombre de nouvelles floraisons dans une parcelle (Figure 1.6), on voit que la première floraison des bananiers se fait de manière relativement synchronisée puis qu'au fur et à mesure du temps, les floraisons sont de moins en moins synchronisées jusqu'à être continues. La deuxième conséquence est que les parcelles sont constituées de plants à différents stades phénologiques (Figure 1.7) et que la diversité des stades phénologiques augmente au cours du temps (Figure 1.8). Ces particularités imposent des contraintes fortes sur la gestion technique des parcelles et notamment sur la gestion des récoltes.

### Gestion des récoltes

Les bananes exportées doivent satisfaire à des standards de taille et de qualité. Par ordre d'importance décroissant, il y a d'abord la tenue du fruit (pas de mûrissement prématuré avant l'arrivée), puis la qualité sanitaire du fruit (présence ou non de moisissures ou pourritures à l'arrivée) enfin vient la qualité cosmétique avec la présence ou non de défauts visibles sur la peau (grattage, choc, pliures de pédoncules,

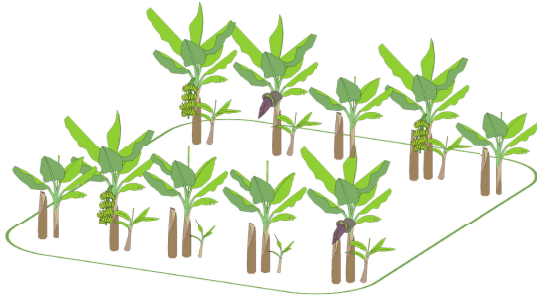


FIGURE 1.7 – Représentation de la diversité des stades phénologiques au sein d'une parcelle

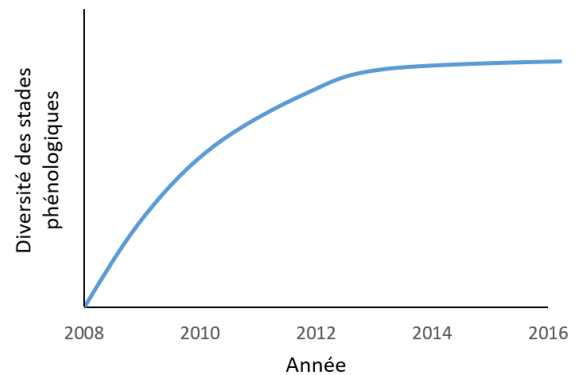


FIGURE 1.8 – Représentation de l'augmentation de la diversité des stades phénologiques au cours du temps. Au moment de la plantation (en 2008) les stades phénologiques sont identiques, puis au fur et à mesure de l'âge de la parcelle et du fait de l'asynchronisme, la diversité des stades phénologiques augmente

etc.). La qualité des soins apportés aux fruits sur la parcelle est donc très importante ainsi que le moment de la récolte et le conditionnement.

Pour connaître l'âge des régimes dans les parcelles, des ouvriers arpentent les parcelles une à deux fois par semaine pour identifier les nouvelles inflorescences et les marquer avec un ruban de couleur. La couleur dépend de la date d'apparition des inflorescences et permet aux ouvriers de connaître l'âge des régimes dans la parcelle. La décision de récolte dépend généralement de deux critères : le diamètre d'un doigt de référence du régime (souvent le doigt central de la deuxième main) et l'âge thermique du régime. L'âge thermique est calculé selon la méthode décrite par Ganry and Chillet (2008) en degrés-jours avec une température de base de 14 °C. Les producteurs adaptent les limites de diamètre et de somme thermique en fonction des plantations et de la période de l'année. Ces décisions résultent de l'évaluation de plusieurs critères tels que : la mesure de la maturité appelée la Durée de Vie Verte (DVV) qui permet de connaître le temps entre la récolte des régimes et leur mûrissement spontané dans des conditions normées, la qualité des régimes mesurée en station, le taux de bananes trop mûres après le transport et enfin les capacités de récolte.

La récolte est manuelle et nécessite deux ouvriers pour chaque régime à récolter. Le premier ouvrier se place sous le régime avec un « berceau » qui permet de recueillir le régime une fois qu'il est coupé par le second ouvrier. Le régime est transporté jusqu'à une station de conditionnement soit par tracteur, soit par cable-way. Le cable-way est un système de transport des bananes constitué d'un câble d'acier fortement tendu qui sert de base de roulement pour les chariots porteurs des régimes, qui sont transportés par trains d'environ 20 à 50 grâce à une traction animale (âne) ou thermique.

### Conditionnement et transport

Toutes les parcelles sont situées à proximité d'une station de conditionnement. Les stations sont des usines qui transforment les régimes bruts en trois flux : (i) des déchets, (ii) des cartons de bananes hors standards qui sont vendus sur les marchés régionaux, (iii) des cartons et des palettes de bananes exportés. Un contrôle qualité est effectué à l'entrée de la station pour écarter les régimes à risque (trop mûrs) et

pour évaluer la qualité de l'itinéraire technique et des soins aux fruits. Les régimes sont ensuite dépatés<sup>1</sup> et triés et pour obtenir des mains puis des bouquets<sup>2</sup> de bananes qui sont mis en carton par type en fonction de leur qualité et de leur taille. Il y a par exemple des qualités « premium » ou « premiers prix » et des « P20 » et « P14 » qui sont des bananes de qualité premium dont les tailles minimales sont respectivement de 20 cm et de 14 cm.

Les cartons sont ensuite envoyés par transport terrestre jusqu'aux ports ou ils sont transportés dans des containers ou dans les cales. Le mûrissement pendant le transport présente un risque économique important. En effet, cette transformation du fruit est associée à un dégagement d'éthylène et de chaleur qui induit, à son tour, le mûrissement des bananes à proximité. Cela peut initier une réaction en chaîne si les capacités de refroidissement et de ventilation sont dépassées et finalement à une perte de la cargaison. Les conditions de transport sont donc très maîtrisées. La température est fixée à 13.5 °C qui est la température minimale permettant de limiter l'évolution des fruits. Une température inférieure entraînerait de la « frisure », c'est-à-dire une altération physiologique irréversible de la banane qui entraîne une décoloration et une diminution de la qualité gustative après mûrissement. Une atmosphère contrôlée est mise en place pour limiter la respiration du fruit et enfin les temps de transport et de stockage sont limités au maximum. Par ailleurs un contrôle fort est mis en place tout au long de la chaîne logistique. Parmi ces contrôles il y a la mesure régulière de la température des palettes pour identifier celles qui commencent à mûrir précocement, la recherche de l'odeur de palettes mûres et des contrôles aléatoires. Enfin, les palettes sont transportées par origine afin d'éviter de disséminer des palettes trop mûres qui seraient difficiles à localiser. L'objectif de ces contrôles est de pouvoir facilement identifier les palettes à risque et de les isoler. A l'arrivée en Europe les bananes sont ensuite mûries au sein de mûrisseries où le pic climactérique est déclenché artificiellement par injection d'éthylène gazeux. Ce système permet de contrôler finement le degré de mûrissement et la couleur des bananes pour satisfaire les demandes des distributeurs.

### Système d'information de la production

Le système d'information de la production repose sur de nombreuses saisies manuelles réalisées par les ouvriers. A l'exception des traitements aériens pour lesquelles la traçabilité des traitements est enregistrée automatiquement, toutes les opérations réalisées sur les parcelles et jusqu'au conditionnement sont saisies manuellement dans des cahiers puis ressaisies dans des bases de données. Une particularité forte du fonctionnement des plantations est l'utilisation de codes visuels pour transmettre de l'information. Pour illustrer cette particularité, je détaille les procédures de comptage des floraisons et des régimes matures car les données issues de ces comptages ont été utilisées à plusieurs reprises dans la thèse.

Nous avons déjà vu que l'âge des régimes dans les parcelles est connu grâce une lanière de couleur qui est accrochée sur l'inflorescence. L'ouvrier « Compteur » part le matin sur des parcelles définies avec autant de sacs que de parcelles. A chaque nouvelle inflorescence, il coupe l'extrémité du bourgeon floral, la « popotte », qu'il laisse dans le sac de comptage. Il grave aussi le numéro de la parcelle sur une banane particulière du régime. En sortant de la parcelle, il indique sur un doigt de banane le nombre de fleurs qu'il a comptées, et le dépose dans le sac. A la fin de la journée, le contrôleur vérifie les nombres annoncés par les compteurs en recomptant le nombre de popottes dans les sacs. Par ailleurs, en cas de chute de la lanière de comptage, le fait que la popotte soit coupée permet de savoir si l'inflorescence a déjà été comptée ou si elle a été oubliée. Le contrôleur enregistre le nombre de nouvelles fleurs par parcelle dans un registre qui est régulièrement transcrit dans une base de données.

La décision de récolte des régimes est réalisée par des ouvriers chargés de l'opération de précalibrage. Cette procédure vise à mesurer le diamètre moyen d'une banane de référence sur les régimes pour décider si le régime est mature. Le deuxième critère est l'âge. Les employés précalibreurs mettent une lanière de couleur, appelée la lanière de précalibrage, dont la couleur et le motif est spécifique du début ou de la fin de semaine sur tous les régimes suffisamment âgés ou qui ont atteint le diamètre de récolte. Ils coupent

1. Dépatage : jargon bananier qui signifie séparer les mains du régime de sa hampe

2. Bouquet : sous-ensemble de bananes issue d'une main. Souvent un bouquet est constitué de 6 bananes

aussi en deux une banane particulière appelée la banane « tire-sève » qui est située en bas du régime et conservent une partie de la lanière ou du tire-sève pour compter le nombre de régimes prêts à être récoltés dans chaque parcelle. Ces informations sont aussi reportées dans un registre qui est régulièrement enregistré dans une base de données.

Les ouvriers chargés des récoltes ne doivent couper que les régimes qui portent une lanière bien précise. Ils notent leur numéro d'équipe sur la hampe de chaque régime qui est ensuite amené en station de conditionnement. Les lanières y sont contrôlées et les lanières ainsi que le numéro d'équipes permettent de savoir s'il y a eu une erreur dans le précalibrage ou dans la récolte. Par ailleurs, l'état de cicatrisation du tire-sève permet d'être certain que le régime a bien été précalibré plusieurs jours avant la récolte même si la lanière de précalibrage est manquante. Le nombre de régimes récoltés ainsi que leur âge et leur diamètre est enregistré par parcelle (on rappelle que le nom de chaque parcelle est gravé sur une banane de chaque régime).

Finalement, ces codes visuels servent aussi bien à renseigner les opérateurs qu'à les contrôler. En effet, des stratégies d'évitement ont été mises en place au fur et à mesure de la vie des plantations par les employés pour diminuer la pénibilité du travail ou pour augmenter leurs profits (souvent les gestes sont payés à la tâche). De nouveaux contrôles sont donc ajoutés pour parvenir à un équilibre et à un niveau de fiabilité de l'information plus ou moins acceptable. On voit donc que le système d'information, même s'il est très peu technologique, est relativement précis et que les observations sont, du moins pour la gestion de la récolte, exhaustives. Ces données sont aussi localisées par parcelle. Ces données représentent donc une information potentiellement intéressantes car en recouvrant l'ensemble de la plantation elles permettent d'avoir des observations sur des situations contrastées de production.

### **1.1.3 La production bananière, une production importante au niveau mondial et soumise à de nouveaux enjeux**

#### **Caractéristiques de la production mondiale**

La banane est le fruit le plus produit au monde avec près de 120 millions de tonnes produites en 2016, juste devant les pastèques (117 millions de tonnes) et les pommes (89 millions de tonnes). L'Inde, la Chine et l'Indonésie sont les pays qui ont la production la plus importante (Figure 1.9). Seulement une petite partie de la production mondiale est exportée, environ 20 %. Les principaux pays exportateurs sont les pays d'Amérique du Sud (Figure 1.10) et les principaux pays qui importent sont les États-Unis, les pays Européens et la Russie (Figure 1.11).

#### **Organisation et mutation du marché mondial**

Le marché est très compétitif et une véritable guerre commerciale s'est organisée entre « les bananes dollars », produites dans les zones d'influence politique des États-Unis, et les bananes « ACP » Afrique Caraïbe Pacifique, produites dans les anciennes colonies européennes. Ces dernières profitent de tarifs douaniers préférentiels en Europe qui les protègent des bananes bon marché d'Amérique latine. Cependant, suite à des actions de justice à l'Organisation Mondiale du Commerce, un compromis a depuis été trouvé pour rectifier ce désavantage concurrentiel. Depuis, les tarifs douaniers pour ces pays diminuent chaque année pour passer de 176 dollars la tonne en 2009 à un objectif de 114 dollars la tonne entre 2017 et 2019. Cela favorisera l'arrivée des bananes d'Amérique latine dans les magasins européens, et entraînera les ACP à diminuer leurs coûts de production pour garder des marges viables. En compensation, l'Union Européenne a créé les Mesures d'Accompagnement de la Banane (MAB) afin d'améliorer la compétitivité des pays ACP. L'enveloppe totale de 200 millions d'Euros a pour objectif d'améliorer la productivité et la « qualité de l'ensemble des facteurs de production ».

Pendant de nombreuses années, de grandes multinationales verticalement intégrées, c'est-à-dire qui contrôlent la production, l'emballage, l'expédition, l'importation et le mûrissage comme Chiquita, Fresh

del Monte Produce, Dole Food, Fyffes et Noboa contrôlaient le marché de la banane. Ces cinq multinationales détenaient à elles seules près de 86 % du marché de la banane dans les années 90. Aujourd'hui, le modèle de production mondiale change et ces compagnies ne détiennent plus que 40 % du marché. Des entreprises nationales se sont lancées dans la production et la grande distribution s'est mise à acheter directement les bananes aux producteurs, à les transporter et à les faire mûrir. De plus en plus ces grandes multinationales réduisent leur rôle dans la production pour devenir uniquement des marchands (acheteurs, transporteurs, mûrisseurs). Ce changement a lieu à cause de l'augmentation des coûts de production mais aussi à cause de l'augmentation des contraintes de qualité environnementale et sociale. De manière générale, les consommateurs et la grande distribution sont de plus en plus regardants sur la qualité socio-environnementale des produits qu'ils consomment. Les distributeurs se sont adaptés et contrôlent leurs filiales d'approvisionnement en leur demandant d'obtenir des certifications comme Tesco Nurture, Global Gap, ISO 14000 ou FairTrade (Lassoudière, 2007). Les audits pour la certification concernent toute la production, les produits phytosanitaires employés, les énergies utilisées ou encore la sécurité des travailleurs, et des documents doivent leur être remis dans un laps de temps réduit (4 h à 24 h).

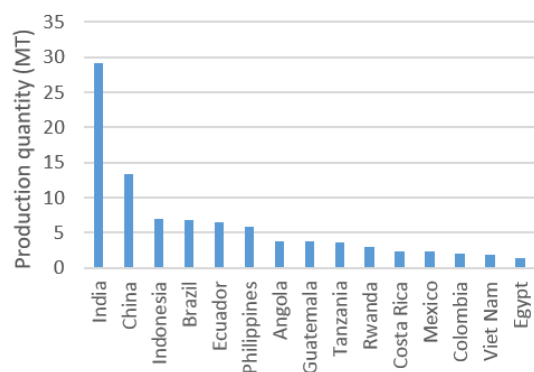


FIGURE 1.9 – Quantités de bananes produites par pays en 2016 en millions de tonnes (source : FAO-STAT 2016)

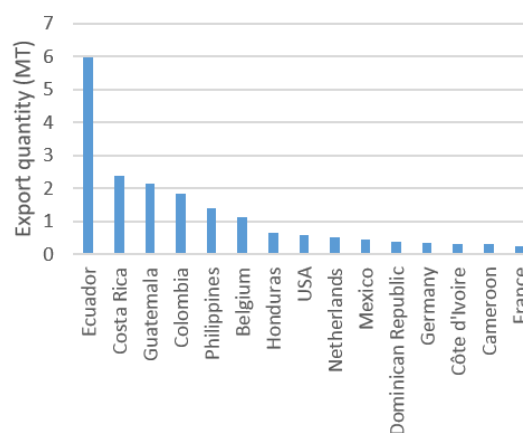


FIGURE 1.10 – Quantités de bananes exportées par pays en 2016 en millions de tonnes (source : FAO-STAT 2016)

### Des risques importants liées à l'apparition et à la propagation de nouvelles maladies

L'industrie bananière a déjà subi plusieurs pandémies. La plus importante à sans doute été celle de la maladie du Panama, déclenchée par un champignon tellurique appelé *Fusarium oxysporum f. sp. cubense*, qui a décimé la production mondiale de bananes dans les années 60 et qui a fait disparaître des plantations industrielles le cultivar Gros Michel qui a ensuite été remplacé par le cultivar Cavendish. Pendant plusieurs décennies, cette maladie n'a plus été une menace et c'est un autre champignon, la cercosporiose noire, qui a posé des difficultés à l'industrie bananière. L'apparition et la dissémination mondiale de ce champignon a forcé l'industrie bananière à dépendre encore plus des fongicides qu'ils ont appliqués jusqu'à 60 fois par an et qui a diminué dans certains cas les rendements par deux. Aujourd'hui, une nouvelle souche du champignon *Fusarium oxysporum* a fait son apparition, la souche TR4 pour Tropical Race 4, qui se propage relativement lentement mais qui rend les sols impropres à la culture de la Cavendish pendant plusieurs décennies sans qu'aucun moyen de lutte ne soit connu, si ce n'est, de nouveau, le remplacement de ce cultivar par un nouveau qui serait résistant.

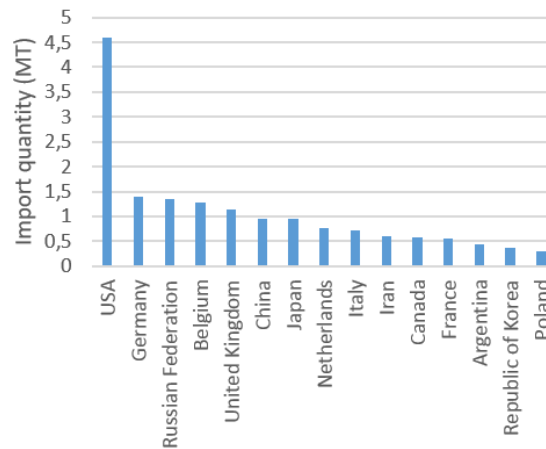


FIGURE 1.11 – Quantités de bananes importées par pays en 2016 en millions de tonnes (source : FAOSTAT 2016)

## 1.2 Contexte de la thèse et objectifs opérationnels

### 1.2.1 Les objectifs de R&D de la Compagnie Fruitière

#### Présentation de la Compagnie Fruitière

La Compagnie Fruitière est une entreprise spécialisée dans l'importation et la distribution de fruits et légumes. Sa gamme de produits est diversifiée. Elle comprend aussi bien du poivre que des fruits exotiques et des productions plus classiques telles que la pomme de terre. Sa production majeure est la banane avec environ 400 000 tonnes produites chaque année dans ses plantations en Afrique (Cameroun, Côte d'Ivoire, Ghana) et en Amérique centrale (Costa Rica). Elle emploie plus de 18 000 personnes à travers ses différentes filiales qui contrôlent l'ensemble de la production et de la logistique depuis les plantations jusqu'à la commercialisation.

#### Objectifs de R&D de la Compagnie Fruitière

Comme nous l'avons vu, le marché de la banane est de plus en plus concurrentiel et de nouvelles contraintes sociétales, agronomiques et techniques pèsent sur la production bananière. Ces mutations affectent aussi la Compagnie Fruitière qui cherche à améliorer ses procédés de production à travers une démarche de R&D. On peut séparer les objectifs de R&D de la Compagnie Fruitière en deux. Le premier objectif est de répondre à des questions stratégiques pour l'entreprise qui sont :

Quel système de culture mettre en place dans les années à venir ? C'est-à-dire, comment repenser celui-ci pour qu'il soit pérenne et rentable connaissant les mutations actuelles et les risques sur la production ?

Comment intégrer le numérique et les nouvelles technologies dans le système de production actuel et dans les systèmes futurs ? C'est-à-dire, quelles technologies et avec quels usages et intérêt ? Avec aussi un questionnement sur les métiers des producteurs : est-ce un risque qu'ils passent moins de temps sur le terrain, et plus de temps derrière un ordinateur ?

Le deuxième objectif pour la Compagnie Fruitière est plus opérationnel :

Améliorer la traçabilité des produits pour répondre aux demandes des distributeurs et des consommateurs mais aussi pour être capable de faire des analyses sur l'efficacité de la chaîne de production

de la plantation jusqu'aux mûrisseries.

Diminuer puis supprimer l'usage des produits phytosanitaires pour des logiques de coûts, d'environnement et d'acceptation sociale.

Moderniser les systèmes de gestion de la production pour rendre les données plus fiables, pour pouvoir les interpréter et en faire un diagnostic et enfin pour décider. L'objectif étant aussi d'attirer une nouvelle génération d'employés qui a été formé à l'usage de ces technologies.

Il y a bien sûr des contraintes fortes liées aux particularités de la production bananière. Comme nous l'avons vu les particularités agronomiques, mais aussi, et c'est important pour toutes les problématiques liées aux nouvelles technologies, des contraintes structurelles. Il faut en effet prendre en compte le fait que les plantations sont situées dans des zones où le tissu industriel est faible ce qui force la Compagnie Fruitière à internaliser beaucoup de métiers et à importer une grande partie des consommables techniques pour résoudre les pannes. Ce sont aussi des zones où la main d'œuvre peut avoir une qualification très hétérogène avec notamment des ouvriers qui peuvent être analphabètes.

### **Moyens mis en œuvre**

La Compagnie Fruitière a initié en 2013 un projet de modernisation du Système d'Information de la Production Agricole (SIPA) visant à fiabiliser les informations et à faciliter leur mobilisation pour améliorer les décisions. Ce projet a pour objectif de mettre en place un outil web permettant d'enregistrer, de visualiser et de mobiliser les observations et indicateurs enregistrés durant tout le processus de production. Cet outil s'appuie sur un système d'information géographique qui a été mis en place après la cartographie de l'ensemble des parcelles en production. Parallèlement aux développements logiciel, les systèmes de mesures sont modernisés pour fiabiliser l'acquisition d'information par la mise en place de systèmes de mesure ou d'enregistrement automatisés (balances électroniques, stations météo connectées, terminaux portables de saisies, etc.). La majorité de ces développements est effectuée par des services internes à la Compagnie Fruitière.

Parallèlement à ces développements, des partenariats ou des prestations ont été initiées avec d'autres organismes comme le CIRAD, l'INRA, Météo France, le CNRS et l'IRSTEA pour effectuer de la recherche. Les 3 thématiques abordées à ce jour sont :

1. l'amélioration agronomique des systèmes de culture, notamment avec le CIRAD ;
2. l'amélioration des traitements phytosanitaires pour la lutte contre la cercosporiose (IRSTEA, CNRS). Ce projet est toujours en cours en 2019 et a été l'objet de la thèse d'Adel Bakache (Bakache et al. 2018 ; Morel et al. 2018) ;
3. l'amélioration de la prise de décision à partir des données de production ou de nouveaux capteurs (IRSTEA, INRA, Météo France).

### **1.2.2 Objectifs opérationnels de la thèse**

Le projet de thèse s'ancre dans le point 3. De manière générale, l'objectif pour la Compagnie Fruitière est de se tenir à jour des avancées technologiques en agriculture. L'entreprise est attirée par les technologies telles que le drone, la télédétection et les capteurs optiques et voudrait les évaluer dans le cadre spécifique de la culture bananière.

L'objectif opérationnel est aussi de proposer des méthodes permettant de valoriser les données de production dans le cadre du projet SIPA. C'est-à-dire, mieux utiliser les données stockées pour le contrôle quotidien ou pour la gestion de la production afin de faire un diagnostic de la production, de fiabiliser les décisions et d'améliorer les prévisions.

## Chapitre 2

# Les difficultés liées à l'asynchronisme pour adopter une démarche d'agriculture de précision

L'objectif de ce chapitre est de définir ce qu'est l'agriculture de précision en tant que discipline scientifique et en tant qu'approche agricole. Les concepts, hypothèses et objectifs de l'agriculture de précision seront détaillés. Une attention particulière sera portée sur la variabilité spatiale des cultures et sur l'intérêt que l'agriculture de précision a donné à cette information pour faire un diagnostic des cultures dans le but d'améliorer la production. Suite à cet état de l'art général, un état de l'art plus spécifique aux recherches publiées en agriculture de précision sur la banane sera présenté. Finalement, les difficultés inhérentes à l'asynchronisme pour étudier la structure spatiale d'une parcelle seront détaillées et les objectifs scientifiques de la thèse seront présentés.

### Sommaire

---

<b>2.1 État de l'art général sur l'agriculture de précision . . . . .</b>	<b>22</b>
2.1.1 Principes et définition de l'agriculture de précision . . . . .	22
2.1.2 Démarche et exemples en agriculture de précision . . . . .	22
2.1.3 Méthodes et disciplines impliquées dans l'agriculture de précision . . . . .	23
2.1.4 Conditions d'application d'une agriculture de précision . . . . .	24
<b>2.2 Les structures spatiales en agriculture de précision . . . . .</b>	<b>24</b>
2.2.1 Origine de la variabilité spatiale des cultures . . . . .	24
2.2.2 Stabilité temporelle de la variabilité spatiale . . . . .	26
2.2.3 Étude et modélisation des structures spatiales . . . . .	26
<b>2.3 État de l'art de l'agriculture de précision sur la banane . . . . .</b>	<b>29</b>
2.3.1 Bibliographie . . . . .	29
2.3.2 Particularités de la culture bananière par rapport aux cultures sur lesquelles s'est développée l'agriculture de précision . . . . .	30
2.3.3 Les difficultés liées à l'asynchronisme pour l'étude de la variabilité spatiale . . .	31
<b>2.4 Objectifs scientifiques de la thèse . . . . .</b>	<b>32</b>

---



## 2.1 État de l'art général sur l'agriculture de précision

### 2.1.1 Principes et définition de l'agriculture de précision

Pour répondre à l'augmentation de la population et à la réduction des surfaces arables sur la planète, de nouvelles approches sont nécessaires pour rendre la production agricole plus durable tout en maintenant une productivité élevée. Parmi les approches proposées, on trouve l'agroécologie, l'agriculture biologique, la biodynamie, l'agriculture raisonnée, l'agriculture intégrée ou encore l'agriculture de précision (Pervanchon and Blouet 2002; USDA 1999). Ces approches sont souvent définies en opposition à ce que l'on appelle l'agriculture « conventionnelle d'après-guerre », basée sur l'usage intensif, systématique et uniforme d'intrants sur des champs monovariétaux de grandes surfaces. Cette agriculture très mécanisée, issue de la « révolution verte », a permis d'améliorer et de pérenniser efficacement les rendements et la sécurité alimentaire. Cependant, elle a engendré des externalités négatives très fortes sur l'environnement et la qualité des aliments et cela n'est plus toléré par les sociétés (Tilman 1999).

Hill and MacRae (1996) ont proposés de classer les pratiques agricoles qui visent à une plus forte durabilité selon trois types E, S et R. La classe E pour « Efficiency increase » regroupe les pratiques qui optimisent l'usage des intrants avec une augmentation associée de la productivité. La classe S pour « Substitution » regroupe les pratiques qui visent à substituer un intrant ou une pratique (par exemple un intrant de synthèse par un intrant d'origine biologique). La dernière classe, R pour « Redesign » est le changement complet de la pratique dans un nouveau système de culture. L'agriculture de précision (AP) se place dans la classe E. Cette stratégie de gestion propose des améliorations méthodologiques et technologiques par rapport à l'agriculture conventionnelle et a été proposée dans les années 1990 pour améliorer les procédés de l'agriculture industrielle (Robert 1993). Le principe de l'AP est de reconnaître que les variabilités spatiales au sein des parcelles sont importantes et qu'en conséquence un itinéraire technique spatialement uniforme n'est pas adapté aux besoins variables des cultures. D'après Oliver (2013b), l'agriculture de précision « uses information and technology to identify and manage the variation within farmed fields ». Il est important de préciser que l'AP est une approche centrée sur l'utilisation et la valorisation des données produites par les exploitants pour améliorer leurs décisions.

Les objectifs de l'AP, en modulant l'itinéraire technique à l'intérieur des parcelles, sont de pouvoir optimiser l'utilisation des intrants et les coûts de production, d'augmenter les rendements, et enfin de diminuer les impacts environnementaux négatifs de l'exploitation agricole (pollution des nappes avec les nitrates et résidus de pesticides, baisse de la biodiversité, etc.). La réalisation de ces objectifs découle de l'utilisation plus adaptée des intrants en supprimant soit les excès, qui augmentent inutilement les coûts de production et les risques environnementaux, soit les déficits qui limitent le rendement potentiel de la culture.

### 2.1.2 Démarche et exemples en agriculture de précision

La décision de modulation d'une culture peut être représentée par les six étapes suivantes (Zwaenepoel and Le Bars 1997; McBratney and Taylor 2000; Öhlmér et al. 1998) :

1. Définition du problème posé ou de la question : quelle étape de l'itinéraire technique voulons-nous moduler ?
2. Observation intra-parcellaire localisée de grandeurs du milieu (sol, plante, atmosphère) ou mobilisation d'observations déjà réalisées
3. Analyse et interprétation des observations. Cette étape vise à transformer les observations en information agronomique utile. Dans cette phase, les modèles et/ou l'expertise peuvent prendre une grande place.
4. Prise de décision. Connaissant l'état de la parcelle, quelle est l'action à faire ? Cette action doit-elle être modulée ?
5. Application modulée ou uniforme en fonction des cas.

6. Enfin la dernière étape est l'évaluation des choix réalisés, étape indispensable pour améliorer continuellement les pratiques

Plusieurs exemples de modulation à toutes les étapes de l'itinéraire technique peuvent être trouvés. La modulation : du travail du sol (Keskin et al. 2011 ; Geoprospectors 2019), des semis (Bullock et al. 1998 ; Shanahan et al. 2004 ; Mehra et al. 2016), de la fertilisation (Coquil and Bordes 2005 ; Diacono et al. 2013) et de l'irrigation (Miller et al. 2017 ; Goldstein et al. 2017), de la taille et de l'éclaircissage des fruits (Bates et al. 2018), de l'épandage des pesticides (Rudd and Roberson 2018), des récoltes (Bramley et al. 2011 ; Antonelli et al. 2017).

Pour un très grand nombre d'applications, la modulation requière l'intervention humaine pour traiter les données et décider avant d'agir. C'est notamment le cas pour toutes les applications où la décision est multi-factorielle. Par exemple, le besoin en irrigation va dépendre du type de sol, du génotype, de l'état de la culture en termes de développement et de maladie, de la météorologie et des objectifs de l'agriculteur. Ce type de modulation est appelé « map-based Variable Rate Application » car l'action modulée est représentée sur une carte de préconisation. L'étude de la variabilité spatiale des observations est donc primordiale et l'expertise de l'agriculteur doit être prise en compte pour conclure sur l'intérêt ou non de moduler l'application. L'application modulée peut se faire avec des appareils classiques mais dans ce cas l'agriculteur doit agir manuellement pour ajuster la dose (par exemple en adaptant la vitesse du tracteur), ou avec des appareils équipés de VRT (variable-rate technology) qui peuvent moduler la dose selon des consignes programmées.

Un nombre plus limité d'applications est aujourd'hui complètement automatisé, c'est-à-dire de la mesure à l'action. Dans ces cas, le capteur mesure une caractéristique de la végétation ou du sol, cette grandeur est interprétée par un modèle, une action est décidée par le contrôleur et enfin l'action est réalisée. Le contrôle automatique de la profondeur de labour, de la fertilisation ou du désherbage en sont des exemples (cf. Grisso et al. (2011) pour une revue des systèmes de modulation en agriculture). Dans ces cas, la géolocalisation n'est pas indispensable et l'intervention de l'agriculteur est faible ou nulle. Ce type de modulation est appelée « sensor-based Variable Rate Application ».

### 2.1.3 Méthodes et disciplines impliquées dans l'agriculture de précision

L'agriculture de précision est une approche ou une pratique agricole qui a été définie dans les sections précédentes. C'est aussi une discipline scientifique animée par une communauté interdisciplinaire qui se réunit lors de conférences internationales (ECPA : European Conference on Precision Agriculture ECPA, ICPA : International Conference on Precision Agriculture) et qui publie dans des revues scientifiques communes (Precision Agriculture, Computers and Electronics in agriculture, etc.)

La recherche et la conception de nouveaux capteurs est primordiale. La démocratisation de l'agriculture de précision a été rendue possible par l'apparition du GPS et des capteurs de rendement qui ont permis d'observer et de quantifier objectivement la variabilité spatiale des cultures. La télédétection a permis de grandes avancées en permettant de caractériser à moindre coût pour l'agriculteur la variabilité du sol et de la végétation sur de grandes surfaces grâce à leur propriétés optiques (Pinter et al. 2003). Des capteurs piétons permettent de mesurer la teneur en chlorophylle des plantes ou le degré de maturité des fruits et peuvent remplacer des mesures longues et coûteuses qui sont réalisées en laboratoire. Malgré cela, la faible disponibilité de capteurs qui permettent de caractériser finement les sols et la végétation dans les champs limite aujourd'hui l'adoption de l'agriculture de précision (Oliver 2013a). Dans ce contexte, de nombreuses recherches sont effectuées pour proposer des capteurs ou des systèmes de mesures destinés à mieux caractériser l'état des végétaux, notamment leur état hydrique et nutritif ou pour détecter précocement les maladies (Koman et al. 2017 ; Oren et al. 2017). Des recherches sont aussi effectuées pour développer des capteurs permettant de mieux caractériser les sols ou la qualité de l'itinéraire technique (exemple : un capteur qui permet de mesurer la qualité de pulvérisation Bendiab et al. 2018). Ces recherches devraient permettre aux agriculteurs et aux scientifiques de faire un diagnostic plus précis des cultures et donc d'améliorer la prise de décision.

L'analyse et l'interprétation des données fait appel aux (géo)statistiques. Cela fait aussi appel aux Systèmes d'Information Géographiques (SIG) pour représenter les données et réaliser des pré-traitements (Leroux et al. 2018a). Si la capacité à interpoler des observations mono-variables et à quantifier leur variabilité spatiale à partir d'échantillonnages dans les parcelles ne pose plus de problème, des recherches sont maintenant effectuées pour réussir à intégrer la diversité des informations disponibles par l'agriculteur afin de prédire une variable d'intérêt agronomique ou préconiser une action (Goldstein et al. 2017). En effet, si la plupart des capteurs permettent de mesurer des effets sur le phénotype (perte de chlorophylle, évaluation de la biomasse, de la LAI, etc.), ils ne permettent souvent pas d'identifier la cause du phénomène observé, qu'elle soit hydrique, nutritive, météorologique ou biotique (Bégué et al. 2008). Il est nécessaire d'utiliser des informations complémentaires pour les identifier. D'après Hatfield and Kitchen (2013), l'agriculture de précision est pour l'instant dans une phase de diagnostic de la variabilité spatiale dans les parcelles et devra passer dans une phase de préconisation.

Bien sûr, le besoin de modulation et de gestion spatialement différenciées conduit à des recherches sur les équipements, pour mettre en œuvre des capteurs, effectuer mécaniquement des opérations précises et rapides, réaliser des algorithmes de contrôle et d'optimisation des opérations culturales (Saddem et al. 2017). Aujourd'hui, si la mécanisation des grandes cultures peut être très poussée du semis à la récolte, des opérations comme la récolte et la manipulation des fruits fragiles restent des problèmes qui sont toujours à l'état de recherche (Vasconez et al. 2019). Les solutions dans ce domaine concernent d'avantage l'assistance que le remplacement de l'homme (Bellon-Maurel and Huyghe 2016). Des recherches sont aussi effectuées sur les robots autonomes avec déjà quelques robots commercialisés pour le désherbage en grande culture, en maraîchage ou en viticulture (Bellon-Maurel and Huyghe 2016).

Des études en sciences de la gestion et en sociologie sont aussi réalisées pour étudier les taux d'adoption de l'agriculture de précision chez les agriculteurs, les freins à leur adoption et les retours économiques (Barnes et al. 2019; Paustian and Theuvsen 2016).

Enfin, l'agronomie est au coeur de la recherche en AP, aussi bien pour évaluer la pertinence des nouveaux capteurs que pour comprendre les phénomènes observés afin de prendre des décisions.

#### 2.1.4 Conditions d'application d'une agriculture de précision

Indépendamment des prérequis matériels, logiciels et humains (disponibilité des capteurs, capacité d'interprétation des données, disponibilité des machines et niveau de formation suffisant des exploitants), plusieurs conditions sont nécessaires pour que la gestion non uniforme d'une parcelle soit pertinente. D'une part, il faut que les conditions de croissance sur la parcelle entraînent une variabilité spatiale significative du comportement de la culture. Sans cela, la gestion uniforme est la plus simple, efficace et économique (McBratney and Taylor 2000). D'autre part, il faut que la variabilité spatiale soit stable dans le temps ou prévisible. Cette deuxième condition est vérifiée si la mesure, la décision et l'action sont réalisées dans un intervalle de temps très court. Par contre, cette condition est très contraignante lorsque les agriculteurs doivent prendre des décisions dont les effets ne seront visibles que plusieurs mois après ou, lorsque la mesure qui permet de décider est prise longtemps avant l'application modulée. C'est par exemple ce qui limite l'utilisation d'une carte de rendement de l'année pour faire des préconisations sur la fertilisation pour l'année (Ping and Dobermann 2005; Blackmore et al. 2003). Cette deuxième condition implique donc d'étudier les évolutions spatiales mais aussi temporelles des propriétés des cultures.

## 2.2 Les structures spatiales en agriculture de précision

### 2.2.1 Origine de la variabilité spatiale des cultures

Chaque plante d'un champ cultivé interagit avec son environnement à travers sa nutrition minérale, hydrique et son accès à la lumière. Ces facteurs affectent le développement des plantes et sont le plus souvent structurés dans l'espace. Ce constat a été fait depuis de nombreuses années par les agriculteurs qui

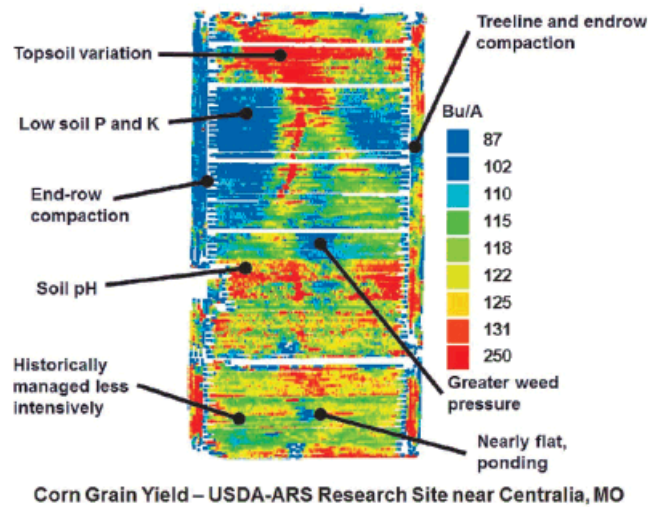


FIGURE 2.1 – Exemple de variabilité spatiale du rendement (exprimé en Bushels/Acre) dans un champ de maïs. Les facteurs qui expliquent les variations locales sont présentés par des flèches. (Source : Shannon 2018)

arpentent leurs parcelles (Smith 1938 ; Sebillotte 1969). La variabilité spatiale des facteurs environnementaux entraîne une variabilité du développement des cultures, avec des conséquences sur leur rendement et leur qualité. La théorie du facteur limitant considère que le rendement est limité par le premier élément qui vient à manquer (nutriment, lumière, chaleur, eau). Si on applique cette théorie à un contexte spatial on peut considérer que la variabilité spatiale du développement de la plante va correspondre à la variabilité spatiale du ou des facteurs limitants.

Certains de ces facteurs sont par nature difficilement contrôlables par l'agriculteur. C'est le cas de la radiation incidente, de la présence d'inoculum, de la profondeur des sols ou de la quantité de pluies. D'autres, par contre, peuvent être contrôlés comme la concentration d'éléments minéraux à travers la fertilisation, la résistance des plantes par le choix des variétés ou la compétition avec les adventices par le désherbage. L'agriculture de précision s'intéresse à la variabilité spatiale des propriétés du milieu et de la culture car une meilleure connaissance de cette variabilité doit permettre d'adapter l'itinéraire technique par zone (McBratney and Pringle 1999).

Les facteurs environnementaux peuvent être classés en plusieurs catégories :

Le sol est souvent considéré comme la caractéristique la plus importante pour prédire la structure spatiale des propriétés des cultures à l'échelle intra-parcellaire (Armstrong et al. 2009). En effet, les propriétés physiques et chimiques du sol conditionnent la nutrition hydrique et minérale des plantes. Les propriétés physiques telles que la profondeur du sol, sa compaction, sa texture vont conditionner la capacité d'enracinement et la réserve utile en eau. Les propriétés chimiques vont conditionner la disponibilité des éléments minéraux et la capacité des plantes à les extraire du sol. Enfin, même si cela a moins été étudié par la communauté AP, les organismes du sol sont aussi importants parce qu'ils parasitent ou, au contraire, rendent plus disponibles certaines substances (Dupuis and Whalen 2007 ; Doré et al. 2011).

Le micro-climat dans les parcelles, qui dépend de la topographie, du type de sol, ou encore de l'architecture des plantes va agir sur la photosynthèse mais aussi sur les risques de gelées ou de brûlures.

Les maladies et les ravageurs, ont aussi des effets structurés dans l'espace à cause des mécanismes de propagation dans les parcelles (la propagation va être influencée par le micro-climat liée à la densité de la canopée, à l'humidité du sol, aux vents dominants, etc.) et par une sensibilité différente des

plantes en fonction de leur vigueur, qui elle-même peut être variable en fonction des caractéristiques du milieu (sol, pente, etc.).

Enfin, les agriculteurs, même si les consignes des opérations cultures sont uniformes, provoquent des variabilités spatiales à cause de leurs pratiques. Ces variabilités peuvent être dues aux passages répétés des machines aux mêmes endroits des parcelles qui provoquent des tassements localisés, à des différences de réglage au sein des parcelles, ou encore à une irrigation ou une fertilisation non homogène.

La variabilité de la culture au sein d'une parcelle reflète les interactions multiples entre les différents facteurs de l'environnement et de l'itinéraire technique. En agriculture de précision, on considère que la variabilité aléatoire, non structurée dans l'espace ou a une très petite échelle spatiale, ne peut pas être utilisée pour moduler l'itinéraire technique, et qu'au contraire la variabilité structurée peut être exploitée pour moduler l'itinéraire technique (Pringle et al. 2003 ; Taylor et al. 2018). Comme nous l'avons vu dans la partie 2.1.4, certaines applications nécessitent aussi que la variabilité spatiale soit stable ou prévisible.

## 2.2.2 Stabilité temporelle de la variabilité spatiale

Les structures spatiales des facteurs qui affectent le développement des plantes sont plus ou moins stables dans le temps. Kaspar et al. (2003) séparent les facteurs environnementaux selon leur stabilité temporelle et différencient les facteurs permanents des facteurs transitoires que l'on peut aussi qualifier d'aléatoires. On considère que les propriétés physiques du sol (pente, profondeur des sols, texture) sont des composantes stables à l'échelle des cycles de production. Au contraire, la composition chimique est souvent considérée comme moins stable, à cause des flux hydriques dans les parcelles mais aussi à cause de la fertilisation, de l'utilisation des minéraux par les cultures et de la dégradation des résidus de culture. La structure spatiale des facteurs biotiques est souvent beaucoup plus aléatoire dans le temps que celle des facteurs abiotiques (Machado et al. 2000). L'évolution temporelle de la structure de ces facteurs va donc modifier la structure spatiale des propriétés des cultures.

Par ailleurs, d'autres facteurs, même s'ils peuvent être uniformes spatialement mais aléatoires dans le temps vont avoir une importance forte en interaction avec des facteurs structurés dans l'espace sur le développement des cultures. Par exemple, la pluviométrie peut être considérée comme uniforme spatialement sur une petite parcelle mais très aléatoire d'une année à l'autre. Cela va avoir une importance forte sur l'instabilité temporelle de la structure spatiale des cultures (Hatfield and Kitchen 2013 ; Maestrini and Basso 2018). D'une année à l'autre, les facteurs limitants dans la parcelle vont être différents. La structure spatiale de la culture pourra être très liée à celle de la capacité de rétention les années sèches, et plus liée à la disponibilité des éléments nutritifs les années avec une pluviométrie non limitante. Ainsi, plusieurs auteurs ont montrés des inversions de motifs spatiaux, appelé effet flip-flop, d'une saison à l'autre (Bégué et al. 2008 ; Leroux et al. 2018c ; Blasch and Taylor 2018) et une relation entre la structure spatiale de la culture et la pluviométrie (Kaspar et al. 2003 ; Hatfield 2012).

En conséquence, la variabilité temporelle du développement des cultures présentera une composante déterministe et répétable, liée à la stabilité de certains facteurs et une composante aléatoire liée à des facteurs plus transitoires qui peuvent être prévisibles, ou non. L'enjeu des observations en AP est de pouvoir caractériser la variabilité spatiale pour identifier celle qui peut être gérée, notamment en différenciant la variabilité aléatoire spatialement de la variabilité structurée et la part de la variabilité qui est stable ou répétable dans le temps.

## 2.2.3 Étude et modélisation des structures spatiales

### Approche générale

Il est possible d'étudier la structure spatiale des facteurs de l'environnement (sol, topographie, etc.) ou de la production (biomasse, rendement, etc.) dans l'objectif d'identifier des zones agronomiques différentes afin de différencier l'itinéraire technique localement.

Lorsque seules les variables de l'environnement sont étudiées, des connaissances et des hypothèses sont nécessaires pour établir les conséquences sur la production finale (qualité et quantité). Les effets de ces facteurs sur la production peuvent être prédits à partir de modèles de connaissances, experts, empiriques ou même mécanistes (cf. Cook and Bramley (1998) pour une liste des modèles qui peuvent être utilisés en AP et des exemples). Le risque de ces approches est de trop simplifier les relations qui existent entre les facteurs mesurés et la production et finalement de manquer certains facteurs ou interactions (Diker et al. 2004). La mesure du rendement ou du développement des cultures (par télédétection notamment) est un moyen simple d'acquérir des informations sur l'effet intégré des interactions entre les facteurs qui influencent le développement des cultures (Diker et al. 2004). Ces mesures sont donc très complémentaires des mesures réalisées sur les facteurs de l'environnement et permettent de produire des cartes interprétables comme celle présentée Figure 2.1.

L'approche qui semble la plus pertinente aujourd'hui pour appliquer une agriculture de précision est une approche qui étudie la stabilité temporelle de la réponse des cultures sur plusieurs années et qui identifie les zones stables et les zones instables dans les parcelles. Grâce à la connaissance des facteurs stables du milieu (topographie, type de sol, etc.) cela permet de définir une stratégie globale de modulation spatiale qui est ensuite adaptée au cours de la période culturale en fonction de la mesure des facteurs instables et non prévisibles comme la pluviométrie (Maestrini and Basso 2018 ; Basso et al. 2011).

### **Théorie des variables aléatoires régionalisées**

En pratique, la modélisation des structures spatiales fait appel à la notion de variable aléatoire régionalisée, telle que décrite par (Matheron 1966). Nous détaillons cette notion ainsi que, plus loin dans le texte, le variogramme car ces outils sont utilisés tout au long de la thèse.

Une variable régionalisée est une variable qui a les propriétés suivantes :

- Elle prend ses valeurs dans un domaine défini de l'espace, par exemple le contour géométrique d'une parcelle

- Elle a un support géométrique d'observation, noté  $S$ , qui peut être ponctuel, surfacique (par exemple la forme de la surface dans laquelle des épis de blés sont comptés) ou volumique (par exemple le volume de sol prélevé). Cette notion de support est importante car les dimensions de celui-ci vont modifier les caractéristiques de la variable mesurée, notamment sa variance.

- Elle possède un caractère irrégulier, présenté comme aléatoire ...

- ... mais aussi des caractéristiques structurées dans l'espace qui reflètent des phénomènes d'intérêts pour l'étude.

Une variable aléatoire régionalisée peut être représentée par le modèle ci dessous (Oliver and Webster 2014) qui est illustré Figure 2.2 :

Avec  $\mu$  l'espérance de la variable  $Z(x)$  qui ne dépend pas de sa position et  $\epsilon(x)$  une variable aléatoire qui est structurée spatialement, de moyenne 0 et de variance  $\sigma^2$ . L'expression  $\Sigma(x, y)$  représente la matrice de variance covariance entre chaque paire d'observations.

Les analyses de cette variable doivent permettre de déterminer les caractéristiques de sa structure spatiale, son caractère aléatoire, les phénomènes à l'origine de ses caractéristiques, sa stabilité temporelle, et la possibilité de contrôler cette variable ou d'en tirer parti.

Le variogramme a été développé historiquement dans le cadre de la prospection minière. Depuis, cet outil a été très souvent utilisé en AP pour étudier les variables aléatoires régionalisées même si d'autres outils sont disponibles comme l'indice de Moran ou le corrélogramme (Mihail 1989 ; Shaukat and Siddiqui 2004 ; Wyse-Pester et al. 2002). Le principe du variogramme, noté  $\gamma(h)$  est d'étudier la variabilité ou « dissemblance » entre des observations effectuées à une distance  $h$  les unes des autres.

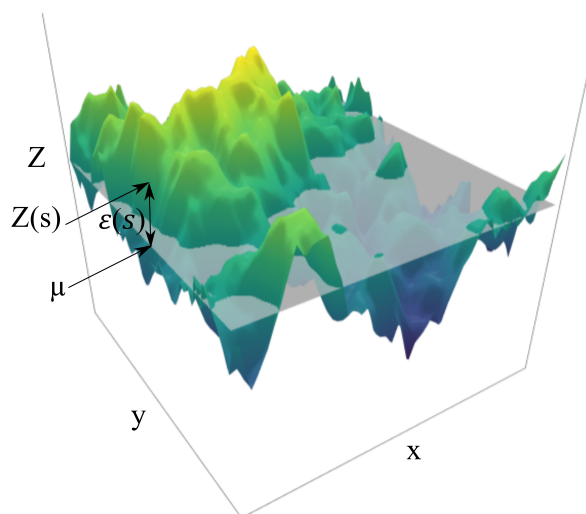


FIGURE 2.2 – Représentation de la variable régionalisée  $Z$ , de son espérance  $\mu$  (surface plane) et de

avec  $\gamma(h)$  le variogramme de  $Z$  qui ne dépend pas de la position  $s$  mais seulement de la distance entre les observations<sup>1</sup>. On s'attend à ce que des observations proches spatialement se ressemblent et que plus la distance qui les sépare augmente moins elles se ressemblent, jusqu'à une certaine distance qui correspond à la portée de la structure spatiale. Passée cette distance, le variogramme atteint une valeur appelée le palier (sill)<sup>2</sup> qui correspond à la variance de la variable ( $\sigma^2$ ). Si les observations réalisées presque au même endroit sont identiques, dans ce cas le variogramme aura une valeur nulle pour une distance  $h$  qui tend vers 0. Les observations réalisées très proches peuvent aussi être un peu différentes, soit à cause de l'erreur de mesure, soit parce qu'il y a des variations à petite échelle. Dans ce cas, le variogramme ne passera pas par 0 mais par une valeur appelée la variance pépète (nugget). L'origine de ce mot vient des gisements aurifères. Si deux mesures sont réalisées à une distance très proche l'une de l'autre mais que la première est réalisée sur une pépète et que la deuxième est réalisée juste à côté mais plus sur la pépète alors les teneurs mesurées peuvent être très différentes. La relation entre la covariance et le variogramme s'écrit<sup>2</sup> :

avec  $C(h)$  la fonction de covariance et  $\sigma^2$  qui correspond à la variance de la variable  $Z$ .

Le variogramme permet donc d'étudier des structures spatiales et de les caractériser. Les paramètres qui permettent de décrire le variogramme expérimental, aussi appelé variogramme empirique, sont obtenues par l'ajustement d'un modèle de variogramme théorique. Ce variogramme théorique peut être l'une des fonctions présentées Figure 2.4 où la somme de plusieurs de ces fonctions. On dit à ce moment là que le variogramme est un variogramme Gignone (nested variogramme). L'intérêt du variogramme provient aussi de son utilisation pour interpoler des mesures grâce à la méthode appelée Krigeage. En effet, lorsque l'on veut estimer une valeur à un endroit de l'espace qui n'a pas été échantillonné, le variogramme permet de savoir quelle pondération donner aux observations situées à côté. Cela permet de faire des estimations non biaisées ce qui n'est pas le cas d'autres méthodes telles que l'interpolation linéaire ou la pondération inverse à la distance.

1. Sous l'hypothèse où  $Z(s)$  est une fonction aléatoire intrinsèque. Voir Oliver and Webster (2014)

2. Sous l'hypothèse de stationnarité d'ordre 2 de  $Z(s)$ . Voir Oliver and Webster (2014)

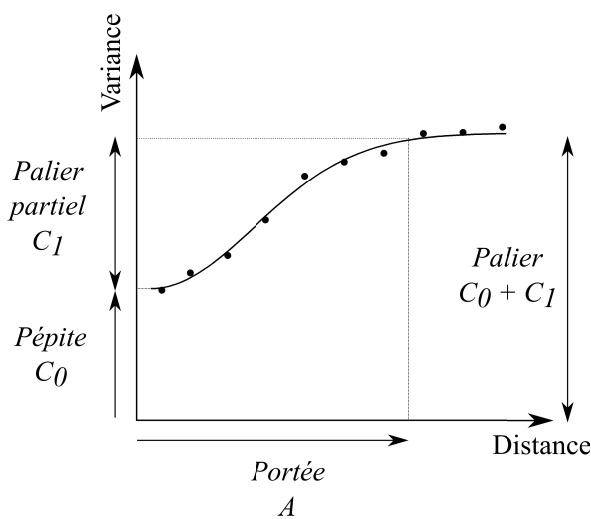


FIGURE 2.3 – Exemple de variogramme empirique (points) et de variogramme théorique (ligne). La variance pépite correspond à la variance non structurée spatialement, le palier partiel correspond à la variance structurée spatialement et le palier correspond à la variance totale. La portée correspond à la distance à partir de laquelle il n'y a plus d'auto-corrélation spatiale entre les données.

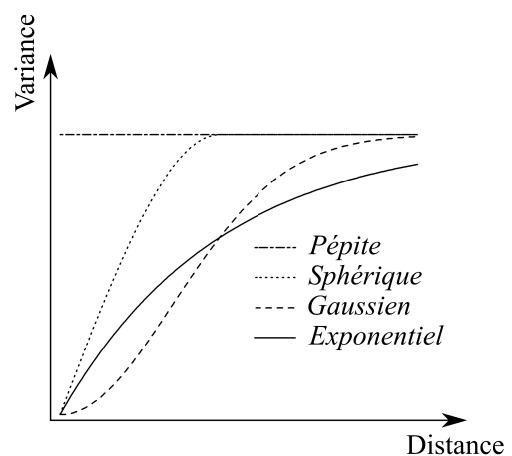


FIGURE 2.4 – Exemples de variogrammes théoriques. Chaque variogramme caractérise des structures spatiales avec des traits structuraux différents. Par exemple, le variogramme pépite correspond à un bruit blanc (absence de structure spatiale), le variogramme gaussien décrit une structure avec une grande régularité spatiale qui apparaît « lisse », le variogramme exponentiel décrit une structure beaucoup plus irrégulière. La somme de variogrammes théoriques permet de modéliser des structures spatiales plus complexes. Par exemple, le variogramme théorique de la Figure 2.3 est la somme d'un variogramme pépite et d'un variogramme gaussien.

## 2.3 État de l'art de l'agriculture de précision sur la banane

L'état de l'art qui vient d'être présenté a été réalisé à partir de l'étude des articles publiés en AP sur de nombreuses cultures différentes de la banane. Je présente maintenant spécifiquement les études qui ont été publiées dans ce domaine de recherche sur la banane.

### 2.3.1 Bibliographie

Peu d'articles traitant de l'usage de l'agriculture de précision en bananeraies ont été publiés. Ils sont présentés succinctement ci-dessous.

J. Stoorvogel a travaillé sur plusieurs projets au Costa Rica pour introduire l'agriculture de précision dans les bananeraies afin de diminuer l'impact environnemental de cette culture et pour augmenter les rendements. Il a notamment publié des articles autour d'un système d'aide à la décision, nommé BanMan, qui doit permettre aux agriculteurs de faire un diagnostic des conditions de production dans la plantation à partir de l'enregistrement géolocalisé des rendements (poids des régimes) et d'une cartographie des sols (Stoorvogel 1998 ; Stoorvogel et al. 2015). Il a montré que le poids des régimes était variable au sein de la plantation et que les conditions de sol ne permettaient pas d'interpréter les cartes de poids des régimes à cause d'autres facteurs tels que les maladies, la qualité du drainage et la fertilité des sols. Il a donc



choisi de représenter l'écart entre le rendement mesuré par site et le rendement attendu dans la zone de sol correspondante et d'observer ensuite au terrain les raisons susceptibles d'expliquer les anomalies. Plusieurs types de mesures correctives ont pu être prises comme l'application localisée de fertilisants ou de pesticides ou même la re-plantation de certaines surfaces pour les endroits où les poids des régimes étaient faibles.

Freitas et al. (2015) ont utilisés les géostatistiques pour identifier la structure spatiale de la sévérité de la maladie fongique induite par la cercosporiose jaune sur les bananiers au sein d'une parcelle. L'objectif de l'étude était de mettre en évidence les corrélations spatiales entre des variables de la parcelle (altitude, propriétés du sol, etc.), des variables de l'état nutritionnel des bananiers (analyses chimiques des macro et micro nutriments) et l'intensité de l'infection. Les mesures de la sévérité de l'infection ont été réalisées tous les mois sur des pieds choisis initialement au même stade phénologique. Cette étude a montré une structure spatiale dans la parcelle pour tous les paramètres étudiés et une corrélation spatiale entre les facteurs du milieu et l'infection.

On peut aussi trouver une étude de faisabilité de l'utilisation des drones dans les bananeraies comme source d'information auxiliaire pour le diagnostic de la production. Machovina et al. (2017) ont mesuré des composantes du rendement comme le poids du régime, le nombre de mains, la longueur du doigt le plus gros du régime pendant 76 semaines sur 23 surfaces d'environ 1 à 3 ha. Pour chacune des 23 surfaces, le sol a été classifié en 4 classes selon leurs caractéristiques physiques (texture, drainage, structure, etc.) et l'indice NDVI obtenu par drone. Ils ont trouvé une corrélation positive entre les composantes du rendement et le NDVI (  $r = 0,45$  pour la relation entre le NDVI et le poids moyen ;  $r = 0,38$  pour la relation entre le NDVI et le nombre de mains et  $r = 0,42$  pour la relation entre le NDVI et la longueur du doigt le plus large). Ils n'ont pas trouvé de relation entre le NDVI et les classes de sol. Dans l'étude, l'âge des parcelles et l'hétérogénéité des stades phénologiques au sein de celles-ci n'étaient pas indiqués.

On remarque que, dans l'ensemble des travaux que nous venons de citer, l'effet de l'asynchronisme n'a pas été mentionné. L'effet de la diversité des stades phénologiques sur le NDVI, l'effet temporel du moment de la récolte sur le poids des régimes n'a pas été étudié.

Nous avons cherché d'autres cultures qui seraient asynchrones comme le bananier pour étendre l'état de l'art. Même si certaines cultures présentent des caractéristiques similaires, nous n'avons pas trouvé d'autres cultures asynchrones. Par exemple, l'ananas est aussi une plante monocarpique tropicale qui produit des rejets et dont la floraison dépend de différents facteurs dont la vigueur du plant. Par contre, sa floraison est synchronisée dans les champs par l'application d'hormone. La canne à sucre est aussi monocarpique et se reproduit par multiplication végétative mais la récolte des champs se fait simultanément pour toute la parcelle un peu avant la floraison ce qui synchronise le développement.

### 2.3.2 Particularités de la culture bananière par rapport aux cultures sur lesquelles s'est développée l'agriculture de précision

Classiquement l'agriculture de précision vise à réduire l'échelle de travail à des zones de gestion intra-parcellaires. Dans la culture bananière, chaque plant de bananier est déjà géré individuellement et le plus souvent manuellement. C'est donc une agriculture extrêmement précise en terme d'échelle de traitement des plants. Au sein d'une parcelle, deux bananiers voisins n'ont pas le même traitement à une date donnée s'ils n'ont pas le même stade phénologique. Par contre, les règles de décisions utilisées pour chaque stade phénologique sont uniformes dans les plantations. L'enjeu de l'agriculture de précision en bananeraie serait donc d'identifier des zones homogènes agronomiques et d'adapter les règles de décision à chacune de ces zones.

Comme nous l'avons vu dans le Chapitre 1, la culture bananière est essentiellement manuelle. Seul le travail de la parcelle avant sa mise en place, les traitements aériens et le transport des régimes des parcelles aux stations de conditionnement sont mécanisés. Cela a des conséquences sur la prise d'observations. L'AP a été rendue possible par la plus grande disponibilité des systèmes automatisés de mesure. Dans le cas de la production bananière, l'essentiel des observations sont des observations humaines avec, en conséquence, une localisation des observations souvent moins précise, à l'échelle de la parcelle ou de la ferme. La faible

mécanisation a aussi des conséquences sur les modes d'application et de modulation. L'AP dans les bananeraies ne peut pas s'appuyer, du moins à court terme, sur les machines pour la modulation. On note par ailleurs que sans re-conception du système de production, le travail des employés ne pourra pas être substitué par des machines, les systèmes de drainage et d'haubanage des bananiers rendant très difficile leur circulation et les gestes demandant beaucoup de savoir-faire (notamment pour la sélection du rejet). Ces raisons expliquent pourquoi Stoorvogel et al. (2015) défendent une agriculture de précision « low/medium Tech » pour cette culture.

Stoorvogel et al. (2015) expliquent aussi qu'il existe moins de modèles mécanistes et de connaissances sur cette culture que pour d'autres cultures occidentales. Cela diminue donc la possibilité d'assister la prise de décision des producteurs dans une démarche d'AP (étapes 3 et 4 de la démarche présentée Section 2.1.2). On observe cependant qu'un grand nombre de modèles ont été développés dans des cadres de recherche scientifique et qu'ils pourraient être adaptés en outils d'aide à la décision. Parmi ces modèles, on peut citer le modèle STICS qui a été adapté au bananier pour étudier l'effet du travail du sol et de l'irrigation sur le premier cycle de floraison (Brisson et al. 1998), le modèle SIMBA (Tixier et al. 2008) produit par le CIRAD pour simuler et comparer des systèmes de culture (Blazy et al. 2009), l'adaptation du modèle CENTURY (Parton 1996) au cas de la culture bananière (Woomer et al. 1998), le modèle de Dens et al. (2008) pour modéliser la croissance des bananiers. Des modèles écophysologiques ont aussi été développés pour modéliser l'assimilation carbonée et la transpiration du bananier (Eckstein and Robinson 1995; Thomas et al. 1998; Siles et al. 2011) mais aussi pour modéliser l'effet des forces sources-puits sur le poids des régimes (Bugaud et al. 2011). Nous pouvons aussi citer le modèle de simulation spatiale de l'épidémiologie du charançon du bananier *Cosmopolites sordidus*, un ravageur important des champs de bananiers (Vinatier et al. 2009).

### 2.3.3 Les difficultés liées à l'asynchronisme pour l'étude de la variabilité spatiale

En plus des particularités structurelles du système de culture, le bananier présente la particularité d'être asynchrone. Cette particularité est critique car l'ensemble des méthodes d'analyse spatiale, indispensables pour identifier l'opportunité de la modulation, ont été développées sur des cultures synchrones.

Classiquement en AP la production de chaque année est analysée comme si c'était une expérience d'homogénéité (Pringle et al. 2003) (Uniformity trial) qui permet de révéler l'effet des conditions variables de l'environnement. Cela est possible dans les cultures annuelles comme le blé parce que chaque année peut-être perçue comme la répétition d'une expérience, avec, certes, des conditions légèrement différentes à cause des facteurs non stables du milieu. Dans les cultures pérennes telles que la vigne, le système peut aussi être perçu comme la répétition d'une expérience avec des conditions réinitialisées chaque année après le passage de l'hiver. En revanche, au contraire des cultures annuelles, les vignes restent en place plusieurs années et stockent des réserves pendant l'hiver ce qui fait que le comportement de l'année dépend de celui de l'année précédente. C'est ce qui semble favoriser la stabilité temporelle de cette culture (Tisseyre et al. 2008).

Dans ces deux cas, chaque année, le développement des plantes est réinitialisé lors du semis, de la plantation ou du redémarrage de la végétation. Au cours de l'année et à un moment donné, les propriétés des cultures peuvent être comparées car elles commencent à se développer en même temps et que les différences spatiales sont expliquées par des conditions différentes de croissance (Figure 2.5a). Dans le cas de la production bananière, le premier cycle de production, c'est-à-dire depuis la mise en place de la parcelle jusqu'à la récolte du premier plant, est similaire à celui d'une culture annuelle. Par contre, dès la récolte du premier régime et à cause de l'asynchronisme, les pieds issus de rejets ne sont plus au même stade de développement. En conséquence, les différences visibles entre des bananiers situés à différents endroits de la parcelle ne sont plus seulement expliquées par des différences de conditions de croissance mais aussi par des différences de stades phénologiques (Figure 2.5b).

Nous nous attendons donc à ce que l'étude d'une propriété du couvert ne mette pas seulement en

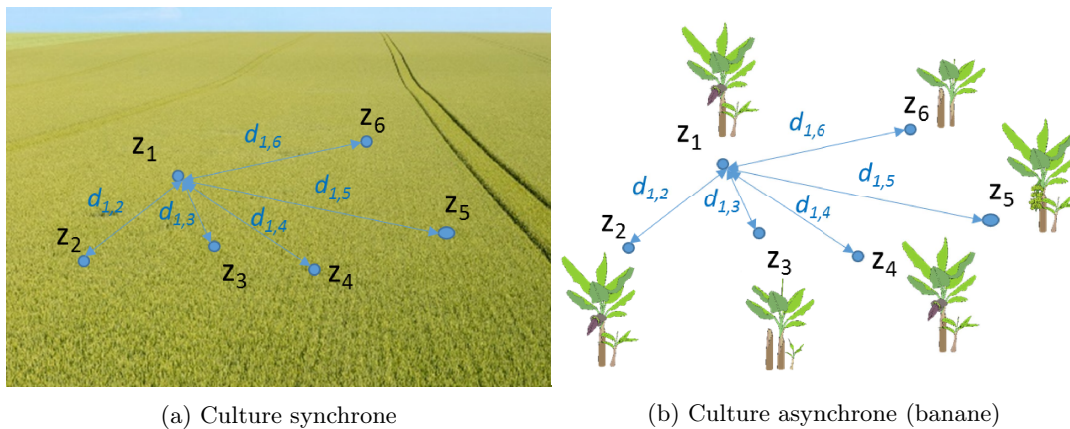


FIGURE 2.5 – Représentation du problème posé par l'asynchronisme pour analyser la variabilité spatiale d'observations faites sur le couvert végétal . a) Les observations sont comparables et reflètent la variabilité des conditions de croissance dans la parcelle. La variabilité spatiale est analysée en comparant la variance de chaque paire de points en fonction de la distance qui les sépare . b) Les observations ne sont plus directement comparables, elles reflètent la variabilité des conditions de croissance dans la parcelle mais aussi la variabilité des stades phénologiques.

évidence l'effet des conditions variables de croissance dans la parcelle mais aussi l'effet des stades phénologiques. Les conséquences attendues sont :

- Un effet du stade phénologique moyen de la parcelle sur la moyenne de la propriété ;
- Un effet sur la variance de la propriété liée à la diversité des stades phénologiques ;
- Un effet sur la structure spatiale de la propriété.

Dans le cas où les mesures seraient réalisées sur un même stade phénologique mais à plusieurs moments, nous nous attendons aussi à des effets similaires. Ces derniers seraient dus aux modifications temporelles des pratiques culturales et de la météorologie qui modifieraient les propriétés mesurées.

## 2.4 Objectifs scientifiques de la thèse

Plusieurs questions scientifiques découlent de cet état de l'art et permettront de répondre à la problématique générale : Est-il possible d'appliquer une approche d'agriculture de précision à la culture bananière malgré l'asynchronisme ?

Les premières questions auxquelles nous avons voulu répondre concernent l'asynchronisme :

- Comment estimer la diversité des stades phénologiques et le stade phénologique moyen des parcelles ?
- A quelle vitesse se met en place la diversité des stades phénologiques dans les parcelles ?
- Est ce que la vitesse dépend des conditions de culture de la parcelle ?

Les questions suivantes concernent la structure spatiale des propriétés des bananeraies sachant que nous avons vu que la modulation n'est pertinente que s'il y a une structure spatiale des propriétés.

- Les propriétés végétatives et reproductives des bananiers sont elles variables dans l'espace ?
- La mesure de la structure spatiale est elle biaisée par l'asynchronisme ?
- Est-il possible de corriger les effets de l'asynchronisme et comment ?

Les données utilisées seront principalement celles qui servent à la gestion des récoltes (observation des floraisons et observation de la maturité des régimes). Cela permettra de répondre plus généralement à la

question : Est ce que les données recueillies par les producteurs pour des besoins de gestion peuvent être utilisées pour faire un diagnostic de la production, c'est-à-dire pour identifier les facteurs qui en affectent le développement ?



## Chapitre 3

# Caractérisation de l'asynchronisme avec les données de production

Le moteur de l'asynchronisme dans la production bananière est la variabilité de la vitesse de développement des plants qui est due à une variabilité naturelle entre les plants, à une compétition entre eux pour les ressources (nutriments, lumière, eau) mais aussi à des différences culturelles résultant de l'action des producteurs (moment de récolte des régimes, méthode de sélection du rejet, etc.) (Tixier et al. 2004). L'asynchronisme s'explique aussi par l'absence de deux phénomènes de synchronisation qui sont souvent présents sur d'autres cultures : (i) la phénologie du bananier n'est pas synchronisée par des variations saisonnières du milieu (photopériode, température) ce qui fait que les floraisons puis les récoltes sont réparties dans le temps, (ii) le début des cycles culturels suivant chaque récolte n'est pas synchrone, car l'apparition des rejets sur les rhizomes est continue et parce que la sélection du rejet-fils par les producteurs n'est pas synchrone.

La première section de ce chapitre étudie la variabilité de la durée des cycles de développement des bananiers à l'échelle d'une plantation. Cette étude a fait l'objet d'un article soumis à *Field Crops Research*, et c'est la version préalable en Français qui est incluse dans ce mémoire. La deuxième section étudie les variabilités des temps de croissance des bananiers à l'échelle intra-parcellaire. La troisième section propose des indices permettant d'évaluer la diversité des stades phénologiques et le stade phénologique moyen des parcelles. Enfin, la dernière section étudie comment évolue la structure spatiale des stades phénologiques au cours du temps. Ces études sont réalisées à partir de la base de données des floraisons de la Compagnie Fruitière.

L'ensemble de ce chapitre s'appuie sur la modélisation probabiliste du cycle de développement des bananiers qui a été réalisée avec Gilles Le Moguédec (INRA). Ce modèle a été construit lors d'une prestation de recherche commanditée par la Compagnie Fruitière auprès de l'INRA et dont j'ai été l'initiateur avant de commencer la thèse. Cette prestation a abouti à la création d'un package R (R Core Team 2018) que nous avons appelé *PreviFlo*, à une note de travail (Le Moguédec 2019) et à des TD pour se former à l'utilisation du modèle. L'ensemble de ces documents ne sont pas présentés dans le mémoire. Les équations qui permettent de modéliser les floraisons des bananiers d'une parcelle sont utilisées dans ce chapitre mais ne constituent qu'une petite partie des équations proposées par Gilles Le Moguédec pour prévoir les floraisons connaissant l'historique des données de floraisons depuis le début de la plantation d'une parcelle ou alors depuis une fenêtre d'observation plus restreinte. Ce chapitre présente une utilisation différente du modèle que celle de la prévision pour, dans ce cas, formaliser ce qu'est l'asynchronisme et analyser les courbes de floraisons des parcelles pour faire un état des lieux de la variabilité de la durée des cycles de floraison.

## Sommaire

---

<b>3.1 Variabilité de la durée moyenne des cycles de développement des parcelles à l'échelle d'une plantation . . . . .</b>	<b>37</b>
3.1.1 Résumé . . . . .	37
3.1.2 Introduction . . . . .	37
3.1.3 Méthodes . . . . .	39
3.1.4 Paramétrisation et qualité d'ajustement du modèle . . . . .	43
3.1.5 Résultats . . . . .	47
3.1.6 Discussion . . . . .	53
3.1.7 Conclusion . . . . .	56
<b>3.2 Variabilité de la durée moyenne des cycles de développement à l'échelle intra-parcellaire . . . . .</b>	<b>57</b>
3.2.1 Introduction . . . . .	57
3.2.2 Matériels et Méthodes . . . . .	58
3.2.3 Résultats et discussion . . . . .	60
3.2.4 Conclusion . . . . .	66
<b>3.3 Formalisation de l'asynchronisme . . . . .</b>	<b>66</b>
3.3.1 Introduction . . . . .	66
3.3.2 Définition des indices . . . . .	66
3.3.3 Matériel et Méthodes . . . . .	67
3.3.4 Résultats et discussion . . . . .	69
<b>3.4 Évolution de la structure spatiale des stades phénologiques au cours du temps . . . . .</b>	<b>71</b>
3.4.1 Introduction . . . . .	71
3.4.2 Analyse de la structure spatiale de données circulaires . . . . .	74
3.4.3 Matériel et méthodes . . . . .	74
3.4.4 Résultats . . . . .	75
3.4.5 Discussion et conclusion . . . . .	75

---

## 3.1 Variabilité de la durée moyenne des cycles de développement des parcelles à l'échelle d'une plantation

### 3.1.1 Résumé

L'utilisation des données produites par les agriculteurs pour générer des connaissances et moduler les décisions de production est l'un des objectifs de l'agriculture de précision. Dans les bananeraies industrielles, les observations du nombre de floraisons par parcelle sont systématiquement enregistrées pour gérer les récoltes. L'utilisation de ces données pour évaluer l'impact de l'environnement et des pratiques des agriculteurs sur la production serait importante pour améliorer les décisions des producteurs de bananes. La durée moyenne entre deux floraisons d'un bananier, appelée *cycle de floraison*, est variable et cette production est dite asynchrone par opposition aux cultures dont la phénologie est synchronisée par les saisons. Le *cycle de floraison* est une propriété intéressante pour comprendre les rendements, qui est longue et coûteuse à étudier par des mesures dans les parcelles. Une analyse a été réalisée à partir des données de floraison de 724 champs situés dans une plantation commerciale camerounaise pour évaluer l'effet des conditions de culture et des choix des agriculteurs sur cette propriété. Un modèle basé sur la description de la séquence de développement du bananier a été construit et ajusté aux données de floraison de chaque parcelle. Cette procédure a permis d'estimer le *cycle de floraison* à partir des données de floraison. Les résultats ont montré que le *cycle de floraison* était variable à l'intérieur de la plantation (moyenne de 209 jours et écart-type de 24 jours) et spatialement structuré (rapport d'auto-corrélation de 0.64). L'effet des choix de production (irrigation, cultivar) et des conditions environnementales (températures, altitude) sur le *cycle de floraison* a été étudié. La prise en compte d'une échelle de temps en degrés-jours par rapport à une échelle en jours juliens n'a pas amélioré l'ajustement du modèle ni l'estimation de *cycle de floraison*. Il a également été démontré que l'altitude, le cultivar et l'irrigation avaient un effet sur *cycle de floraison*. Pour les producteurs, cela souligne l'importance de prendre en compte cette propriété lors de l'analyse des rendements et de ne pas seulement étudier le poids des régimes à la récolte. La forte auto-corrélation spatiale de *cycle de floraison* offre également des perspectives intéressantes pour moduler les décisions des agriculteurs par zone.

### 3.1.2 Introduction

Depuis les années 1990, la disponibilité de systèmes de mesure permettant d'acquérir des données géoréférencées à haute résolution spatiale a fait émerger le concept d'agriculture de précision (AP). L'AP a pour objectif d'utiliser la variabilité mise en évidence par l'étude des données des producteurs pour augmenter la performance des exploitations en adaptant l'itinéraire technique par zone (Basso et al. 2016). Cette pratique appelée Site Specific Management (SSM) doit permettre d'optimiser l'utilisation des intrants en fonction des besoins variables des cultures et ainsi de diminuer les charges économiques et l'impact négatif de la sur-utilisation d'intrants (Balafoutis et al. 2017).

Le SSM a été étudié initialement sur les grandes cultures pour lesquelles les données de rendements étaient disponibles. Mais de nombreuses autres cultures ont ensuite été étudiées. Parmi elles, le coton (Bronson et al. 2006), les kiwis (Taylor et al. 2007), la canne à sucre (Singh et al. 2006), la vigne (Tisseyre et al. 2008), etc. A notre connaissance un seul article (Stoorvogel et al. 2000) évalue la possibilité de faire du SSM en banane. Cette stratégie de gestion serait pourtant intéressante pour réduire la quantité d'intrants utilisés dans la culture bananière destinée à l'export. Ce type de production repose en effet sur des plantations industrielles qui s'étendent sur de grandes surfaces et qui sont le plus souvent monovariétales et intensives en intrants (Ploetz and Evans 2015). Les bénéfices environnementaux qui peuvent être attendus par une utilisation plus précise des intrants sont très forts sachant que la banane représente une production importante en volume au niveau mondial. Elle fait en effet partie des fruits et légumes les plus exportés, devant les pommes de terres, les pommes, ou encore les tomates (FAOSTAT 2016).

La banane présente des particularités qui doivent être prises en compte pour étudier les rendements et en faire un diagnostic. Premièrement, le bananier est une plante monocarpique qui ne produit qu'une



seule inflorescence et donc un seul régime de bananes. Deuxièmement, la production bananière est semi-pérenne. « Pérenne » car les plants de bananiers qui sont mis en place dans les parcelles produisent un rhizome qui reste indéfiniment en place, « semi » car chaque bananier produit par le rhizome ne vit que pendant une durée déterminée, jusqu'au mûrissement de son régime. Dans les bananeraies, les producteurs contrôlent le nombre de rejets qu'ils laissent se développer simultanément sur chaque rhizome de façon à ce que seulement deux bananiers coexistent, le bananier parent et le bananier fils. Le bananier fils assure la pérennité de la production après la récolte du pied parent. Troisièmement, les bananiers d'une parcelle se développent à leur propre rythme et ne sont pas synchrones (Tixier et al. 2004). L'étendue temporelle de la première floraison d'une parcelle de bananiers est assez courte puis au fur et à mesure des cycles de floraison, cet intervalle augmente jusqu'à ce que la floraison devienne continue. Le moment des récoltes suit la même allure générale avec seulement un décalage d'une dizaine de semaines.

L'asynchronicité du développement des pieds de bananiers s'explique par l'absence de deux phénomènes de synchronisation qui sont souvent présents sur d'autres cultures : (i) la phénologie du bananier et notamment sa floraison n'est pas synchronisée par des variations saisonnières du milieu (photopériode, température) ce qui fait que les floraisons puis les récoltes sont réparties dans le temps. (ii) le début des cycles culturaux suivant chaque récolte n'est pas synchrone, car l'apparition des rejets sur les rhizomes est continue et parce que la sélection du rejet fils par les producteurs n'est pas synchrone.

Dans la plupart des cultures saisonnières sur laquelle s'est développée l'AP les périodes de récoltes sont constantes ce qui fait que la durée d'un cycle de production n'est pas étudiée. Dans le cas de la production bananière la durée du cycle (CD) dépend de plusieurs facteurs de l'environnement mais aussi de l'itinéraire technique. Tixier et al. (2004) ont montré que la méthode de sélection du rejet a un effet direct sur  $CD$  et qu'il est possible par ce moyen de contrôler le moment de la récolte. La disponibilité de la lumière qui dépend à la fois de la densité de plantation et de la taille des bananiers a un effet sur la croissance des pieds et sur  $CD$  (Eckstein and Robinson 2000; Israeli et al. 1995). A titre d'exemple, l'effeuillage du pied parent juste avant la récolte permet au rejet de croître plus rapidement. Cottin et al. (1987) ont aussi montré un effet de l'altitude. La durée entre la plantation des parcelles situées en altitude et le moment de leurs premières floraisons est plus longue que pour des parcelles plus près du niveau de la mer.

Au sein d'une plantation commerciale, les facteurs du milieu et de l'itinéraire technique sont donc susceptibles de varier et de modifier le rendement. Il est donc important pour les producteurs d'avoir des outils leur permettant de mesurer cela pour comprendre les causes de variabilité du rendement et adapter localement leurs itinéraires techniques. L'approche proposée dans cet article vise à évaluer cette composante du rendement et à analyser ces sources de variabilité à l'échelle d'une plantation commerciale en utilisant les données de production. On suppose par ailleurs que la variabilité du poids des régimes peut être étudiée en suivant la méthode proposée par Stoorvogel et al. (2000). Ces deux types d'informations (le poids moyen et le  $CV$ ) sont complémentaires pour expliquer le rendement et toutes les deux pertinentes car la relation entre le rendement et le poids des régimes est complexe. Dans certains cas ces deux composantes sont corrélées négativement (le rendement diminue et le poids augmente), dans d'autres positivement ce qui entraîne un effet antagoniste sur le rendement (Dorel et al. 2016).

La mesure du rendement en production est longue et coûteuse. Elle nécessite de suivre individuellement plusieurs plants pendant plusieurs cycles de floraisons pour à la fois estimer fiablement le rendement de chaque parcelle, mais aussi pour estimer la variabilité sur l'ensemble des parcelles. Cottin et al. (1987) ont montré qu'il était possible d'utiliser les données des producteurs pour évaluer la vitesse de croissance de chaque parcelle. Les données qu'ils ont utilisées sont les observations de floraison, qui sont réalisées systématiquement depuis la plantation des parcelles jusqu'à leur mise en jachère pour gérer les récoltes. Le modèle qu'ils ont proposé est un modèle de forme qui permet de modéliser l'allure de la courbe de floraison de chaque parcelle. Cette approche est intéressante car elle permet aux producteurs d'estimer à moindre coût quelques caractéristiques de leurs parcelles. Par contre, ce modèle ne permet pas d'estimer directement le rendement et nécessite un nombre de paramètre qui augmente proportionnellement avec le nombre de cycles de floraison ce qui rend cet outil difficilement utilisable en production. Le modèle SIMBA

proposé par Tixier et al. (2008), et particulièrement le module SIMBA-POP (Tixier et al. 2004) permet de prendre en compte le fonctionnement des cycles du bananier et notamment l'asynchronisme. Cependant, ce modèle a été créé pour concevoir des systèmes de culture par simulation, et est moins adapté pour une caractérisation statistique des parcelles à partir des données de production. Il est donc nécessaire de proposer une nouvelle approche qui permette à partir des données de comptage de floraison et d'un modèle statistique d'estimer pour chaque parcelle d'une plantation.

Cette étude a deux objectifs : (i) concevoir un modèle probabiliste qui permet d'estimer à partir des observations de floraison réalisées en production et (ii) d'étudier la variabilité de et d'en expliquer les origines.

### 3.1.3 Méthodes

#### Présentation du modèle

##### Représentation du cycle de développement des bananiers

Le cycle de développement des bananiers pendant une période de production (PP), depuis la plantation des parcelles jusqu'à leur mise en jachère a été décrit par plusieurs auteurs (Tixier et al. 2004 ; Lassoudière 2007 ; Turner et al. 2007). Les parcelles sont généralement plantées avec des clones du cultivar Cavendish qui sont obtenus *in vitro*. Les vitro-plants créés en laboratoire sont acclimatés dans des serres où ils continuent leur croissance et gagnent en robustesse avant d'être plantés sur des parcelles. Lorsqu'ils sont plantés, ils sont donc âgés de plusieurs semaines. Ces plants vont de nouveau subir une phase d'acclimatation avant de commencer leur croissance végétative. Ils vont produire environ 30 feuilles avant que le bourgeon apical ne se différencie en bourgeon floral (Turner et al. 2007 ; Vargas et al. 2009). Au moment de la différenciation, environ 11 feuilles sont déjà initiées dans le pseudo-tronc et vont être émises progressivement avant que la fleur n'apparaisse (Turner et al. 2007). La floraison apparente termine la phase végétative et initie la phase reproductive. Durant cette phase, l'inflorescence va croître pour donner un régime de bananes. Cette phase se termine à la récolte du régime et le pied qui serait mort naturellement, est coupé. La pérennité de la production est assurée par la sélection d'un seul rejet parmi l'ensemble des rejets produits continuellement par le rhizome du pied parent. Tous les rejets émis sont supprimés régulièrement jusqu'au moment de la sélection du rejet fils qui est réalisée plusieurs semaines avant la récolte. Le rejet fils va assurer le cycle de floraison suivant et passe par les mêmes phases de développement que le vitro-plant. Ce cycle se répète ensuite indéfiniment jusqu'à la mort du rhizome. Une différence significative existe entre le premier cycle et les suivants, notamment parce que les rejets sont en compétition avec les pieds parents pour la lumière, ce qui n'est pas le cas pour le vitro plant. La PP d'une parcelle se termine par une mise en jachère avant une nouvelle plantation. La représentation simplifiée du cycle de développement d'un bananier est donnée Figure 3.1.

##### Formalisation probabiliste de la séquence de production pour simuler les floraisons d'un ensemble de pieds plantés en même temps

Le principe du modèle qui est décrit ci-dessous est illustré Figure 3.1. C'est un modèle probabiliste qui s'appuie sur la description simplifiée de chaque phase de développement du bananier par une variable aléatoire dont la loi de probabilité est définie *a priori*. Le principe du modèle ainsi que les lois choisies sont donnés ci-après. L'objectif est de modéliser le nombre de nouvelles fleurs qui apparaissent entre deux dates d'observations et sur une parcelle et de déduire à partir de ces observations.

Considérons un bananier, appelé , qui est planté à une date . La durée entre la plantation du vitro-plant et la récolte de son régime constitue le premier cycle de floraison noté (3.1). Pendant , va subir une phase d'acclimatation appelée phase juvénile et notée avant de commencer sa croissance végétative . Cette phase végétative débute par une période notée où le rejet qui va être sélectionné n'est pas encore apparu. La seconde période, notée est initiée à l'apparition du rejet sélectionné et se termine par l'apparition de l'inflorescence. On note le moment d'apparition de l'inflorescence. L'inflorescence de

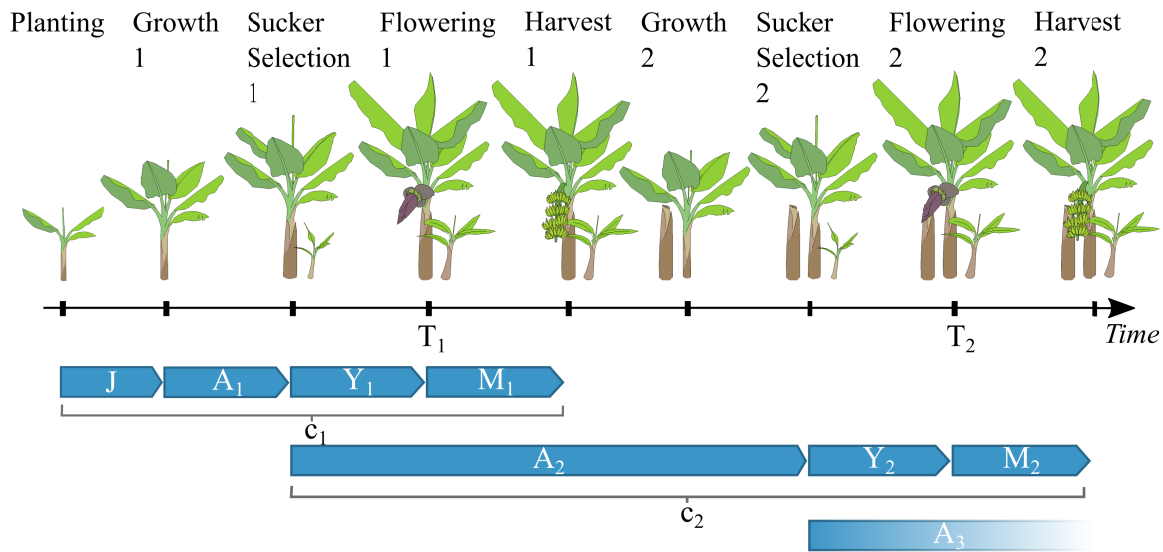


FIGURE 3.1 – Représentation des phases de développement d'un bananier dans une plantation commerciale.  $J$  représente la durée de la phase juvénile,  $A_1$  la durée de la phase après la phase juvénile et avant l'apparition du premier rejet,  $Y_1$  la durée de la phase entre l'apparition du rejet  $Y_1$  et le rejet  $M_1$ , la phase entre l'apparition du rejet  $M_1$  et la floraison  $A_2$ , la durée entre la floraison  $A_2$  and  $M_2$  le  $k^{\text{ème}}$  cycle de floraison.

va ensuite se développer avant d'être récoltée, ce qui constitue la phase  $A_2$ . Le rejet qui a été sélectionné sur  $M_1$  va assurer le deuxième cycle ( $M_2$ ). Il suit à son tour une phase végétative séparée en deux périodes  $J$  et  $A_1$  et une phase reproductive  $Y_1$ . Le moment de floraison de ce rejet est notée  $T_2$ . Ce schéma se répète pour tous les autres  $M_k$ .

La date de la floraison  $T_k$  du bananier  $k$  correspond donc à la somme des durées des phases successives  $J, A_1, Y_1, M_1, A_2, Y_2, M_2, \dots, A_k$ , avec  $k$  qui va de  $1$  à  $k$  et  $T_k$ .

$$(3.1)$$

Si on considère un ensemble de dates d'observations de floraisons, noté  $\{T_k\}$ , on appelle  $P_k(t_1, t_2)$  la probabilité que la  $k^{\text{ème}}$  fleur soit observée entre les deux dates  $t_1$  et  $t_2$ . Les lois de distributions des variables  $T_k$ ,  $P_k(t_1, t_2)$  et  $P_k(t_1, t_2)$  permettant de calculer cette probabilité sont définies ultérieurement.

$$(3.2)$$

On suppose que l'intervalle de temps entre deux observations de comptage est courte, ce qui fait qu'un bananier n'a pas le temps de fleurir deux fois dans cet intervalle. En conséquence, la probabilité d'observer une fleur, quel que soit  $k$ , est la somme des probabilités  $P_k(t_1, t_2)$  pour toutes les valeurs de  $k$ .

$$(3.3)$$

On introduit maintenant la mortalité dans le modèle et on considère que la durée de vie de  $k$  à partir de sa date de plantation peut être décrite par une variable aléatoire  $X_k$ . La probabilité que  $k$  meurt après une date  $t$  est appelée la fonction de survie et est notée  $S_k(t)$  définie par  $S_k(t) = P(X_k > t)$ .

On suppose que les mécanismes de floraison sont indépendants de ceux déterminant la survie de  $X$ , de sorte que la variable aléatoire  $X$  est indépendante des variables aléatoires  $Y$ . La probabilité  $P$  qu'une floraison soit observée entre les dates  $t_1$  et  $t_2$ , quel que soit  $X$ , s'écrit alors :

$$(3.4)$$

Toutes ces équations s'appliquent au cas d'un seul bananier. On s'intéresse maintenant à la modélisation d'une population de  $N$  bananiers plantés en même temps, indépendants, et soumis aux mêmes conditions de croissance. Le nombre de nouvelles fleurs observées entre deux dates  $t_1$  et  $t_2$ , selon ces hypothèses, peut alors être modélisé par une loi Binomiale de paramètres  $N$  et  $P$ , et est noté  $X$ .

$$(3.5)$$

Dans la suite de l'article, la série temporelle des nombres de fleurs observées  $X_t$  associée aux dates  $t$  est appelée une courbe de floraison CF. Afin de faciliter la lecture de certains graphiques, nous avons aussi représenté l'espérance du nombre de fleurs observées entre deux dates notée  $E(X)$ .

$$(3.6)$$

Le temps entre deux floraisons sur le même bananier correspond à  $T$ . En utilisant l'équation 3.1 du modèle,  $T$  s'écrit :

$$(3.7)$$

La durée moyenne des cycles de floraison, notée  $\mu$ , est l'espérance de  $T$  :

$$(3.8)$$

L'estimation de  $\mu$  à partir des données d'observation  $X_t$  d'une parcelle, revient donc à estimer l'espérance de  $T$ .

#### Choix des lois de distributions du modèle

Jusqu'à présent les lois de probabilités des variables aléatoires  $X$ ,  $Y$ , et  $T$  nécessaires pour simuler et pour estimer  $\mu$  n'ont pas été définies. On propose de les définir *a priori* en s'appuyant sur les travaux de Tixier et al. (2004), pour  $X$ ,  $Y$  et  $T$  et en utilisant une loi classique pour  $S$ . Des lois Gamma d'espérance  $\mu$  et d'écart-type  $\sigma$  ont été choisies<sup>1</sup> pour représenter respectivement les trois variables  $X$ ,  $Y$  et  $T$  (Équations 3.9, 3.10 et 3.11). La loi Gamma ( $r, \lambda$ ) possède une distribution d'allure similaire à la loi log-normale utilisée par Tixier et al. (2004). De plus, cette loi possède des propriétés d'additivité intéressantes pour calculer la loi de distribution de  $X$ . Quel que soit  $X$  on considère donc que  $X$  peut être représentée par une loi Gamma de paramètres identiques  $\mu$  et  $\sigma$ . On considère que quel que soit  $Y$ ,  $Y$  peut être représentée par une loi Gamma d'espérance  $\mu$  et d'écart-type  $\sigma$ . Une exception a été retenue pour la variable  $T$  qui suit une loi Gamma d'espérance  $\mu$  et d'écart-type  $\sigma$ . Le paramètre correctif  $r$  a été utilisé pour différencier l'écart-type de  $T$  de l'écart-type de  $X$ ,  $Y$  en raison de la particularité du premier cycle de floraison qui est assuré par un vitro-plant et non pas par un rejet.

$$(3.9)$$

$$(3.10)$$

1. Usuellement, la loi Gamma utilise deux paramètres, un paramètre d'échelle  $\lambda$  et un paramètre de forme  $r$ . Cependant, comme les variables manipulées sont des durées et qu'il est souvent plus facile de raisonner en termes d'espérance et d'écart-type sur ces durées, il a été choisi de transformer des paramètres en espérance  $\mu$  et en écart-type  $\sigma$ . Le passage d'un paramétrage à l'autre est le suivant : 
$$\begin{cases} \mu = r\lambda \\ \sigma = \sqrt{r} \cdot \lambda \end{cases} \leftrightarrow \begin{cases} r = (\mu/\sigma)^2 \\ \lambda = \sigma^2/\mu \end{cases}$$

(3.11)

Pour  $X$ , la loi de survie, une distribution de Weibull à trois paramètres a été choisie. Il s'agit d'une fonction classique pour modéliser une durée de vie (Lawless 2002).

(3.12)

La loi de distribution de  $Z$  (Équation 3.1) qui est la somme de variables aléatoires qui suivent des lois Gamma est calculée grâce aux propriétés d'additivité (éventuellement avec approximation) de la loi Gamma<sup>2</sup> (Famoye et al. 1995). Cela a permis de calculer toutes les lois de probabilités (Équations 3.2 à 3.4) qui permettent ensuite de simuler le résultat d'un comptage de nouvelles floraisons entre deux dates d'observation (Équation 3.5). Au total, 10 paramètres sont utilisés pour modéliser  $Z$ .

## Description du site d'étude et des données utilisées

### Site d'étude

Les données sont issues du processus de production de la société de production bananière Plantation du Haut Penja (PHP) située au Cameroun, à Njombé (WGS84, E : 4.612, N : 9.639). Le climat de cette région est tropical et se caractérise par de faibles variations des températures moyennes dans l'année et de fortes variations des précipitations mensuelles. La saison des pluies s'étend d'avril à octobre avec des précipitations mensuelles moyennes de 330 mm et une température moyenne de 26,8 °C. La saison sèche s'étend de novembre à mars avec des précipitations mensuelles moyennes de 72 mm et une température moyenne de 27,9 °C. La zone d'étude s'étend sur 3200 ha et est principalement située sur des andosols. Chaque parcelle est une unité homogène en termes d'opérations culturales (date de plantation, date de mise en jachère, fertilisation, traitements phytosanitaires, etc.). La surface moyenne d'une parcelle est d'environ 4.5 ha.

### Données de floraison et prétraitements associés

Les observations de floraison utilisées dans cette étude ont été effectuées manuellement entre janvier 2002 et avril 2017 dans le cadre d'un suivi habituel de la production. Ces observations sont effectuées une à deux fois par semaine dans chaque parcelle. Chaque inflorescence observée est marquée par un ruban ce qui permet d'éviter toute confusion entre deux passages successifs. Une même inflorescence ne peut donc pas être comptée deux fois. La fréquence de passage dans les parcelles a été variable en fonction des contraintes de gestion des ressources humaines. Pour chaque parcelle nous avons extrait une série de dates de comptages  $t_i$  et le nombre d'inflorescences  $n_i$  associé à chacune de ces dates. Certaines de ces séries étaient composées de plusieurs périodes de production (PP) séparées par des périodes sans comptage d'une durée supérieure à 2 mois. Ces périodes correspondaient à des mises en jachère suivies de replantation. Lors du prétraitement des données, ces périodes sans comptage ont été utilisées pour individualiser les PP. L'étude se concentre sur les deux dernières PP des parcelles. Par ailleurs, seules les PP de plus de 500 jours ont été conservées, car elles permettaient d'observer au moins  $t_i$  et  $n_i$  ce qui est une condition nécessaire pour ajuster le modèle. En concordance avec la Section 3.1.3, l'ensemble des dates de comptage et des comptages correspondant à une parcelle et à une période de production est appelée une courbe de floraison et est noté CF dans la suite de l'article.

2. Soit  $Z$  la somme de deux variables aléatoires indépendantes  $U$  et  $V$  qui suivent une loi Gamma de paramètres de forme  $r_U, r_V$  et de paramètres d'échelle  $\lambda_U, \lambda_V$ . Si  $X$  et  $Y$  possèdent le même paramètre d'échelle  $\lambda_U = \lambda_V$  alors  $Z$  suit une loi Gamma de paramètre de forme  $r_U + r_V$  et de paramètre d'échelle  $\lambda_U$ . Si  $\lambda_U \neq \lambda_V$ , alors, dans la gamme des valeurs de paramètres que nous sommes amenés à manipuler, on peut raisonnablement approximer la loi de  $Z$  par une loi Gamma de paramètre d'échelle  $\lambda_{Z*} = \frac{r_U \lambda_U^2 + r_V \lambda_V^2}{r_U \lambda_U + r_V \lambda_V}$  et de paramètre de forme  $r_{Z*} = \frac{[r_U \lambda_U + r_V \lambda_V]^2}{r_U \lambda_U^2 + r_V \lambda_V^2}$ .

### Qualification de la qualité des données de floraison

A l'issue de cette procédure, les CF ont été vérifiées pour s'assurer de leur cohérence par rapport aux résultats théoriques publiés dans la littérature. C'est-à-dire que l'on s'attendait à ce que pieds plantés en même temps fleurissent de manière relativement synchrone lors de , puisqu'au fur et à mesure des floraisons s'espacent dans le temps pour former une CF cyclique avec des pics de plus en plus étalés dans le temps (Tixier et al. 2004).

Les données ont été recueillies par plusieurs personnes pendant plusieurs années à des fins de gestion. Nous nous attendions à ce que certaines données ne soient pas fiables. En effet, nous nous sommes aperçus que le nombre de floraison du premier pic de floraison pouvait être très différent du nombre de pieds plantés renseigné dans la base de données ( ). Nous avons considéré que n'était pas fiable. Enfin, les dates de plantation enregistrées pouvaient être postérieures aux premières floraisons. Lorsque les dates étaient raisonnables (plus de 40 j entre la date de plantation et la date de première floraison) celles-ci ont été conservées comme le début de la PP. Dans les autres cas, la date de plantation a été définie automatiquement (date de première floraison - 50 j) pour faciliter la procédure de calibration du modèle.

Il est important de noter que les données ne comportaient pas d'informations sur la mortalité ordinaire (progressive et régulière) ni sur la mortalité soudaine liée à des vents forts, à un fort parasitisme ou à la gestion de la production, ni sur les replantations locales. C'est ce qui a motivé le fait de modéliser la mortalité dans le modèle. Il est aussi important de noter que les pieds ne sont pas identifiés, si bien que les floraisons observées ne sont pas associées à un bananier mais bien à une parcelle.

### Données sur le contexte de production des parcelles

Des données supplémentaires telles que le cultivar (William ou Grande Naine), la présence d'un système d'irrigation, les coordonnées géographiques des parcelles ont aussi été notées et utilisées dans cette étude. L'altitude et la pente des parcelles ont été obtenues à partir des cartes d'élévation ASTER Global Digital Elevation Model, produites par la NASA et METI (Figure 3.2).

Les données climatiques nécessaires pour évaluer l'effet des sommes thermiques sur ont été obtenues à partir d'une station météorologique située dans la plantation (thermomètre à mercure sous abri). Le calcul des sommes thermiques a été réalisé en utilisant une température de base de 14 °C (Tixier et al. 2004).

## **3.1.4 Paramétrisation et qualité d'ajustement du modèle**

### Présentation des paramètres du modèle

En théorie, 10 paramètres sont nécessaires pour définir les lois de distribution des variables du modèle (Section 3.1.3). Toutefois, la nature des données disponibles nous a conduit à reconsidérer le nombre de paramètres. En effet, les données ne permettent pas d'individualiser les pieds de bananier si bien qu'il n'était pas possible d'identifier séparément et : seule leur somme était identifiable. C'est pourquoi nous avons choisi de fixer l'écart-type de à la valeur 0, transformant ainsi cette variable aléatoire en une constante identifiable. Par ailleurs, compte tenu du fait que la date de plantation était une date estimée par le prétraitement, la variable est devenue une constante de calage pour la date de démarrage des floraisons. Les ajustements étant peu améliorés par l'utilisation du paramètre de décalage de la loi de Weibull, celui-ci a été fixé à 0. A noter que c'est lors de cette phase que la nécessité de différencier le premier pic de floraison des suivants en ajoutant le paramètre a été identifiée. L'examen des données nous a également amené à estimer le paramètre (Section 3.1.3). Ce paramètre ajusté n'est pas directement interprétable et prend certaines fois des valeurs très importantes. Seul le nombre de pieds encore vivants au moment de la première floraison estimée par est directement interprétable.

Les données d'entrée du modèle, les variables du modèle et les paramètres du modèle sont présentées dans les Tables 3.1, 3.2 et 3.3. Il est important de noter que parmi les 9 paramètres ajustés, seul qui

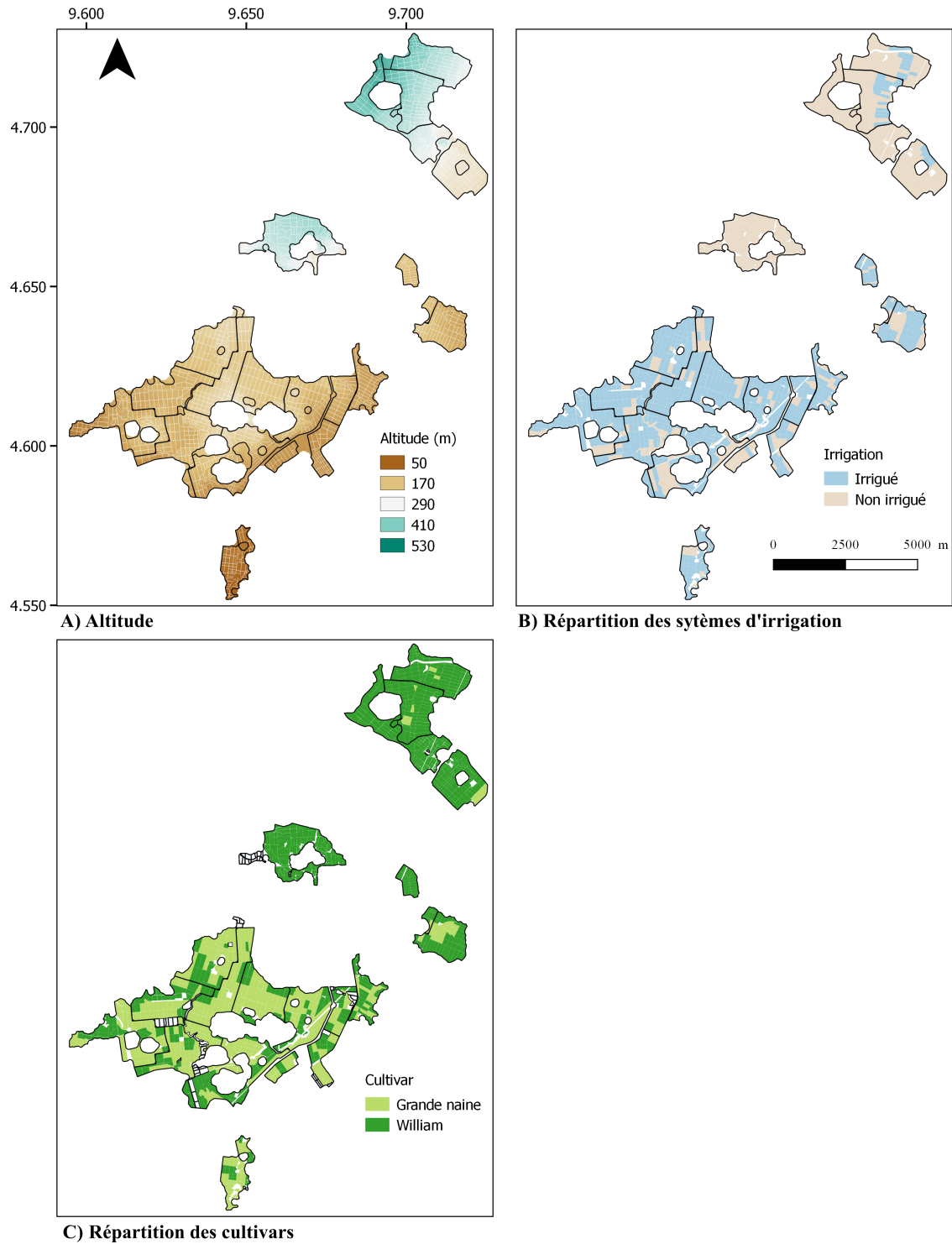


FIGURE 3.2 – Cartes de la zone d'étude montrant : A) l'altitude des parcelles de la plantation, B) les parcelles irriguées et non irriguées et C) la répartition des cultivars dans la plantation

TABLE 3.1 – Description des entrées du modèle

Donnée d'entrée	Description	Unité
D	Ensemble des dates de comptage des nouvelles inflorescences d'une parcelle	j
X	Ensemble du nombre de nouvelles inflorescences d'une parcelle	Inflorescences

TABLE 3.2 – Description des variables du modèle

Variable	Description	Unité	Distribution
	Durée de la phase juvénile	j	Constante
	Durée entre la fin de la phase juvénile du vitro-plant et l'apparition du premier rejet	j	Gamma
	Durée entre l'apparition successive de deux rejets sélectionnés	j	Gamma
	Durée de la phase entre l'apparition d'un rejet et la floraison du pied	j	Gamma
	Durée de survie des bananiers depuis leur plantation	j	Weibull

correspond à  $\theta$  est étudié.

#### Critères d'ajustement et méthode d'ajustement

Pour ajuster les paramètres à estimer, les seules informations utilisées sont  $D$  et  $X$ . Comme le nombre de fleurs comptées entre deux dates est modélisé par une loi Binomiale (Équation 5), en supposant l'indépendance des réalisations de ces lois pour des intervalles de dates différents, nous pouvons écrire la logvraisemblance de l'ensemble des données comme la somme des logvraisemblances pour chacun des intervalles de dates :

$$(3.13)$$

Où  $\theta$  désigne le vecteur de l'ensemble des paramètres du modèle,  $n$  est le nombre estimé de pieds encore en vie à la date  $t_j$ ,  $p_j$  est la probabilité de floraison entre les dates  $t_{j-1}$  et  $t_j$  calculée à partir des

TABLE 3.3 – Description des paramètres du modèle

Paramètre	Description	Unité	Valeur
	Espérance de	j	Calibrée
	Ecart-type de	j	0
	Espérance de	j	Calibrée
	Ecart-type de	j	Calibrée
	Coefficient de correction de $\theta$ pour la durée	-	Calibrée
	Espérance de	j	Calibrée
	Ecart-type de	j	Calibrée
	Paramètre de décalage de	j	0
	Paramètre de forme de	-	Calibrée
	Paramètre d'échelle de	j	Calibrée
	Nombre de pieds encore vivant à la date de première floraison	bananiers	Calibrée



mêmes paramètres et  $n$  est le nombre de fleurs observées à la date  $t$ . Les estimations des paramètres sont celles issues de la méthode Maximum de Vraisemblance c'est-à-dire les valeurs des composantes de  $\hat{\theta}$  qui maximisent la fonction de l'équation 3.13. Il s'agit d'une optimisation sous contrainte puisque la plupart des paramètres doivent prendre des valeurs positives. La durée moyenne des cycles de floraisons  $\mu$  a été contrainte à prendre une valeur comprise entre 100 et 400 jours. Ces bornes ont été fixées de manière à faciliter la convergence de l'algorithme d'estimation en l'empêchant d'explorer des régions de l'espace des paramètres qui ne peuvent pas contenir une solution optimale raisonnable. En pratique, nous avons simplement utilisé une bijection de l'espace admissible des paramètres sur  $\mathbb{R}$  pour pouvoir appliquer des procédures d'optimisation sans contrainte. Par exemple pour un paramètre  $\mu$  contraint à la positivité, nous avons réécrit la vraisemblance en l'exprimant en fonction de  $\ln \mu$  et optimisé la fonction résultante. Nous obtenons alors une estimation de  $\ln \mu$  dont il suffit de prendre l'exponentielle pour obtenir une estimation de  $\mu$  qui est alors nécessairement positive.

#### Critères de validation de l'estimation de $\theta$

Lorsque l'ajustement n'a pas convergé ou que le  $\mu$  estimé correspondait à une borne d'ajustement (100 j ou 400 j) les CF n'ont pas été considérées dans l'étude.

Par ailleurs, les CF pour lesquelles le  $\mu$  était trop éloigné du centre de la distribution des estimés ont été considérées comme aberrantes. Un seuil de 5 % a été fixé ce qui a permis de conserver 95 % des données estimées. Dans tous les autres cas, pour évaluer l'aptitude du modèle à estimer  $\mu$  pour chaque parcelle et chaque PP, il aurait fallu idéalement pouvoir comparer le  $\mu$  estimé à la valeur vraie, ce que nous n'avions pas dans la base de données. On rappelle par ailleurs que ce paramètre aurait été coûteux à estimer même expérimentalement.

Nous avons donc défini d'autres critères : (i) la capacité du modèle à reproduire la forme attendue d'une courbe de floraison, puis (ii) la cohérence spatiale de l'estimation des  $\mu$  sur des parcelles voisines, l'indice choisi étant l'indice de Cambardella (Section 3.1.4) et (iii) la cohérence de  $\mu$  avec les ordres de grandeurs et les sources de variabilités issues de la littérature.

### **Analyse des sources de variabilité du $\mu$**

#### Analyse de la variabilité spatiale

La structure spatiale de  $\mu$  a été étudiée pour la dernière PP des parcelles (la PP toujours en cours en avril 2017). Pour chaque  $\mu$  estimée, les coordonnées du centre de la parcelle correspondante lui ont été attribuées. La structure spatiale de  $\mu$  a ensuite été étudiée en calculant le variogramme expérimental selon l'équation 3.14.

$$\gamma(h) = \frac{1}{2n} \sum_{i,j=1}^n (\mu_i - \mu_j)^2 \mathbb{1}_{\|x_i - x_j\| \leq h} \quad (3.14)$$

Dans cette équation  $\mu_i$  et  $\mu_j$  sont les valeurs des  $\mu$  estimés aux endroits  $x_i$  et  $x_j$ . La variable  $x$  désigne le centre de la parcelle sur laquelle le  $\mu$  a été estimé et qui est défini par ses coordonnées cartésiennes  $(x, y)$ . La variable  $h$  représente la distance qui sépare deux parcelles.  $n$  représente le nombre de  $\mu$  qui peuvent être comparés à une distance  $h$ . L'ensemble des  $\gamma(h)$  obtenus en prenant une séquence de  $h$  forme le variogramme expérimental. Le calcul du variogramme expérimental a été fait en utilisant 11 valeurs de  $h$  s'étalant régulièrement de 100 à 1100 m. Le  $h$  minimum a été choisi à 100 m car cela correspond à la distance moyenne entre les centres des parcelles voisines.

La variance pépète  $\sigma^2$  qui correspond à la variance non structurée spatialement a été estimée par la valeur du variogramme au plus petit  $h$ . La variance palier  $\sigma^2_{\text{total}}$  est la somme de la variance pépète et de la variance du palier partiel  $\sigma^2_{\text{spatial}}$  qui correspond à la variance structurée spatialement. La somme  $\sigma^2_{\text{total}}$  a été estimée par la valeur du variogramme au  $h$  le plus élevé. La proportion de variance structurée

spatialement a été évaluée par le ratio défini par l'équation 3.15 qui représente la proportion de variance structurée spatialement.

(3.15)

Le variogramme expérimental a été modélisé en utilisant la somme d'un variogramme pépité et de deux variogrammes gaussiens afin de modéliser un variogramme avec deux structures spatiales de deux portées différentes. La somme de ces fonctions a constitué le variogramme théorique. Les variances pépité, du palier partiel et du palier du variogramme théorique sont notées  $\sigma_p^2$ ,  $\sigma_{p1}^2$  et  $\sigma_{p2}^2$ , respectivement. Ces variances permettent de calculer le ratio théorique noté  $\rho$ . Les  $\rho$  ont été interpolés par krigeage ordinaire à partir du variogramme théorique sur une grille de points régulièrement espacés de 30 m dans l'enveloppe géographique des parcelles.

#### Analyse de l'effet des facteurs du milieu

Les effets de l'altitude, du cultivar et de l'irrigation sur  $\rho$  ont été analysés en utilisant un modèle linéaire général (Équation 3.16).

(3.16)

Avec  $\rho$ , la durée moyenne des cycles de floraison estimée par le modèle sur une parcelle de centre  $i$ , la constante de référence de la population, l'effet de l'altitude en notée  $\alpha$ , l'effet du cultivar en noté  $\beta$ , l'effet de l'irrigation en noté  $\gamma$  et les résidus du modèle que l'on suppose suivre un processus gaussien spatialisé.

(3.17)

Avec  $\sigma_p^2$  la variance de  $\rho$  et  $\Sigma$  la matrice de corrélation spatiale. La structure de corrélation spatiale entre les  $\rho$  a été modélisée par un variogramme gaussien selon la méthode proposée par Pinheiro and Bates (2000). L'effet des facteurs du modèle a été évalué en utilisant une approche « Top down » selon le Critère d'Information Bayésien (BIC). Le résidu  $\epsilon$  a été cartographié selon la même méthode que celle présentée pour  $\rho$  pour représenter la variabilité résiduelle du modèle, c'est-à-dire celle non expliquée par les facteurs irrigation, cultivar et altitude.

#### Analyse de l'effet des températures sur les floraisons

Pour chaque CF, l'effet des températures a été testé en comparant le BIC du modèle calibré en utilisant une échelle temporelle en jours au BIC du modèle calibré en utilisant une échelle temporelle en degrés jours.

#### **Logiciel utilisé**

Le modèle a été codé sous R et la procédure d'ajustement a été construite en utilisant les fonctions « optim » et « mle ». L'étude du variogramme et le krigeage ont été réalisées en utilisant les packages « sp » (Pebesma and Bivand 2005) et « gstat » (Pebesma 2004). La régression linéaire générale a été faite avec la fonction « gls » du package « nlme » (Pinheiro et al. 2018).

### **3.1.5 Résultats**

#### **Description et qualification de la qualité des données de floraison**

Les 1281 courbes de floraison (CF) ont été obtenues pour 724 parcelles différentes, certaines parcelles présentant une ou deux périodes de production (PP). Une grande majorité des CF de la base de données

a présenté une allure générale semblable. C'est à dire que les nuages de points formaient une succession de pics régulièrement espacés qui s'étaient de plus en plus dans le temps jusqu'à devenir indiscernables (Figure 3.3 A et B). Cette allure générale a correspondu aux CF obtenues par différents auteurs dans des bananeraies expérimentales ou commerciales (Tixier et al. 2004 ; Cottin et al. 1987).

Malgré une allure générale semblable, les CF ont présenté des différences en termes de nombre d'observations, de nombre moyen de fleurs, de durées et de nombre de pics visibles. Le nombre moyen de fleurs a dépendu notamment de la surface des parcelles qui a déterminé le nombre de pieds plantés (la densité de plantation étant approximativement la même pour toutes les parcelles du domaine). La durée des périodes de production (PP) a été variable et a notamment dépendu de la perte de vigueur de certaines parcelles qui ont été mises en jachère précocement. La fréquence d'observation dans les parcelles a pu varier entre 1 et 15 jours en fonction des parcelles et des périodes de l'année. Cela a dépendu des contraintes de gestion des exploitants et de la disponibilité de la main d'œuvre. Cet écart variable au sein des CF a entraîné une irrégularité de . En effet, à une date , plus le temps depuis le dernier comptage est élevé, plus le nombre de fleurs qui va être compté est important.

Certaines CF n'ont pas présenté le comportement attendu (Figure 3.3 C et D). Ces cas sont particuliers et ont été observés moins fréquemment dans la base de données. La CF de la Figure 3.3 C présente des pics de floraison irrégulièrement espacés avec un pic de floraison soudain fin 2016. Ce cas peut s'expliquer par une mortalité soudaine suivie d'une replantation. La CF de la Figure 3.3 D présente une augmentation de la hauteur des pics de floraison au cours de la vie de la plantation. Ce cas pourrait s'expliquer par un agrandissement de la parcelle qui aurait entraîné une augmentation du nombre de pieds (et de fleurs) observés. En l'absence de métadonnées permettant de connaître les événements subis par les parcelles, ces CF, bien qu'atypiques, ont été conservées dans la base de données.

### Présentation du modèle et de son ajustement sur les données de floraison

Un exemple de simulation d'une CF à partir du modèle est donné Figure 3.4. Le modèle permet de réaliser un tirage aléatoire du nombre de nouvelles floraisons comptées (points noirs) mais également de calculer l'espérance de la loi de probabilité associée (ligne rouge). Quatre exemples d'ajustement du modèle sur les CF ont été représentés Figure 3.3 (ligne rouge). Pour 97 % des CF de la base de données (1251/1281), l'ajustement a convergé et a permis d'estimer les paramètres du modèle.

L'histogramme des  $\mu$  estimés par le modèle est représenté sur la Figure 3.5. La médiane des  $\mu$  est de 209 j et l'écart-type est de 24 j. Parmi les  $\mu$  estimés, 95 % est situé entre 172 j et 257 j. Ce résultat met en évidence une forte variabilité de  $\mu$  à l'intérieur de la plantation. Les  $\mu$  situés à l'extérieur de cet intervalle (63/1251 CF) n'ont pas été utilisés pour la suite de l'analyse.

### Analyse de la structure spatiale de

Le variogramme expérimental des  $\mu$  est présenté Figure 3.6. Les variances  $\sigma^2$  et  $\sigma$  valent respectivement 100 j<sup>2</sup> et 10 j et le ratio  $\rho$  vaut 0.64. La forme du variogramme expérimental met en évidence deux structures spatiales avec des portées différentes. En conséquence, le variogramme théorique a été modélisé en utilisant la somme de plusieurs variogrammes : un variogramme pépète de variance  $\sigma^2$  de 79 j<sup>2</sup>, un variogramme gaussien à faible portée de paramètres  $\sigma^2$  et  $\rho$  de 217 m et 113 j respectivement, et un variogramme gaussien avec une plus grande portée de paramètres  $\sigma^2$  et  $\rho$  de 669 m et 93 j respectivement. Le ratio théorique est plus élevé que le ratio expérimental,  $\rho_{th} = 0.7$ . Ces résultats montrent que le  $\mu$  est auto-corrélé spatialement ( $\rho$  et  $\sigma$  sont relativement faibles comparativement à  $\mu$ ). Un exemple de CF de deux parcelles voisines est présenté Figure 3.7. On voit que l'allure générale des deux CF est très similaire et que l'espacement entre chacun des pics est semblable ce qui illustre la corrélation entre les  $\mu$  de parcelles proches. La carte de  $\mu$  sur la plantation est présentée Figure 3.8. Cette carte permet d'illustrer la variabilité des  $\mu$  à l'échelle du domaine. Le nord de la plantation présente des  $\mu$  de près de 260 j alors que les valeurs de la partie centrale de la plantation sont fréquemment proches de 180 j.

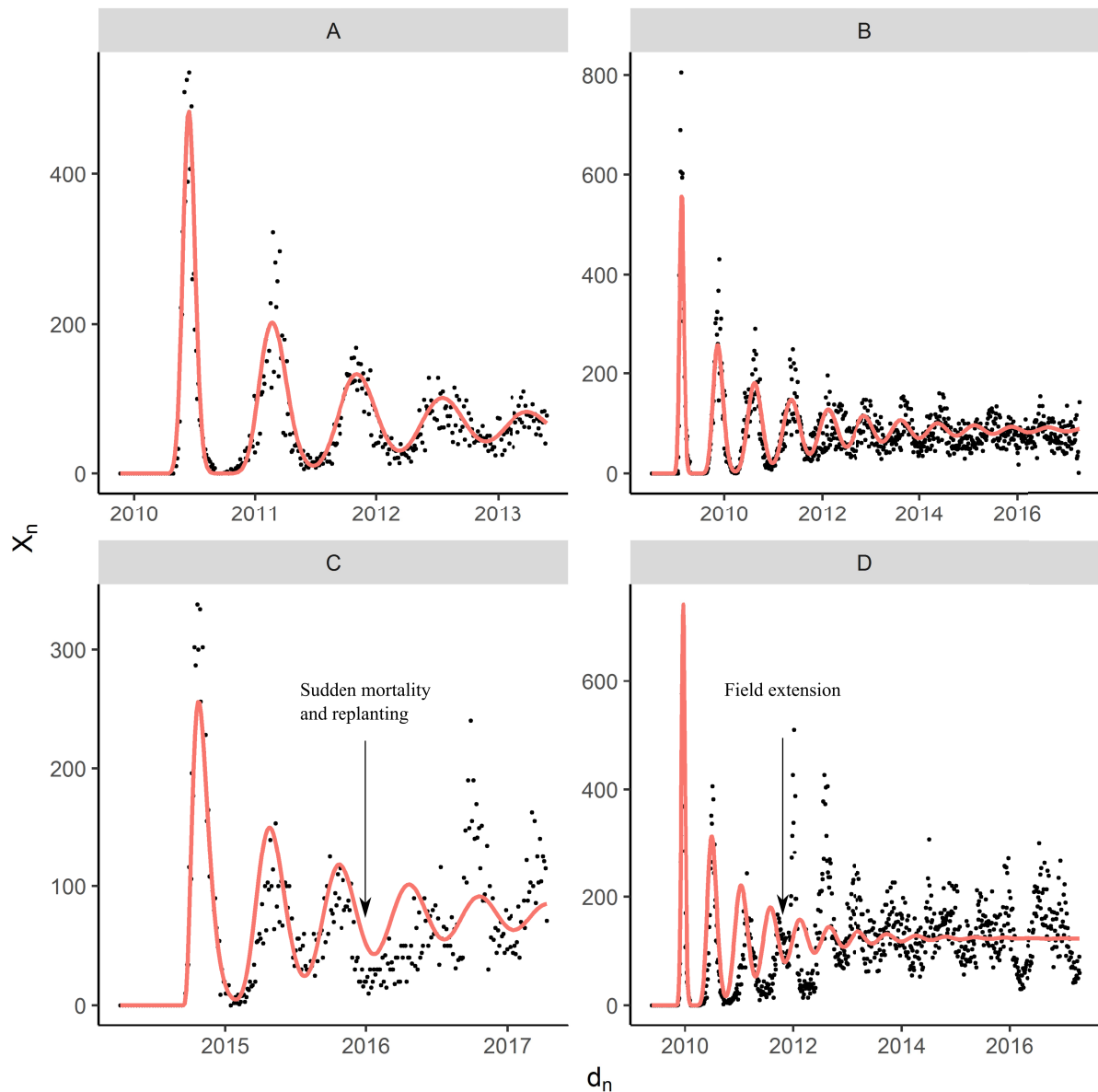


FIGURE 3.3 – Exemples de CF observées (points). Simulation de  $X_n$  (ligne rouge) pour des dates de comptage régulièrement espacées, d'un écart constant égal à l'écart moyen des dates de comptage de la CF. Les CF A et B correspondent à des courbes types attendues. Les CF C et D correspondent à des cas atypiques où les hypothèses du modèle ne sont pas respectées (mortalité soudaine, replantation, mélange de populations) et où il est difficile d'estimer le  $d_n$ .

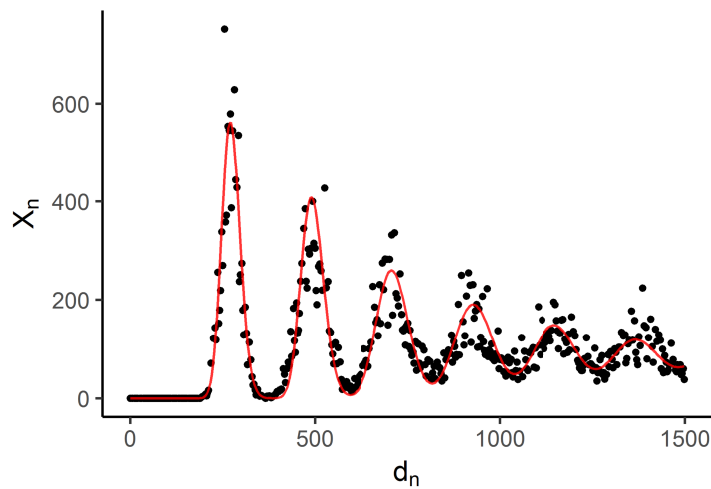


FIGURE 3.4 – Simulation de  $X_n$  pour des dates d'observation irrégulièrement espacées (points) et simulation de  $X_n$  pour des dates d'observations régulièrement espacées, d'un écart constant égal à l'écart moyen des dates d'observations de la CF (ligne rouge). Les paramètres utilisés pour la simulation sont les suivants :  $\mu = 100$ ,  $\sigma = 100$ ,  $\lambda = 0.001$ ,  $\alpha = 0.001$ ,  $\beta = 0.001$ ,  $\gamma = 0.001$ ,  $\delta = 0.001$ ,  $\epsilon = 0.001$ ,  $\zeta = 0.001$ ,  $\eta = 0.001$ ,  $\theta = 0.001$ ,  $\iota = 0.001$ ,  $\kappa = 0.001$ ,  $\lambda = 0.001$ ,  $\mu = 0.001$ ,  $\nu = 0.001$ ,  $\xi = 0.001$ ,  $\omicron = 0.001$ ,  $\pi = 0.001$ ,  $\rho = 0.001$ ,  $\sigma = 0.001$ ,  $\tau = 0.001$ ,  $\upsilon = 0.001$ ,  $\phi = 0.001$ ,  $\chi = 0.001$ ,  $\psi = 0.001$ ,  $\omega = 0.001$ ,  $\delta = 0.001$ ,  $\epsilon = 0.001$ ,  $\zeta = 0.001$ ,  $\eta = 0.001$ ,  $\theta = 0.001$ ,  $\iota = 0.001$ ,  $\kappa = 0.001$ ,  $\lambda = 0.001$ ,  $\mu = 0.001$ ,  $\nu = 0.001$ ,  $\xi = 0.001$ ,  $\omicron = 0.001$ ,  $\pi = 0.001$ ,  $\rho = 0.001$ ,  $\sigma = 0.001$ ,  $\tau = 0.001$ ,  $\upsilon = 0.001$ ,  $\phi = 0.001$ ,  $\chi = 0.001$ ,  $\psi = 0.001$ ,  $\omega = 0.001$ .

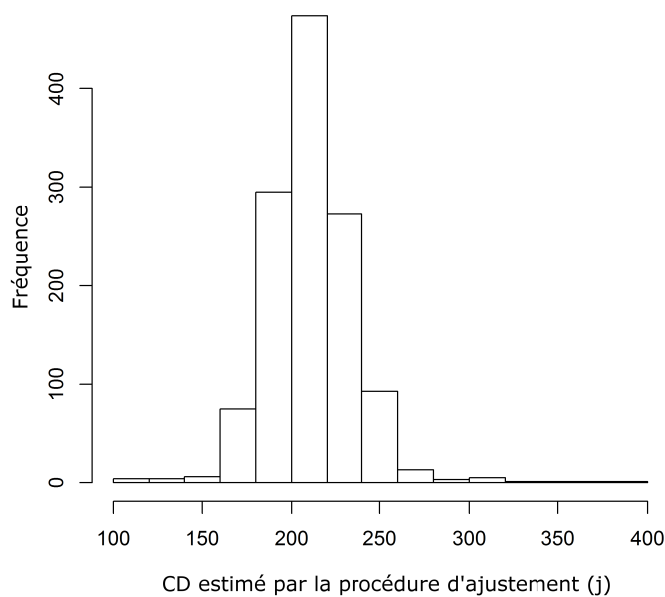


FIGURE 3.5 – Histogramme des  $CD$  estimés sur l'ensemble des CF par la procédure d'ajustement du modèle.

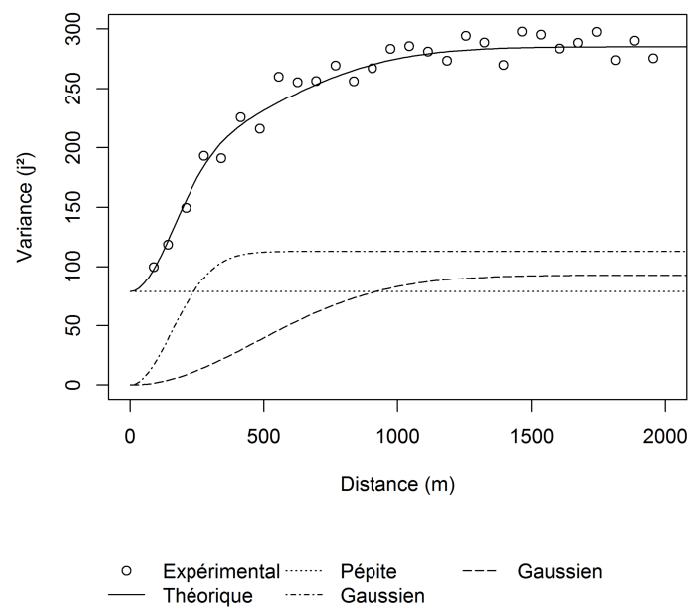


FIGURE 3.6 – Variogramme expérimental (points) et variogramme théorique (ligne pleine). Le variogramme théorique est la somme d'un variogramme pépité et d'un variogramme gaussien qui sont représentés par des tirets.

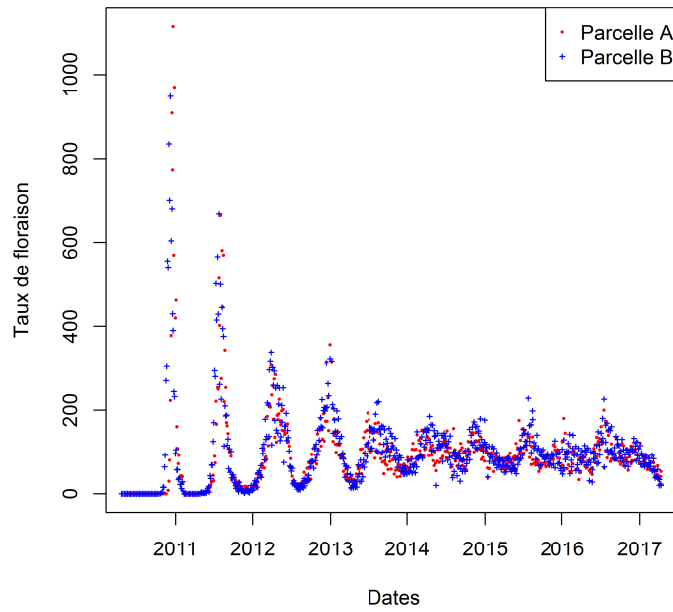


FIGURE 3.7 – Courbes de floraisons normalisées de deux parcelles voisines.

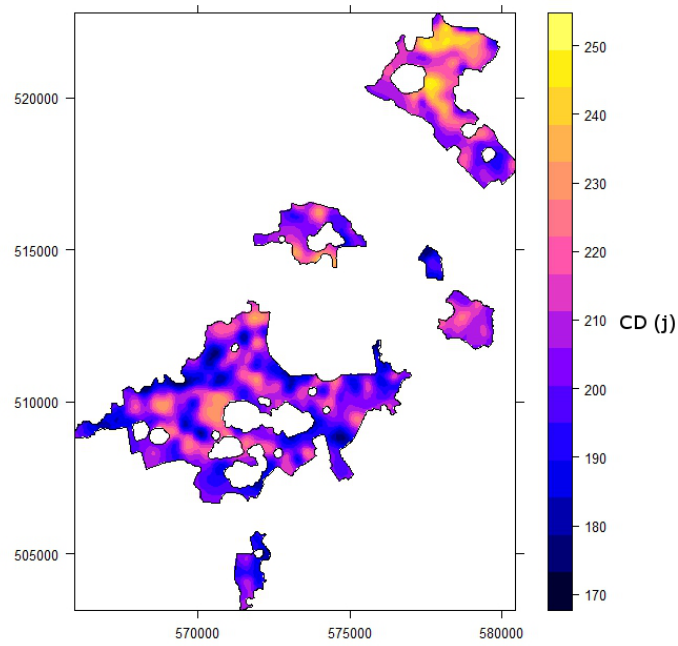


FIGURE 3.8 – Carte des  $CD$  interpolés. Les coordonnées sont affichées en UTM 32 N

TABLE 3.4 – Coefficients de la régression linéaire générale présentée équation 3.16. Avec  $\beta_0$  la constante de référence de la population (Altitude 0 m, cultivar Williams, irrigué),  $\beta_1$  l'effet de l'altitude,  $\beta_2$  l'effet du cultivar Grande Naine,  $\beta_3$  l'effet de l'irrigation,  $\sigma^2$  la variance des résidus,  $\sigma^2_{sp}$  la portée de la structure spatiale des résidus,  $\sigma^2_{pep}$  la variance pépète des résidus. \*\* p-value  $\leq 0.005$ , \* p-value  $\leq 0.05$

Paramètre	Estimation	Ecart-type
$\beta_0$ (j)	192.41**	2.29
$\beta_1$ (j/m)	0.10**	0.011
Grande Naine ( $\beta_2$ )	-14.26**	1.40
Non irrigué ( $\beta_3$ )	3.26*	1.40
$\sigma^2$ (j <sup>2</sup> )	209.87	-
$\sigma^2_{sp}$ (m)	320.72	-
$\sigma^2_{pep}$ (j <sup>2</sup> )	100.04	-

### Analyse de la variabilité de

Les résultats du modèle linéaire général sont présentés Table 3.4. L'altitude a un effet sur  $\beta_0$ , une augmentation de 100 m d'altitude entraîne une augmentation de  $\beta_0$  d'environ 10 j. L'effet du cultivar est aussi important, le cultivar William présente des  $\beta_2$  en moyenne de 14 j plus courts que la variété Grande Naine. L'effet de l'irrigation est significatif mais très faible, le fait de pouvoir irriguer diminue le  $\beta_3$  de 3.2 j. Les résidus ( $\sigma^2$ ) présentent une variance totale de 210 j<sup>2</sup> et une variance pépète de 100 j<sup>2</sup> ce qui donne un  $\sigma$  de 0.52. Cela montre que lorsque les effets de l'altitude, du cultivar et de l'irrigation sont pris en compte, environ 50 % de la variabilité structurée spatialement reste non expliquée par ces facteurs. On peut noter que  $\sigma^2_{sp}$  est plus petit pour  $\beta_0$  que pour  $\beta_1$  (110 j<sup>2</sup> contre 180 j<sup>2</sup>) mais que  $\sigma^2_{pep}$  est sensiblement le même pour ces deux variables (100 j<sup>2</sup> dans les deux cas). Les facteurs pris en compte dans le modèle linéaire général permettent donc surtout d'expliquer une part de la variance structurée spatialement. Ce résultat permet de supposer que la variabilité erratique (non structurée dans l'espace) provient de phénomènes à courtes distances (différences naturelles entre les parcelles, choix culturaux, variabilité d'autres facteurs non pris en compte) et de l'incertitude d'estimation de  $\beta_0$ . Cette variabilité non structurée est visible sur la carte de  $\beta_0$  (Figure 3.9) qui présente des outliers locaux c'est-à-dire des valeurs extrêmes par rapport à leur voisinage local (quelques une de ces valeurs extrêmes sont mises en évidence par des flèches).

L'utilisation d'une échelle temporelle en degrés jours plutôt qu'en jours a diminué le BIC dans 7.5 % des cas. Dans tous les autres cas le BIC a été plus faible pour une échelle en jours ce qui montre que les degrés jours calculés avec une température de base de 14 °C n'améliorent pas la qualité de l'ajustement du modèle.

### 3.1.6 Discussion

#### Représentation probabiliste d'une courbe de floraison

Le modèle a été établi sur la base d'une représentation de la durée des phases de développement du bananier et, excepté pour la loi de mortalité, tous les paramètres ont été définis en termes d'espérances ou d'écart-types homogènes à des durées (Table 3.3). Ces choix de modélisation ont été effectués pour que les paramètres estimés à partir de la procédure de calibration sur les données des producteurs soient facilement interprétables.

Le modèle reproduit bien le comportement général des CF, c'est à dire une première floraison relativement synchrone, avec une durée entre l'apparition de la première fleur de la parcelle et de la dernière assez regroupée dans le temps, puis un étalement progressif de cette durée pour les cycles de floraison ultérieurs. Cet étalement est connu et provoqué par l'asynchronisme des bananiers. Par rapport à d'autres



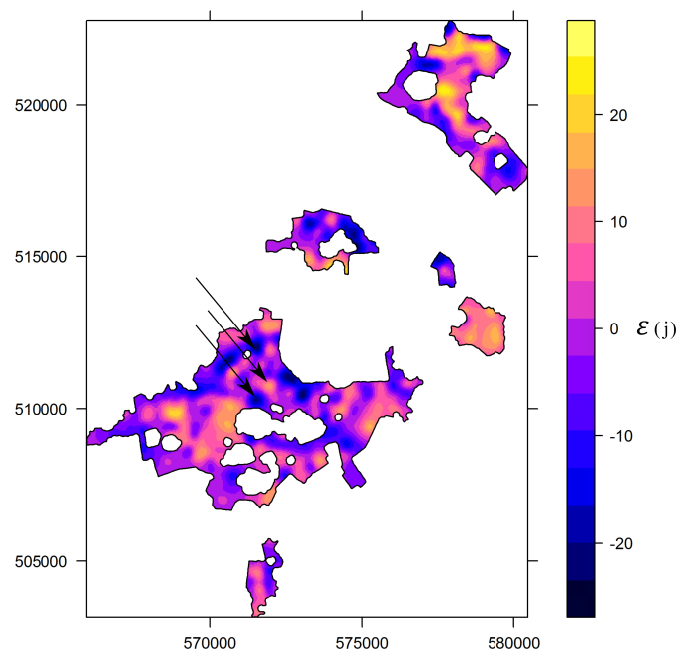


FIGURE 3.9 – Cartographie des résidus du modèle linéaire général présenté équation 3.16. Les coordonnées sont affichées en UTM 32 N. Les flèches représentent quelques extrema locaux.

approches, telles que le modèle de forme proposé par Cottin et al. (1987), le modèle offre l'avantage de représenter un nombre de cycles de floraison potentiellement infini. Ceci permet de s'adapter à des PP variables comme c'est le cas en production.

Les équations et lois choisies dans le modèle (Équations 3.1 à 3.12) impliquent plusieurs hypothèses sur le comportement d'une population de bananiers plantés en même temps : (i) la mortalité est régulière et indépendante de la durée des phases  $\tau_1$  et  $\tau_2$  (Équations 3.4 3.12), (ii) les pieds sont considérés indépendants et soumis aux mêmes conditions de croissance (Équation 3.5), (iii) il n'y a pas de replantation au cours de la vie de la parcelle (Équation 3.5), et (iv) la distribution des  $\tau_1$  ne dépend pas du cycle de floraison ( $\tau_1$ ) (Équation 3.10).

Les hypothèses de mortalité régulière et d'absence de replantation sont simplificatrices par rapport aux phénomènes connus. En effet, les producteurs sont amenés à remplacer les pieds faibles où les pieds tombés. Cette pratique a lieu tout au long des PP des parcelles. En règle générale, ces opérations ne sont pas tracées, si bien qu'en production, ces informations ne sont pas disponibles. La mortalité ajustée par la procédure de calibration de notre modèle correspond donc à une mortalité apparente, c'est-à-dire au bilan du nombre de pieds morts qui n'ont pas été remplacés.

L'hypothèse d'indépendance des pieds et d'homogénéité des conditions de croissance dans les parcelles est aussi vraisemblablement simplificatrice car il a été montré sur d'autres propriétés culturales du bananier comme la circonférence, la maturité ou des maladies que celles-ci peuvent être structurées dans les parcelles (Freitas et al. 2015; Leroux et al. 2018a; Lamour et al. 2017). Les données n'étant pas géolocalisées il n'est pas possible de modéliser la corrélation entre les plants dans cette étude. L'hypothèse d'identité de la distribution des  $\tau_1$  quel que soit  $\tau_1$  (choix de la loi de distribution et de ses paramètres) est l'hypothèse la plus simple pour modéliser les floraisons sachant que le cycle de floraison n'est pas connu en pratique. C'est une hypothèse qui implique que le comportement des bananiers est suffisamment stable dans le temps pour que les variations météorologiques (hydrique et température notamment) ou d'itinéraire technique ne modifient pas fortement  $\tau_1$  d'un cycle à l'autre. Ces hypothèses simplificatrices étaient nécessaires pour établir un modèle permettant de calculer  $\tau_1$  et pour estimer les paramètres du modèle à partir des données disponibles. Malgré ces hypothèses, le modèle proposé s'ajuste correctement aux CF observées sur un grand nombre de parcelles (Figure 3.3 A et B). Ces hypothèses sont donc raisonnables dans les conditions agronomiques de cette plantation, du moins tant que les replantations ne sont pas trop importantes (Figure 3.3 C et D). Par ailleurs, on observe une structure spatiale ( $\tau_1$ ) importante du  $\tau_1$  estimé à partir du modèle sur l'ensemble du domaine constitué de parcelles différentes ce qui confirme la capacité du modèle à estimer de manière cohérente le  $\tau_1$  sur des jeux de données différents. Dans cette étude, seul le  $\tau_1$  a été présenté mais d'autres propriétés des courbes de floraisons telles que le nombre de pieds productifs ou l'hétérogénéité des temps de croissance au sein des parcelles ( $\tau_1$  ou  $\tau_2$ ) pourraient constituer des sources d'information utiles aux producteurs pour faire un diagnostic de leur production. On peut aussi noter qu'il serait possible d'utiliser ce modèle pour faire des prévisions de floraisons.

### Sources de variabilité du

La durée moyenne des cycles de floraisons  $\tau_1$  obtenue dans cette étude (209 j) est similaire à celles que Arantes et al. (2017) et Dens et al. (2008) ont obtenu respectivement dans la région de Bahia au Brésil et au Honduras. Par contre, elle est en moyenne plus courte que celles obtenues par Cottin et al. (1987) et Dorel et al. (2016) respectivement en Martinique (306 j) et en Guadeloupe (249 j). Elle est plus longue que celle obtenue par Weber et al. (2017) dans la région de Ceará au Brésil (180 jours). Les différences de vitesse de cycle de floraison entre des localités avaient déjà été montrées par d'autres auteurs (Turner et al. 2007). Par contre, la variabilité spatiale au sein d'une plantation n'avait jamais été mise en évidence. La présence d'une variabilité spatiale structurée importante du  $\tau_1$  montre que ce dernier dépend fortement des conditions locales et que cela doit être pris en compte pour comprendre le rendement des parcelles et, *in fine*, adapter les pratiques par site.

Des liens ont été observés entre le  $\tau_1$  estimé par le modèle et les propriétés de l'environnement

et de l'itinéraire technique (Table 3.4). L'analyse de la variance a en effet montré un effet du cultivar, de l'altitude et de l'irrigation. Cette analyse est faite sur des données de production et donc sans plan d'expérience. Cela ressemble donc à une étude observationnelle comme cela peut-être fait en foresterie (von Gadow et al. 2016) en écologie (Tilman 1989; Shaffer and Johnson 2008) ou encore en médecine (Rubin 2006). L'absence de plan d'expérience augmente le risque de confondre l'effet des facteurs et de se tromper sur leur estimation à cause de facteurs cachés. Les résidus du modèle linéaire général présentent une structure spatiale avec un ratio d'auto-corrélation de 0.52. La modélisation de l'auto-corrélation spatiale des résidus est donc pertinente et permet de limiter les risques de biais dans l'estimation de l'effet des facteurs fixes (cultivar, altitude et irrigation). On suppose que cette variabilité résiduelle structurée spatialement est due à un facteur de l'environnement qui n'a pas été pris en compte dans cette étude. Le sol est un facteur qui agit sur le (Brisson et al. 1998) et pourrait expliquer une part de cette variabilité résiduelle.

L'absence d'irrigation augmente légèrement la durée des cycles de floraison. Cela a été montré par Brisson et al. (1998) sur le premier cycle de floraison de bananiers aux Antilles françaises. Dans leur expérience, le délai sur la floraison a été de 6 et 12 jours. Dans des conditions semi-arides, un déficit d'irrigation peut entraîner des délais encore plus longs, entre 42 et 79 jours (Goenaga and Irizarry 1995). Les estimés par le modèle semblent montrer un délai plus faible de seulement 3.3 jours (Table 3.4). Cela pourrait s'expliquer par des différences de contraintes hydriques moins fortes entre les parcelles irriguées et les parcelles non irriguées de cette plantation. La base de données ne contient pas d'information sur les volumes d'eau apportés, ce qui ne permet pas d'évaluer l'état de stress hydrique des parcelles.

L'augmentation du avec l'altitude est très similaire à ce qui a été mis en évidence par Cottin et al. (1987). L'augmentation de la durée des cycles avec l'altitude s'explique probablement par une diminution de la température. En effet, les températures moyennes de la partie Nord de la plantation sont 1.15 °C plus faibles que celles de la partie centrale de la plantation (résultat non montré).

Notre étude montre une différence significative de 14 j sur le estimé entre les variétés. L'étude de Weber et al. (2017), qui a suivi un plan d'expérience randomisé complet (3 répétitions de 20 plants par cultivar), n'a pas montré de différence entre les cultivars Grande Naine et Williams pour le temps entre la plantation et la 3<sup>ème</sup> récolte. Deux hypothèses pourraient expliquer cette différence entre les deux études : (i) il y a une différence agronomique réelle entre les cultivars et elle peut être mise en évidence dans cette plantation grâce au nombre important de parcelles d'études (ii) la différence résulte d'un facteur caché, comme par exemple la décision du producteur de raisonner l'implantation d'un cultivar en fonction de ce facteur et d'en adapter l'itinéraire technique en conséquence. La présente étude n'a pas permis de tester ces hypothèses.

Le fait d'utiliser une échelle temporelle en degrés jours par rapport à une échelle en jours juliens n'améliore pas le modèle. L'effet temporel de la température a été testé en utilisant une seule station météo car les évolutions saisonnières au sein de la plantation sont très similaires (résultat non montré). Plusieurs études ont montré que la température a un effet sur la croissance du bananier (Ganry and Sioussaram 1978) et l'effet de l'altitude mis en évidence dans cette expérimentation confirme cette tendance. Par contre, la somme des températures ne permet pas de prédire (voir la revue des travaux sur l'effet des températures sur le par Turner et al. (2007)). Comme nous l'avons vu dans la partie 3.1.6, l'hypothèse de stabilité des quel que soit le cycle n'est pas remise en cause par les données, malgré l'alternance d'une saison sèche et humide. L'absence d'effet des degrés jours et le faible effet du stress hydrique qui ont été mis en évidence sont cohérents avec ce résultat.

### 3.1.7 Conclusion

Nous avons montré qu'il était possible de valoriser des données acquises à des fins de gestion dans une bananeraie pour faire un diagnostic agronomique *a posteriori*. Pour cela, nous avons utilisé une modélisation probabiliste. Basé sur la représentation de la séquence de développement des pieds de bananier, le modèle synthétise chaque courbe de floraison par 9 paramètres, qui ont été choisis pour leur signification agronomique. Dans cette étude, la durée des cycles de floraison ( ) a été étudiée. Ce paramètre est

une des composantes d'élaboration du rendement et son estimation offre des perspectives opérationnelles intéressantes pour les producteurs. Dans une démarche expérimentale classique, le modèle et le plan d'expérience sont construits *a priori*. La démarche présentée dans cette étude diffère de ce schéma et se rapproche plutôt d'une étude observationnelle. Ce sont des données de production qui ont été mobilisées, des données massives mais élaborées avec un autre objectif que la modélisation. En particulier, les données nécessaires à la qualification de la qualité du modèle et de l'estimation quantitative du rendement de chaque pied étaient manquantes. Nous avons donc proposé des critères de validation de l'estimation de rendement basés sur la cohérence spatiale de l'estimation et la cohérence par rapport aux ordres de grandeurs trouvés par d'autres études. L'analyse des sources de variabilité du rendement tel qu'il peut être estimé par le modèle a mis en évidence un effet de l'environnement et des pratiques culturales. Les tendances mises en évidence sont similaires à celles obtenues lors d'expérimentations agronomiques pour l'effet de l'altitude et de l'irrigation. Cette étude a aussi montré un effet du cultivar, effet que l'on suppose local, dû soit au mode de décision des planteurs, soit aux caractéristiques biologiques du cultivar. La disponibilité des données sur l'état hydrique des parcelles pourrait permettre de mieux évaluer l'effet d'un défaut d'irrigation sur le rendement. De la même manière, des données sur les propriétés agronomiques des sols permettraient de prendre en compte l'effet de la variabilité pédologique. Le développement des systèmes d'information en production ainsi que la plus grande accessibilité économique des capteurs permettant d'obtenir des informations sur le milieu (état hydrique, type de sol) pourraient permettre dans l'avenir de généraliser ce type d'approche en agriculture.

## 3.2 Variabilité de la durée moyenne des cycles de développement à l'échelle intra-parcellaire

### 3.2.1 Introduction

Dans cette section est présentée une extension de l'étude présentée en Section 3.1. Il s'agit ici d'évaluer comment le même modèle probabiliste peut être utilisé à une échelle intra-parcellaire afin d'étudier la mise en place de l'asynchronisme. On ne reprendra pas la description du modèle ni les notations utilisées pour décrire ses variables et ses paramètres, qui resteront identiques. De même, on ne reviendra pas sur la procédure d'ajustement utilisée pour estimer les paramètres du modèle à partir d'un jeu de données. Dans la section précédente, la variabilité de la durée des cycles de floraison a été étudiée à l'échelle d'une plantation avec comme plus petite unité la parcelle. Cette étude a montré qu'au sein d'une plantation la variabilité du rendement était majoritairement structurée spatialement et qu'une partie de la variabilité s'expliquait par des facteurs du milieu et de l'itinéraire technique (altitude, cultivar, irrigation). Faute d'une géolocalisation précise des observations de floraisons, la variabilité intra-parcellaire de rendement n'a pas été étudiée. Les pieds avaient été considérés spatialement indépendants, c'est-à-dire sans corrélation spatiale entre eux. C'est une hypothèse que nous avons voulu tester avec un jeu de données sur lesquelles cette fois les floraisons étaient géolocalisées, pour comprendre comment se met en place l'asynchronisme au sein des parcelles.

D'après Tixier et al. (2004) plusieurs facteurs sont susceptibles d'agir sur l'asynchronisme. Ces auteurs identifient la variabilité naturelle entre les plants et la variabilité des pratiques des producteurs au moment de la sélection des rejets ou de la récolte des régimes. Ils identifient aussi la variabilité spatiale des facteurs environnementaux de la parcelle qui modifient la nutrition hydrique ou minérale ainsi que le parasitisme. On s'attend donc à ce qu'il y ait une dépendance spatiale entre les pieds et donc une auto-corrélation spatiale de rendement au sein des parcelles ce qui, à notre connaissance, n'a jamais été montrée dans la littérature scientifique.

Pour étudier la variabilité intra-parcellaire de rendement, ce dernier est estimé dans cette section en utilisant la même procédure que celle présentée Section 3.1 pour des Courbes de Floraisons (CF) mesurées à plusieurs endroits d'une parcelle. Ces CF obtenues à l'intérieur des parcelles possèdent moins d'observations. C'est une difficulté pour estimer les neuf paramètres du modèle. Des précautions ont donc été

prises pour diminuer le nombre de paramètres à estimer et assurer une estimation robuste. La variance des estimés est ensuite étudiée pour mettre en évidence la part d'auto-corrélation spatiale dans la variance totale. L'hypothèse que nous faisons est que l'auto-corrélation spatiale s'explique par l'effet de facteurs eux mêmes structurés dans l'espace. Cela inclut les facteurs liés à la composition du sol (texture, profondeur, composition) et au micro-climat. Nous supposons par ailleurs que la variabilité non structurée spatialement provient des facteurs non structurés dans l'espace où des facteurs dont la portée spatiale est très faible. Cela inclut la compétition entre les pieds pour les ressources et l'aléa des opérations culturales (méthode de sélection du rejet, moment de coupe du régime, etc.). Nous verrons que l'éclaircissement joue également un rôle qui peut être structuré spatialement du fait de la disposition des parcelles.

### 3.2.2 Matériels et Méthodes

#### Parcelle d'étude

La parcelle étudiée est située dans la plantation commerciale appelée Plantation du Haut Penja (PHP) localisée à Njombe au Cameroun (WGS84, E : 4.612, N : 9.639). Cette parcelle faisait partie d'une expérimentation dans laquelle les inflorescences étaient géolocalisées par GPS. Elle a été choisie pour l'étude car c'était la seule parcelle pour laquelle il y avait un historique géolocalisé des comptages depuis sa plantation. Elle a une surface de 5 ha et est délimitée par une piste qui permet le passage des tracteurs.

#### Enregistrement des dates de floraison

Les nouvelles floraisons ont été comptées plusieurs fois par semaine par des ouvriers comme cela se fait classiquement dans la production bananière. Chaque nouvelle inflorescence a été géolocalisée avec un smartphone et marquée avec une lanière pour éviter qu'elle ne soit comptée deux fois. La précision de la localisation des inflorescences a été estimée à 5 m.

#### Prétraitement des données

Le nombre de fleurs comptées a été sommé par semaine et associé à la date du mercredi. La position GPS des observations n'a pas toujours été enregistrée par le smartphone à cause d'un défaut de l'application au moment de cette expérimentation. Les semaines où moins de 90 % des observations ont été géolocalisées ont été éliminées des données utilisées, pour ne pas biaiser l'estimation de  $\rho$ .

#### Analyses statistiques

La procédure d'ajustement du modèle de floraison présentée dans la Section 3.1 a été appliquée pour estimer les paramètres de floraison de la CF. Il est important de noter que la période de production PP de cette parcelle était courte (1 an et demi) et qu'en conséquence nous n'avons pas jugé pertinent de prendre en compte un phénomène de mortalité des plants. Les deux paramètres de mortalité du modèle ont donc été fixés à l'infini. La procédure d'ajustement a permis de calibrer les 7 paramètres restants du modèle, dont le  $\rho$  et le nombre de pieds total  $N$ . La densité de plantation  $D$  a été calculée comme le ratio entre le  $N$  estimé et la surface de la parcelle.

Afin d'évaluer la variabilité spatiale de  $\rho$ , la parcelle a été divisée selon un maillage régulier constitué de mailles carrées de 15 m de largeur (Figure 3.10). Pour chaque  $U$ , toutes les fleurs comptées dont les coordonnées GPS étaient situées à l'intérieur ont été extraites pour former une Courbe de Floraison notée  $C_U$ . Les  $C_U$  contenant moins de 80 floraisons n'ont pas été conservées dans l'étude car nous avons considéré qu'en dessous de cette limite, la quantité d'information n'était pas suffisante pour estimer valablement  $\rho$ . Par ailleurs, la densité de plantation étant homogène sur la parcelle, le nombre de pieds de chaque  $U$  a été estimé comme le produit de la surface de chaque  $U$  par  $D$ . Finalement, pour chaque  $U$  la procédure d'ajustement a été appliquée et les 6 paramètres du modèle ont été évalués. Le  $\rho_U$  estimé pour chaque  $U$  a été associé aux coordonnées du centre du  $U$  correspondant.

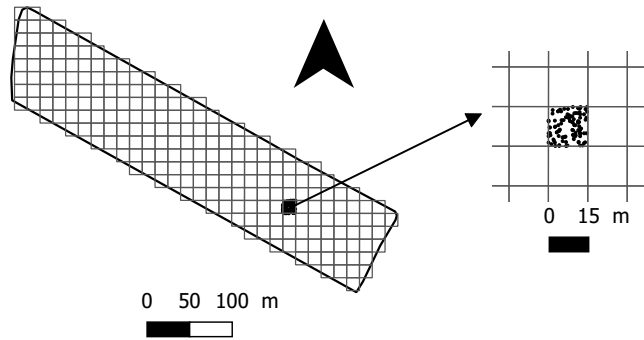


FIGURE 3.10 – Sélection des floraisons dans des mailles carrés d'une largeur de 15 m

Le variogramme de  $\gamma(h)$  noté  $\gamma$  a été calculé selon l'équation :

$$\gamma(h) = \frac{1}{2n} \sum_{i,j} (z_i - z_j)^2 \quad (3.18)$$

Dans cette équation  $z_i$  et  $z_j$  sont les valeurs de  $z$  estimés aux endroits  $i$  et  $j$ ,  $h$  désigne la position du centre d'un  $h$  et  $n$  représente la distance qui sépare deux estimations de  $z$  différentes. Le terme  $n$  représente le nombre de  $z$  qui peuvent être comparés à une distance  $h$ . L'ensemble des  $\gamma(h)$  obtenus en prenant une séquence de  $h$  a formé le variogramme expérimental. Le calcul du variogramme expérimental a été fait en utilisant 5 valeurs pour  $h$  allant de 15 m à 75 m. La distance minimale (15 m) a été imposée par l'écartement minimum entre les  $z_i$  et la distance maximale (75 m) a été choisie pour éviter de privilégier un angle d'analyse spatiale particulier sachant que la parcelle avait une forme allongée.

La variance pépité  $\sigma^2$  correspond à la variance non structurée spatialement et a été estimée par la valeur du variogramme au plus petit  $h$ . La variance du palier  $\sigma^2_s$  est la somme de la variance pépité  $\sigma^2$  et de la variance du palier partiel  $\sigma^2_p$  qui correspond à la variance structurée spatialement. La somme  $\sigma^2_s$  a été estimée par la valeur du variogramme au  $h$  le plus élevé. La proportion de variance structurée spatialement a été évaluée par l'indice  $I$  :

$$I = \frac{\sigma^2_p}{\sigma^2_s} \quad (3.19)$$

Le variogramme empirique ne permet pas d'étudier l'auto-corrélation spatiale pour des  $h$  inférieurs à 15 m qui correspondent à la largeur des mailles. Plusieurs types de variogrammes théoriques ont été ajustés au variogramme expérimental : un variogramme sphérique, un variogramme exponentiel et un variogramme gaussien. Ces variogrammes ont des allures différentes pour des petites distances. Le choix du variogramme théorique qui représente le mieux la structure spatiale est discuté. Pour chacun d'entre eux une variance pépité  $\sigma^2$ , un palier partiel  $\sigma^2_p$  et une portée spatiale ont été évaluées. La portée spatiale correspond à la distance à partir de laquelle il n'y a plus de corrélation spatiale entre les  $z_i$ . La portée effective de chacun des variogrammes, notée  $h_{95}$  a été calculée comme la distance à partir de laquelle 95 % de  $\sigma^2_s$  était atteinte. La portée effective a permis de comparer la portée de chacun des variogrammes théoriques de manière plus fiable qu'en utilisant la portée non corrigée (McBratney and Pringle 1999).

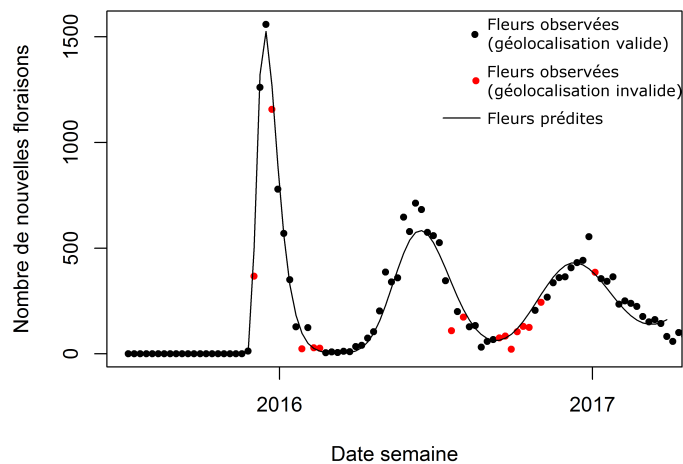


FIGURE 3.11 – Courbe de floraison de la parcelle et son ajustement. Les floraisons observées dont la géolocalisation est valide correspondent aux floraisons des semaines où au moins 90 % des observations étaient géolocalisées. La courbe des fleurs prédites correspond à la courbe modélisée ajustée.

### Cartographie de la variabilité de CD

Les ont été interpolés par krigeage sur une grille de points régulièrement espacés de 1 m dans l'enveloppe géographique de la parcelle. Pour cela, chacun des variogrammes théoriques ajustés a été utilisé, ce qui a permis de produire 3 cartes.

### Logiciel utilisé

L'ensemble de cette étude a été réalisée avec le logiciel R (R Core Team 2018). Le calcul du variogramme expérimental, l'ajustement des variogrammes théoriques et le krigeage ont été réalisés en utilisant les packages « sp » (Pebesma and Bivand 2005) et « gstat » (Pebesma 2004).

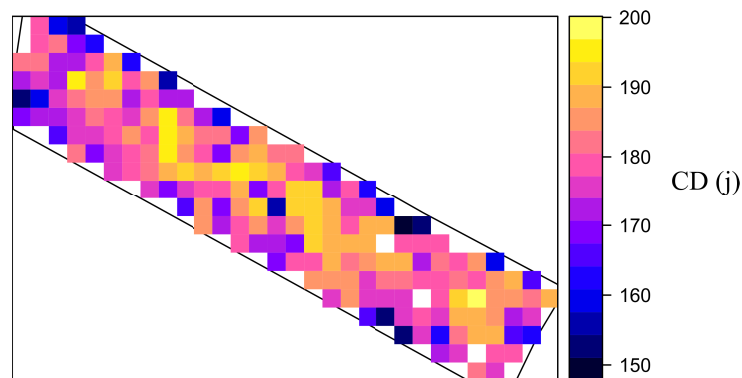
## 3.2.3 Résultats et discussion

### Présentation de la courbe de floraison de la parcelle

La CF de la parcelle entière est donnée Figure 3.11 ainsi que la courbe ajustée par le modèle. On constate que le modèle s'est bien ajusté ce qui avait déjà été montré par l'étude précédente sur un plus grand nombre de parcelles (Section 3.1). Les paramètres ajustés du modèle sont donnés Table 3.5. Parmi ces paramètres, correspond à et vaut 180 j. Le nombre total de pieds estimé par le modèle est de 6779 pieds ce qui donne un de 1360 pieds/ha. Cette densité est un peu plus faible que la consigne de plantation qui était de l'ordre de 1450 pieds/ha. Nous avons vérifié que le estimé était cohérent avec le nombre de fleurs du premier pic de floraison et du deuxième pic de floraison qui sont facilement identifiables (6604 et 6953 floraisons, respectivement). Nous avons donc considéré que le et estimés étaient cohérents et nous avons fixé le nombre de pieds de chaque à 31 pieds.

TABLE 3.5 – Paramètres ajustés pour la CF de la parcelle

Paramètre	Description	Valeur	Ajustement
	Espérance de la durée de la phase Juvénile (J)	-42 j	Ajusté
	Ecart-type de la durée de la phase Juvénile (J)	0 j	Fixé
	Espérance de la durée entre l'apparition de deux rejets successifs	180.05 j	Ajusté
	Ecart-type de la durée entre l'apparition de deux rejets successifs	29.55 j	Ajusté
	Coefficient de correction de pour le premier cycle de floraison	0.003	Ajusté
	Espérance de la durée de la phase entre l'apparition d'un rejet et la floraison du pied	25.58 j	Ajusté
	Ecart-type de la durée de la phase entre l'apparition d'un rejet et la floraison du pied	14.03 j	Ajusté
	Paramètre de décalage de la loi de survie des pieds	0 j	Fixé
	Paramètre de forme de la loi de survie des pieds	+ Inf	Fixé
	Paramètre d'échelle de la loi de survie des pieds	+ Inf	Fixé
	Nombre de pieds dans la parcelle	6779 bananiers	Ajusté

FIGURE 3.12 – CD estimé par  $\hat{\mu}$ , unité en jours. Les carrés en blanc correspondent aux  $\hat{\mu}_{ij}$  où le nombre d'observation était inférieur à 80



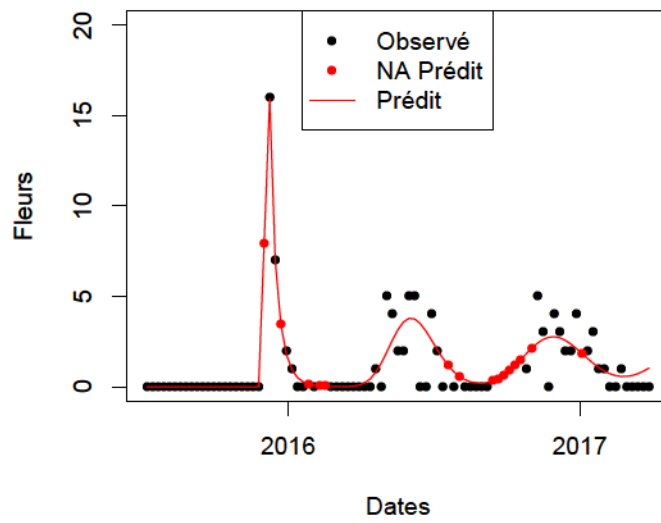


FIGURE 3.13 – Exemple d'ajustement du modèle à la courbe de floraison d'une maille. Les fleurs observées sont symbolisées par des points. Les données des semaines pour lesquelles la géolocalisation était déficiente (comptages géolocalisés inférieurs à 90 %) n'ont pas été pris en compte dans l'ajustement par le modèle (NA). Les valeurs correspondantes (points en rouge) ont pu en revanche être estimées par le modèle ajusté.

### Estimation de $\mu$ dans chaque maille de la grille

Un exemple de résultat d'ajustement d'une  $\mu$  est donné Figure 3.13. Les estimations de  $\mu$  par  $\mu$  sont représentées Figure 3.12. Dans plusieurs  $\mu$  situés le plus souvent au bord de la parcelle, le nombre de floraisons a été insuffisant pour ajuster le modèle ( $< 80$  floraisons). Ces carrés apparaissent en blanc sur la Figure 3.12. Les 3  $\mu$  qui apparaissent en blanc au centre de la parcelle sont des  $\mu$  pour lesquels la procédure d'ajustement n'a pas convergé. Au total, 193  $\mu$  ont été estimés au sein de la parcelle. La moyenne de  $\mu$  estimée en faisant la moyenne des  $\mu$  de chaque  $\mu$  est de 178 j ce qui correspond bien à la valeur évaluée par le modèle pour la parcelle entière (Table 3.5) et la variance est de 96 j<sup>2</sup>.

### Analyse variographique

Le variogramme expérimental est présenté Figure 3.14. La variance  $\sigma^2$  est de 59 j<sup>2</sup> et  $\mu$  vaut 95 j. La portée du variogramme expérimental est de 50 m. Le ratio d'auto-corrélation spatiale  $\rho$  est de 38 %. Cela montre qu'il existe bien une corrélation spatiale entre les  $\mu$  estimés dans des mailles proches, corrélation qui diminue avec la distance.

Les 3 modèles de variogrammes théoriques ajustés sont montrés Figure 3.14. L'ajustement des modèles théoriques est assez similaire entre 15 et 90 m. La portée effective  $\mu$  des variogrammes théoriques est similaire, de 41 m et 43 m pour les variogrammes sphériques et gaussien et de 50 m pour le variogramme exponentiel. Les modèles diffèrent principalement pour les distances inférieures à 15 m. Le  $\mu$  est de 0 j, 31 j et 45 j pour les variogrammes exponentiel, sphérique et gaussien, respectivement. La valeur de  $\mu$  obtenue par le variogramme exponentiel n'est pas vraisemblable du fait du caractère aléatoire de l'événement de floraison. Les  $\mu$  obtenus à partir de  $\mu$  des variogrammes gaussien et sphériques sont de 55 % et 68 %, respectivement. Cela montre que des facteurs structurés spatialement affectent la vitesse de croissance des plants, ce qui explique en partie l'asynchronisme au sein de la parcelle.

### Cartographie de $\mu$

Les cartes interpolées de  $\mu$  obtenues à partir des 3 variogrammes théoriques sont présentées Figure 3.15. Par rapport à la carte Figure 3.12 ces cartes apparaissent lissées ce qui est un résultat attendu par le krigeage. Les deux cartes obtenues à partir des variogrammes gaussien et sphériques sont très similaires. La carte obtenue à partir du variogramme exponentiel est un peu différente, avec notamment plus de bruit, c'est-à-dire plus de points avec des valeurs discordantes par rapport à leur environnement. Cela s'explique par la variance  $\sigma^2$  qui est nulle pour le variogramme exponentiel ce qui entraîne une carte faiblement lissée.

La gamme de variation du  $\mu$  est d'environ 40 jours. Les valeurs les plus faibles sont présentes plus souvent sur les bords de la parcelle, ce qui veut dire que la vitesse de développement des pieds est plus rapide. Au contraire, les valeurs les plus élevées sont présentes au centre de la parcelle. La cause la plus probable qui explique dans ce cas un  $\mu$  plus faible sur les bordures est une luminosité plus importante. Cet effet est connu par les producteurs qui observent que les premières floraisons apparaissent régulièrement sur les bordures des parcelles (Lassoudière 2007).

### Validité de l'estimation des $\mu$

Avec des  $\mu$  de cette surface, les  $\mu$  sont beaucoup moins régulières que les CF étudiées dans l'étude précédente. La taille des carrés choisie a été réduite le plus possible pour pouvoir avoir un nombre de  $\mu$  estimés suffisant pour pouvoir calculer un variogramme sachant que l'on conseille d'avoir minimum 100 points et préférentiellement autour de 150 ou 225 points (Webster and Oliver 1992). On constate cependant (i) que la procédure d'ajustement a convergé dans la majorité des cas (98 %) vers une valeur de  $\mu$  raisonnable, ce qui est un premier critère de qualité, (ii) que les  $\mu$  estimés sur des  $\mu$  voisins sont cohérents.

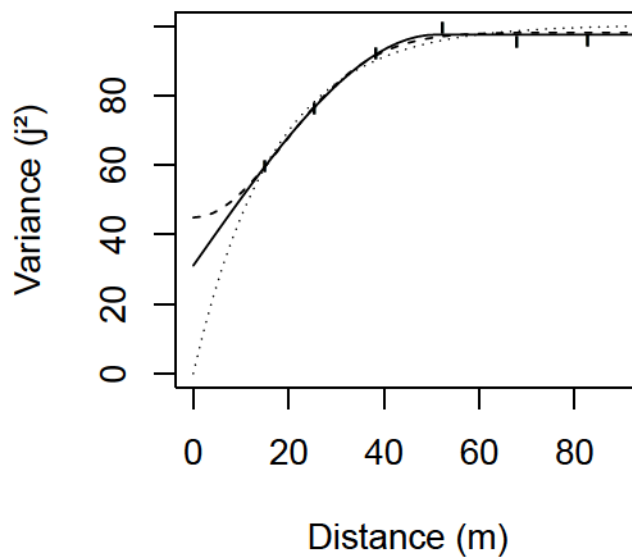


FIGURE 3.14 – Variogramme expérimental et variogrammes théoriques de *CD*. Les points correspondent au variogramme expérimental, la ligne correspond à un modèle sphérique, la courbe en tirets correspond à un modèle gaussien et la ligne en pointillés correspond à un modèle exponentiel

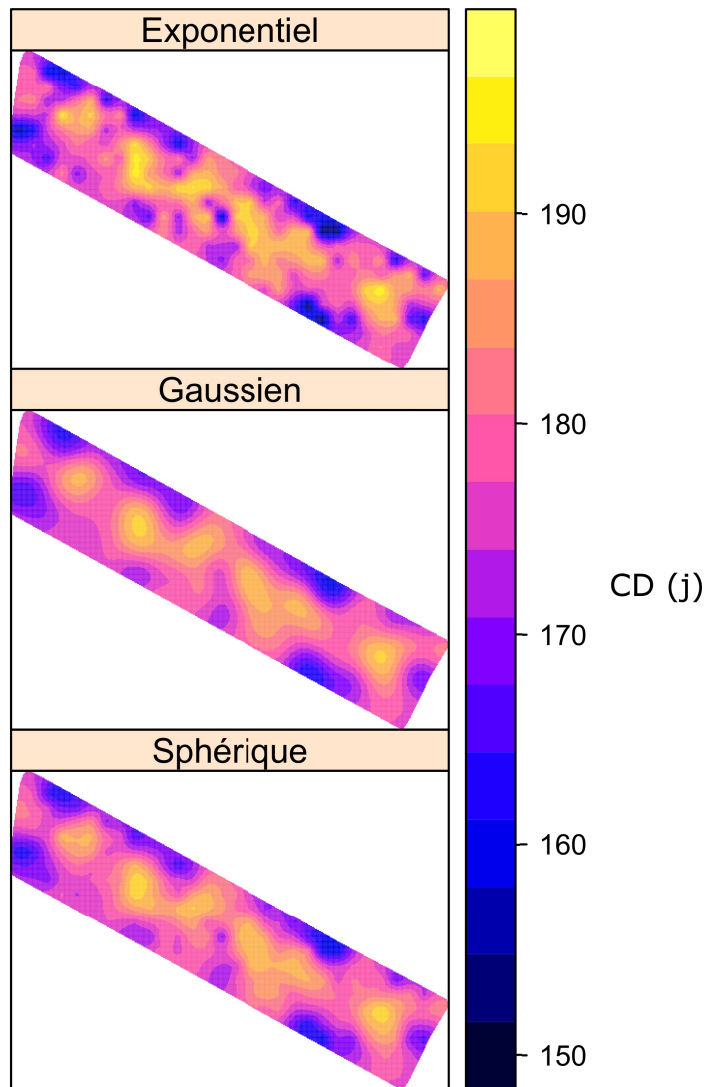


FIGURE 3.15 – Interpolation des estimés suivant les 3 modèles théoriques de variogramme (sphérique, gaussien et exponentiel)

### 3.2.4 Conclusion

Cette étude montre qu'une grande partie de la variabilité de la vitesse de croissance des pieds dans la parcelle est structurée spatialement. L'effet de bordure mis en évidence pourrait être expliqué par un meilleur accès à la lumière, c'est en tout cas l'explication corroborée par les producteurs. Cette étude montre un exemple d'utilisation du modèle en intra-parcellaire quand les floraisons sont géolocalisées mais cette étude devrait être réalisée sur un plus grand nombre de parcelles pour s'assurer du caractère général de l'auto-corrélation mise en évidence.

## 3.3 Formalisation de l'asynchronisme

### 3.3.1 Introduction

L'objectif de cette étude est de mieux formaliser ce qu'est l'asynchronisme en raisonnant à partir des équations du modèle présentées dans l'article de la Section 3.1.

Nous avons vu que le moteur de l'asynchronisme était l'hétérogénéité des conditions de croissance qui mène à une variabilité des vitesses de croissance au sein des parcelles. Nous avons aussi vu que l'asynchronisme entraîne une diversité des stades phénologiques, diversité qui augmente avec le temps depuis la plantation des parcelles. Nous supposons que la vitesse d'augmentation de la diversité des stades phénologiques dépend de la durée des cycles de floraison mais aussi de l'hétérogénéité des conditions de croissance des parcelles. La description de l'asynchronisme et de ses conséquences par des indices objectifs pourrait être intéressante pour les producteurs afin de faire un diagnostic de leurs parcelles.

L'objectif de cette étude est double. Premièrement, il s'agit de proposer 3 indices qui permettent de caractériser (i) l'hétérogénéité des conditions de croissance des parcelles (ii) le stade phénologique moyen des parcelles et (iii) la diversité des stades phénologiques. Pour cela, nous avons réutilisé les équations du modèle pour proposer ces 3 indices. Deuxièmement, il s'agit de présenter le comportement de ces indices sur les CF de parcelles représentatives de conditions de production variées.

### 3.3.2 Définition des indices

#### Indice d'hétérogénéité des conditions de croissance

L'équation 3.7 du modèle (Partie 3.1) définit le temps entre deux floraisons sur un même pied de la manière suivante :

(3.7)

La durée moyenne des cycles de floraison ( ) a été définie comme l'espérance de :

(3.8)

La variabilité des cycles de floraison s'exprime comme la variance de :

(3.20)

Cette variance ne dépend pas de et est stable dans le temps. Cette hypothèse du modèle n'a pas été remise en cause par les ajustements présentés Section 3.1 et nous avons vu que, dans un grand nombre de cas, le modèle permettait bien de reproduire les formes des CF mesurées. Nous ne remettons donc pas en cause cette hypothèse ici et considérons qu'elle est valide dans les conditions de cette plantation.

L'indice d'hétérogénéité choisi, noté , est écrit :

(3.21)

Nous avons choisi de transformer la variable par le logarithme afin de normaliser sa distribution.

### Indice de stade phénologique moyen et de diversité des stades phénologiques

Afin de définir le stade phénologique moyen et la diversité des stades phénologiques à l'échelle parcellaire nous avons tout d'abord défini l'indice d'avancement de chaque bananier qui permet de représenter son stade phénologique en fonction de sa date de floraison passée et de sa date de floraison future. L'indice d'avancement de chaque bananier à une date , noté , est défini par l'équation suivante :

$$\text{---} \quad (3.22)$$

Avec l'avancement du pied à une date , la date de floraison future du bananier en place à la date et la date de floraison précédente, c'est-à-dire la date de floraison du bananier parent. Si le bananier correspond au vitro-plant, c'est-à-dire au premier plant qui est planté sur la parcelle, alors on considère que est la date de plantation de la parcelle. Par construction, est donc compris entre 0 et 1. Cet indice tend vers 1 si le bananier est tout proche de fleurir et est proche de 0 si le bananier vient tout juste de fleurir.

La distribution des a été étudiée pour définir le stade phénologique moyen et sa variabilité à l'échelle parcellaire. Pour cela, les théories développées dans le domaine des statistiques directionnelles ont été utilisées. Ce domaine permet d'étudier des données comme l'heure où les angles qui ne peuvent pas être étudiées comme d'autres données linéaires. En effet, des estimateurs comme la moyenne ou la variance n'ont plus de sens, par exemple parce que la moyenne entre deux angles valant  $359^\circ$  et  $1^\circ$ , ne vaut pas  $180^\circ$  mais  $0^\circ$ . Les statistiques circulaires permettent de calculer des moyennes et des dispersions sans ces effets de seuil, ce qui a permis d'étudier la distribution des d'une parcelle sans l'effet de seuil lié au passage de la floraison. Pour cela, nous avons placé les sur un cercle qui a pour origine 0 et pour circonférence 2 . Avec 0 qui correspond à un avancement de 0 et 2 qui correspond à un avancement de 1. Cette représentation est illustrée Figure 3.16. La variable peut s'écrire sous la forme d'un nombre complexe :

$$(3.23)$$

Avec l'angle du point sur le cercle.

Le centre de gravité de l'ensemble des points du cercle est noté et correspond à :

$$\text{---} \quad (3.24)$$

L'écriture complexe de est la suivante :

$$(3.25)$$

Avec , , la distance entre le centre du cercle et le centre de gravité et l'angle du centre de gravité. La variance circulaire notée se calcule par l'équation suivante :

$$(3.26)$$

Lorsque est situé sur le bord du cercle, , lorsque est situé au centre du cercle, est maximal et vaut 1.

Les deux indices représentant respectivement le stade phénologique moyen et la diversité des stades phénologiques sont et . Une représentation de ces deux indices est donnée Figure 3.16.

### 3.3.3 Matériel et Méthodes

#### Données

Les données de l'étude présentée dans l'article Section 3.1 ont été reprises ainsi que les résultats des ajustements du modèle. Parmi ces données, les 1188 CF qui avaient été jugées fiables et les paramètres

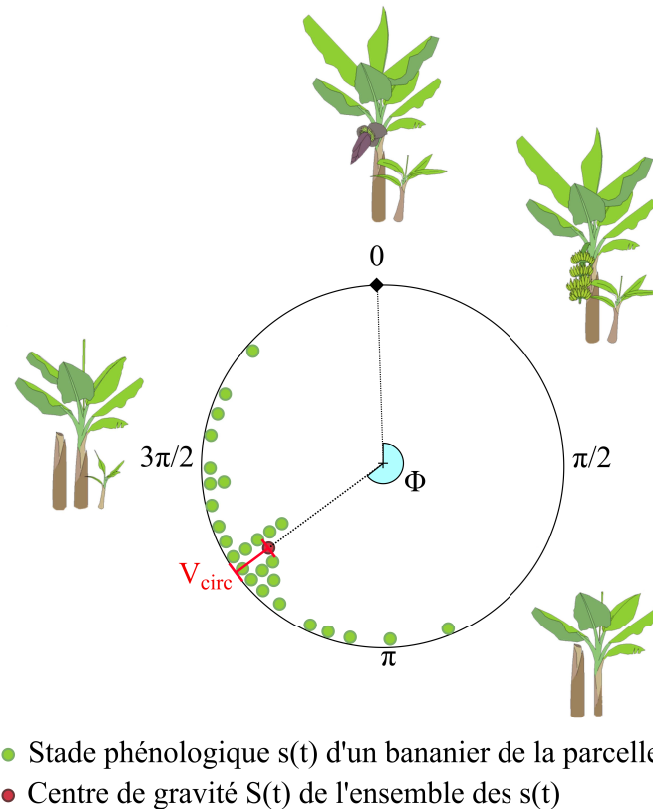


FIGURE 3.16 – Représentation de l'avancement (points verts) des bananiers sur un cercle. Pour simplifier le schéma les points sont représentés superposés alors qu'ils devraient tous être placés sur le périmètre du cercle. Le stade phénologique moyen de la parcelle est donné par l'angle  $\Phi$  et la variabilité est obtenue à partir de la distance la plus courte entre le centre de gravité des  $s(t)$  et le périmètre du cercle.

des ajustement associés ont constitué la base de données de cette étude. Une nouvelle échelle des temps des CF a été calculée pour exprimer le temps avec comme origine le moment de la première floraison au lieu du moment de la plantation. Cette nouvelle échelle de temps est plus pratique pour comparer des CF qui ont des caractéristiques différentes et dont la date de plantation était parfois mal renseignée (Voir Section 3.1.3).

### Choix des CF et calcul des indices

Pour chacun des ajustements,  $\mu$  a été calculé selon l'équation 3.21. Ces CF ont ensuite été triées selon les valeurs de  $\mu$  et  $\sigma$  estimées par l'ajustement. Quatre CF ont été choisies pour représenter le comportement de courbes de floraisons correspondant à des situations contrastées de production en termes de  $\mu$  et de  $\sigma$ . Pour cela, deux classes ont été définies pour  $\mu$  et pour  $\sigma$  : une classe qui réunissait 15 % des valeurs les plus faibles et une classe qui réunissait 15 % des valeurs les plus hautes. Cela a permis de définir 4 classes : Classe 1 :  $\mu$  courte – fort ; Classe 2 :  $\mu$  courte – faible ; Classe 3 :  $\mu$  longue – élevé ; Classe 4 :  $\mu$  longue – faible. Dans chacune de ces 4 classes nous avons choisi une CF pour laquelle le modèle s'était bien ajusté et qui avait une PP d'au moins 1400 j. Ces CF ont ensuite été utilisées pour étudier  $\mu$  et  $\sigma$  afin d'étudier le comportement de ces indices dans ces 4 conditions de production.

Le calcul de la variable  $\mu$  à partir de laquelle sont calculés les indices  $\mu$  et  $\sigma$  utilise les dates de floraison  $t_i$  de chacun des pieds des parcelles (Équations 3.25 et 3.26). Or les  $t_i$  ne sont pas connus, seul le nombre de floraisons global des parcelles est observé ( $N$ ). Par contre le modèle peut simuler les  $t_i$  de  $N$  pieds tel que la somme correspond (en espérance) à la CF ajustée. Nous avons utilisé cette fonctionnalité pour simuler les dates de floraison  $t_i$  nécessaires au calcul de  $\mu$ ,  $\sigma$  et  $\rho$  pour chaque CF. L'évolution temporelle de ces 3 indices a ensuite été calculée à chaque pas d'une séquence de  $\Delta t$ . Cette séquence a été initiée à la date de plantation et a été terminée 1200 j après la date de première floraison. La durée entre chaque  $t_i$  de la séquence temporelle a été fixée à 7 j. Cette procédure a été répétée 100 fois pour chaque CF afin de calculer l'intervalle de confiance à 95 % de ces 3 indices.

### 3.3.4 Résultats et discussion

#### Description des CF choisies

La Figure 3.17 montre la distribution des  $\mu$  et de  $\sigma$  sur l'ensemble des 1188 CF de la base de données. Les points situés à gauche du graphique correspondent aux CF où les  $\mu$  sont courtes tandis que les points situés à droite correspondent aux CF avec des  $\mu$  plus longs. Les points situés vers le bas de l'axe des ordonnées correspondent aux CF où  $\sigma$  est faible et les points situés vers le haut correspondent aux CF où  $\sigma$  est élevé. Les limites de classes sont représentées par des droites. Les 4 CF choisies pour représenter des situations contrastées de la base de données en termes de  $\mu$  et de  $\sigma$  sont indiquées en rouge.

On remarque que le nuage de points ne présente pas de tendance particulière ce qui signifie qu'il n'y a pas de corrélation entre  $\mu$  et  $\sigma$ . Cela montre que, quelle que soit la durée des cycles de floraison dans les parcelles, l'hétérogénéité des conditions de croissance peut-être faible ou forte. On suppose que les conditions générales du pédo-climat et du système de culture (système d'irrigation, cultivar) agissent sur  $\mu$  à l'échelle des parcelles indépendamment des conditions plus locales qui agissent sur l'hétérogénéité, comme la pente, la variabilité du sol à l'intérieur de la parcelle, l'effet des bordures, etc. L'étude précédente (Section 3.2) a mis en évidence que la variabilité intra-parcellaire de  $\mu$ , représentée par l'indice  $\rho$  dans cette étude, présentait une structure spatiale. Les deux composantes de la variance (la variance pépète  $\rho$  et la variance du palier partiel  $\sigma$ ) peuvent expliquer la variabilité de  $\mu$ . En disposant d'observations géolocalisées des floraisons sur un plus grand nombre de parcelles il serait possible d'étudier si une augmentation de  $\mu$  est due principalement à une augmentation de  $\rho$  ou plutôt à une augmentation de  $\sigma$ . Dans les conditions actuelles c'est-à-dire sans géolocalisation des floraisons, cet indice pourrait déjà



Paramètre	Description	CF1	CF2	CF3	CF4
	Espérance de la durée de la phase Juvénile (J)	4.34	-1.98	-116.86	-13.85
	Ecart-type de la durée de la phase Juvénile (J)	0	0	0	0
	Espérance de la durée entre l'apparition de deux rejets successifs	181.97	188.22	247.39	239.31
	Ecart-type de la durée entre l'apparition de deux rejets successifs	30.20	29.72	40.38	30.33
	Coefficient de correction de pour le premier cycle de floraison	0.42	0.38	0	0.51
	Espérance de la durée de la phase entre l'apparition d'un rejet et la floraison du pied	35.57	0	135.00	0
	Ecart-type de la durée de la phase entre l'apparition d'un rejet et la floraison du pied	25.38	0.12	21.29	0.18
	Paramètre de décalage de la loi de survie des pieds	0	0	0	0
	Paramètre de forme de la loi de survie des pieds	0.05	2.29	+ Inf	4.89
	Paramètre d'échelle de la loi de survie des pieds	0	1862	+Inf	1853
	Nombre de pieds encore vivant à la date de première floraison	7234	10526	2530	2855
	Indice d'hétérogénéité de la parcelle	7.72	6.78	7.84	6.86

TABLE 3.6 – Paramètres de chacune des 4 CF choisies

être utilisé par les producteurs pour identifier les parcelles très hétérogènes sur lesquelles un diagnostic pourrait être mené pour savoir s'il est possible d'homogénéiser la vitesse de développement des pieds.

L'examen des CF avec des  $\alpha$  supérieurs à 8.5 a montré que celles-ci étaient aberrantes (résultat non montré), soit parce qu'elles correspondaient à des parcelles plantées en plusieurs fois, soit parce que la date de plantation réelle était bien antérieure à la date de plantation renseignée. Cet indice pourrait donc être utilisé pour évaluer la fiabilité des CF.

Les 4 CF choisies sont représentées Figure 3.18 et les paramètres de ces courbes ainsi que  $\alpha$  sont donnés Table 3.6. CF1 et CF2 ont un  $\alpha$  court d'environ 185 j et présentent environ 7 pics de floraison pendant la période d'observation de 1200 j. CF3 et CF4 présentent un  $\alpha$  plus long d'environ 243 j et présentent 5 pics.

### Représentation du comportement des indices d'asynchronicité

Un exemple de simulation des  $\alpha$  correspondant à la CF4 est donné Figure 3.19. Par construction, si tous les points  $\alpha$  sont situés au même endroit sur le cercle leur centre de gravité  $\alpha$  est alors positionné sur le bord du cercle. Cela correspond à une diversité des stades phénologiques nulle et dans ce cas  $\alpha$  et tous les bananiers sont au même stade phénologique. Si tous les points  $\alpha$  sont distribués uniformément sur le cercle, leur centre de gravité est situé au centre du cercle et il n'y a pas d'orientation privilégiée. Dans ce cas la diversité des stades phénologiques est maximale  $\alpha$  et la notion de stade phénologique moyen n'est plus pertinente car aucun stade n'est plus représenté que les autres.

L'évolution de  $\alpha$  pour chacune des CF choisie est représentée Figure 3.20. Les courbes de  $\alpha$  sont croissantes tout au long de la PP des parcelles. L'indice proposé permet bien de représenter l'évolution attendue de la diversité sans effet de seuil lié au passage de la floraison.

L'évolution de  $\alpha$  présente une inflexion vers 200 j pour CF3 et CF4 puis l'évolution est plus régulière. Il existe aussi une inflexion moins visible sur le graphique pour CF2 vers 5 j. Cette inflexion est due à la particularité du premier cycle de floraison. Nous avons vu par simulation que l'allure de cette

inflexion dépendait du temps entre la plantation et la première floraison et du paramètre du modèle  $\beta$ . Cette inflexion peut même avoir une allure en cloche avec une décroissance dans les cas où la durée entre la plantation déclarée et la première floraison est très courte. Après cette inflexion, la pente des courbes diminue au cours du temps jusqu'à ce que  $\beta$  se rapproche de 1.

On observe, toujours sur la Figure 3.20 que la courbe de  $\beta$  correspondant à la CF1 est supérieure à toutes les autres. Cela montre que la diversité des stades augmente plus rapidement que pour les autres parcelles. Ce résultat est cohérent avec les caractéristiques de cette CF dont le  $\beta$  est court et pour laquelle  $\beta$  est élevé. La courbe de  $\beta$  correspondant à la courbe CF4 est inférieure à toutes les autres. Cela est aussi cohérent avec le fait que son  $\beta$  est long et son  $\beta$  plus faible que les autres CF. Ces résultats montrent que la diversité des stades phénologiques n'augmente pas à la même vitesse dans toutes les parcelles, autrement dit que les parcelles ne se désynchronisent pas toutes à la même vitesse.

Pour toutes les CF où le modèle s'est bien ajusté, on s'attend à ce que les estimations de  $\beta$  et  $\beta$  soient cohérentes. Pour les CF qui présentent des évolutions irrégulières à cause d'une mortalité soudaine ou d'un changement de pratique on s'attend à ce que la méthode soit moins fiable. Notamment, il est possible que des parcelles se resynchronisent après une mortalité soudaine ce qui ne peut pas être simulé par la procédure présentée ici. Cette approche donne donc des tendances générales mais pourrait être améliorée.

## Conclusion

Les équations du modèle de floraison ont été utilisées pour calculer un indice de variabilité de la durée des cycles de floraison au sein des parcelles. On suppose que cet indice traduit l'hétérogénéité intra-parcellaire des conditions de croissance. Nous avons montré que cet indice était indépendant de la durée moyenne des cycles de floraison. Deux indices ont aussi été proposés pour représenter le stade phénologique moyen et la diversité des stades phénologiques d'une parcelle. Nous avons vu que la vitesse à laquelle les pieds se désynchronisent est variable d'une parcelle à l'autre. Ces outils pourraient servir aux producteurs à mieux caractériser l'état de développement des parcelles à partir de leurs données de floraison.

## 3.4 Évolution de la structure spatiale des stades phénologiques au cours du temps

### 3.4.1 Introduction

Nous avons vu que la vitesse de développement des pieds pouvait être structurée spatialement à l'échelle intra-parcellaire. Cela a été démontré dans la Section 3.2 en étudiant la durée des cycles de floraisons  $\beta$  des bananiers situées dans des sous-ensembles carrés de la parcelle appelés mailles et notés  $\beta$ . En conséquence, on en déduit que les stades phénologiques seront aussi structurés spatialement lors du premier cycle de développement des bananiers de la parcelle. L'objectif de cette section est de montrer que même si la structure spatiale de  $\beta$  est stable dans le temps, la variabilité des stades phénologiques sera de moins en moins auto-corrélée spatialement au fur et à mesure du temps depuis la plantation. C'est-à-dire qu'au bout d'un moment, la connaissance du stade phénologique d'un bananier à un endroit de la parcelle ne donnera plus d'information sur le stade phénologique d'un autre bananier à proximité. Pour démontrer cela, nous réutilisons les indices développés dans la section précédente qui permettent de caractériser l'asynchronisme et nous les appliquons *in silico* au cas de la parcelle présentée dans la Section 3.2. Un nouvel outil appelé le cosineogramme est nécessaire pour analyser la structure spatiale de données circulaires et est aussi présenté.

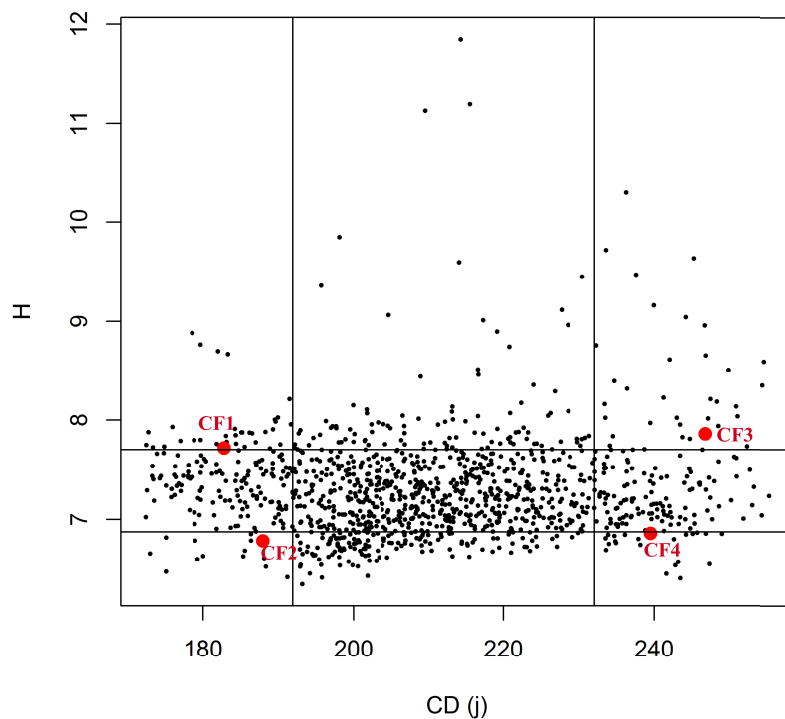


FIGURE 3.17 – Représentation de la durée d'un cycle de floraison et de l'hétérogénéité des conditions de croissance estimés par la procédure d'ajustement du modèle pour chacune des 1188 CF. Les lignes verticales correspondent aux quantiles à 15 % et à 85 % de et les lignes horizontales correspondent aux quantiles à 15 % et à 85 % de . Les points rouges correspondent aux CF qui ont été choisis pour la suite des analyses.

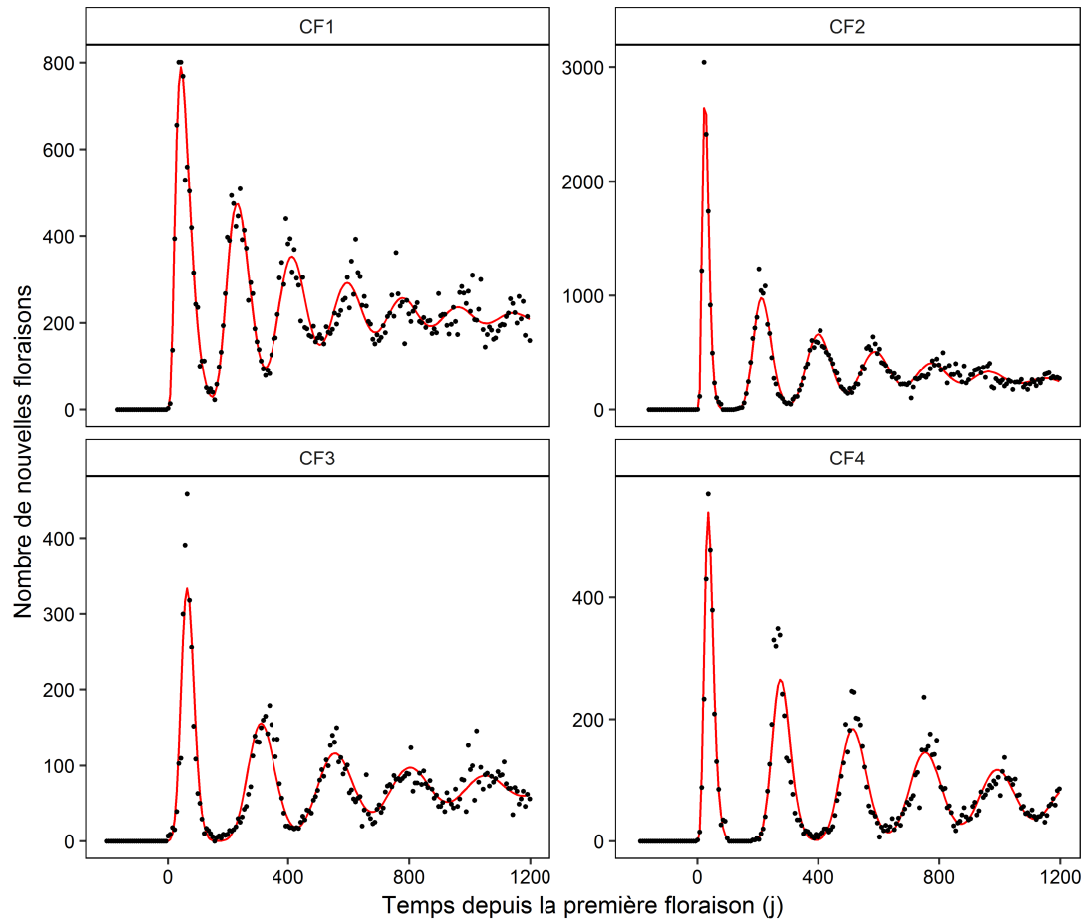


FIGURE 3.18 – Représentation des 4 CF choisies (points) et de leur ajustement par le modèle (ligne rouge).

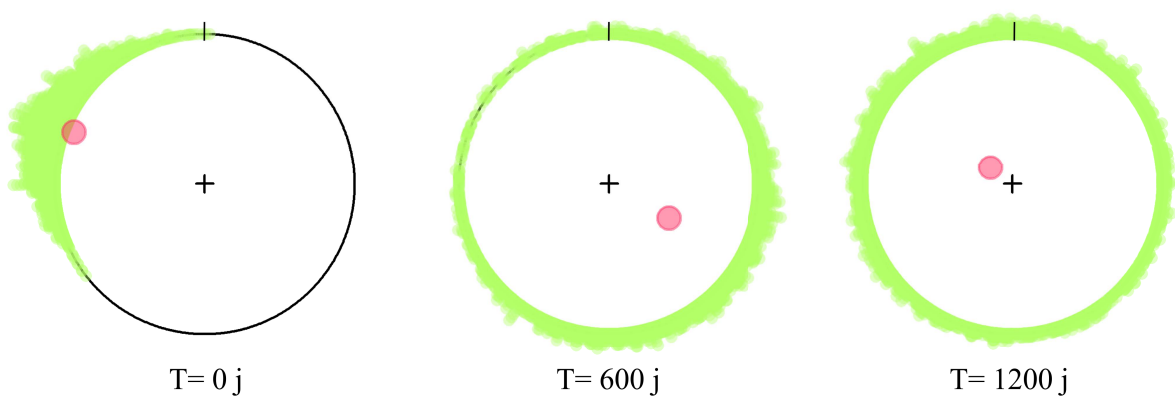


FIGURE 3.19 – Représentation de  $T=0$  j et de  $T=1200$  j à 3 temps depuis la première floraison de la CF4

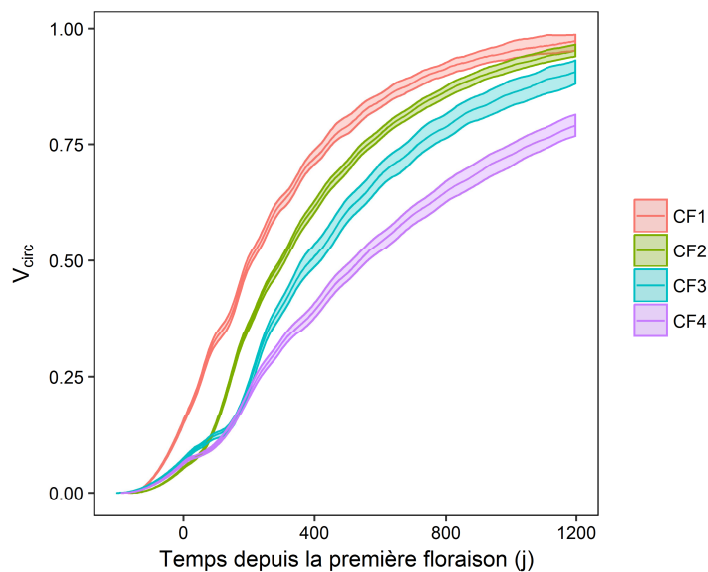


FIGURE 3.20 – Représentation de l'évolution de la diversité des stades phénologiques au cours du temps pour chacune des 4 CF. L'enveloppe à 95 % de est aussi représentée.

### 3.4.2 Analyse de la structure spatiale de données circulaires

On rappelle que le stade phénologique de chaque pied d'une parcelle,  $\theta_i$ , a été représenté sur un cercle (Figure 3.19) avec un angle  $\theta_i$  (Équation 3.23). La corrélation spatiale de données circulaires, ici les  $\theta_i$ , peut être étudiée à partir du cosineogramme inversé présenté par Morphet (2009) (« Inverted cosineogram ») qui est l'analogue du variogramme pour des données circulaires. Morphet (2009) ne présente que l'équation du cosineogramme mais l'équation du cosineogramme inversé peut-être facilement déduite de son travail et s'écrit :

$$\text{---} \tag{3.27}$$

Dans cette équation  $\theta_i$  et  $\theta_j$  sont les angles correspondants aux stades phénologiques des bananiers aux endroits  $i$  et  $j$ .  $d_{ij}$  représente la distance qui sépare deux bananiers.  $n$  représente le nombre de  $\theta_i$  qui peuvent être comparés à une distance  $d_{ij}$ . L'ensemble des  $d_{ij}$  obtenus en prenant une séquence croissante de  $d_{ij}$  forme le cosineogramme inversé. On remarque que si les angles sont en phases alors  $d_{ij}$  vaut 0, et que si les angles sont indépendants alors  $d_{ij}$  vaut 1.

A l'instar du variogramme théorique, un modèle appelé modèle cosinus peut être ajusté au cosineogramme pour caractériser la dépendance spatiale d'observations circulaires. Ce modèle peut être utilisé pour interpoler des observations circulaires par Krigeage circulaire (Morphet 2009).

### 3.4.3 Matériel et méthodes

Les résultats de la Section 3.2 sont repris dans cette étude, notamment : les paramètres ajustés de la courbe de floraison de la parcelle entière, le nombre de pieds estimé dans les mailles (31), ainsi que la carte des  $\theta_i$  interpolés par le modèle gaussien à l'intérieur de la parcelle.

La méthode utilisée consiste à simuler l'évolution au cours du temps du stade phénologique moyen de chaque maille de cette parcelle et d'étudier l'auto-corrélation spatiale des  $\theta_i$  au cours du temps. La courbe de floraison de la parcelle n'avait été observée que sur une période d'environ 1,5 années

mais afin d'évaluer l'évolution de l'auto-corrélation spatiale sur une plus grande période de temps, les simulations sont réalisées sur une période comprise entre la date de plantation de la parcelle et 1500 jours après.

Pour simuler l'évolution de  $\sigma^2$ , nous avons simulé les dates de floraisons  $t_i$  correspondantes à chacun des 31 bananiers de chaque  $U$ . Les paramètres utilisés pour ces simulations ont été les paramètres de la courbe de floraison de la parcelle entière présentés Table 3.5. Par contre, le  $\sigma^2$  de chaque  $U$  a été fixé à la valeur interpolée au niveau du centre de  $U$  par le modèle gaussien dont la carte est présentée Figure 3.15. La simulation des  $t_i$  de chacun des bananiers de chaque maille nous a permis de calculer  $\sigma^2$  et  $\rho$  en utilisant les équations 3.22 et 3.23. Le stade phénologique moyen  $\bar{t}$  de chaque  $U$  a ensuite été calculé selon les équations 3.24 et 3.25.

Le cosinogramme inversi  $\rho$  a été calculé selon l'équation 3.27 en utilisant 5 valeurs de  $r$  allant de 15 m à 75 m. Le modèle théorique gaussien a été ajusté à  $\rho$  ce qui a permis d'estimer la portée de la structure spatiale ainsi que la variance pépite  $\sigma^2$  et le palier partiel noté  $\rho_0$ . Nous avons choisi un modèle gaussien car c'est le modèle qui avait été choisi pour interpoler les  $t_i$  de cette parcelle. Le calcul de  $\rho$  et l'ajustement du modèle théorique ont été effectués pour toutes les semaines depuis la date de plantation de la parcelle jusqu'à 1500 jours.

Le ratio  $\rho$  qui correspond à la proportion de variance structurée spatialement (Équations 3.28) a été calculé pour chaque semaine et son évolution a été étudiée.

$$\rho = \frac{\sigma^2}{\sigma^2 + \rho_0} \quad (3.28)$$

Dans cette équation, la variance pépite  $\sigma^2$  correspond à  $\sigma^2$ . La variable  $r$  été évaluée à 15 m et non pas à une distance de 0 m car cela correspond à la distance minimale entre les centres de deux  $U$ . La variance  $\sigma^2$  a été obtenu par l'ajustement du modèle gaussien.

#### 3.4.4 Résultats

La Figure 3.21 montre la courbe de floraison de la parcelle ainsi que la courbe de floraison simulée. L'évolution de la diversité des stades phénologiques est aussi présentée. On voit que dès 1200 jours, la diversité maximale ( $H'$ ) est atteinte.

Quatre cosinogrammes inversis calculés à plusieurs temps en jours après la plantation de la parcelle sont présentés Figure 3.22. On observe que  $\rho$  augmente au cours du temps jusqu'à atteindre la valeur maximale de 1 à 1260 jours. Cela est attendu car la diversité des stades phénologiques augmente au cours du temps. On observe aussi que la proportion de variance structurée spatialement diminue au cours du temps. Cela est confirmé par l'évolution globale du ratio  $\rho$  pendant les 1500 jours depuis la date de plantation qui est représentée Figure 3.23. On observe que lors du premier cycle de floraison ce ratio relativement élevé ( $\rho \approx 0.8$ ) ce qui montre que les stades phénologiques sont structurés spatialement. Ce ratio diminue dès le début du deuxième cycle de floraison et il devient nul à environ 1400 jours, c'est-à-dire après le 6<sup>ème</sup> pic de floraison (Figure 3.21). À ce moment, la variabilité des stades phénologiques correspond donc à un effet pépite pur (bruit blanc). L'évolution de  $\rho$  n'est pas totalement régulière et présente notamment une augmentation autour de 1100 jours. On suppose que cela est dû à l'incertitude de l'estimation des paramètres du modèle théorique cosine par la procédure d'ajustement.

#### 3.4.5 Discussion et conclusion

L'examen des cosinogrammes de cette étude *in silico* suggère que les stades phénologiques sont uniformément répartis dans une parcelle âgée, même s'il existe une structure spatiale de la durée des cycles de développement. Ce résultat est important, car lorsque l'on étudiera une propriété du couvert qui dépend du stade phénologique et des conditions de croissance, le mélange uniforme des stades n'induera pas de structure spatiale sur la propriété étudiée. Ainsi, si une structure spatiale est mise en évidence, elle proviendra uniquement des conditions de croissance.

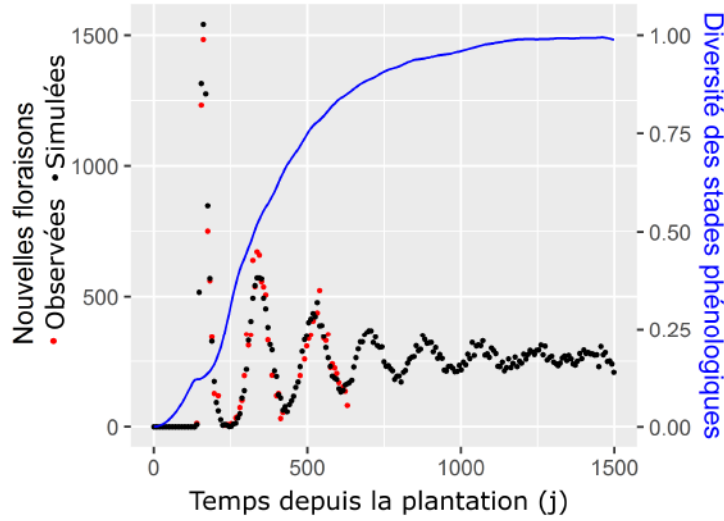


FIGURE 3.21 – Courbe de floraison de la parcelle. Les points rouges correspondent aux floraisons observées et les points noirs aux floraisons simulées. La ligne bleue représente l'évolution de la diversité des stades phénologiques au cours du temps évaluée à partir de l'indice  $V_{circ}(t)$

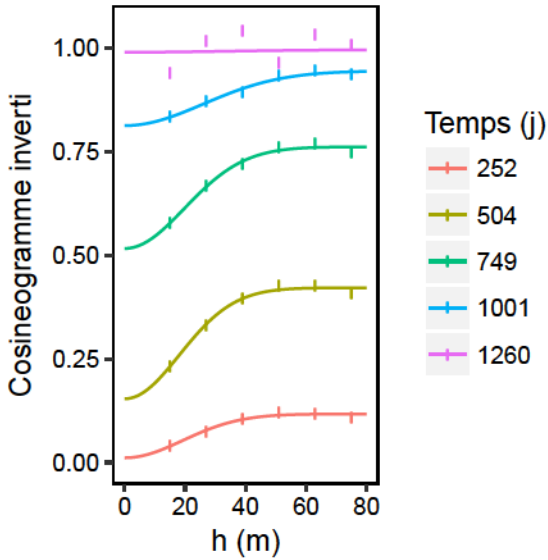


FIGURE 3.22 – Cosineogrammes inversés  $\zeta(h)$  des stades phénologiques simulés dans la parcelle pour différents temps après la plantation de la parcelle. Les points correspondent aux cosineogrammes inversés et les lignes correspondent aux modèles théoriques ajustés

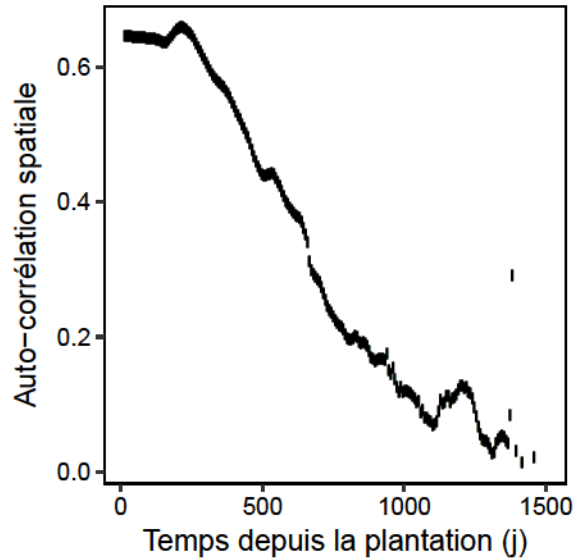


FIGURE 3.23 – Évolution du ratio d'auto-corrélation spatiale des stades phénologiques  $r$  au cours du temps depuis la plantation de la parcelle

## Chapitre 4

# Proposition d'une méthode de correction des biais liés à l'asynchronisme sur les données de récolte

Dans ce chapitre, je présente l'article que nous avons soumis en septembre 2018 au journal « Precision Agriculture ». Son objectif était de présenter la culture bananière qui est atypique dans la communauté Agriculture de Précision (AP) et de montrer que les concepts d'analyse spatiale développés sur d'autres cultures devaient être adaptés. Pour cela, nous avons repris la théorie classiquement utilisée en AP et nous l'avons légèrement modifiée pour prendre en compte la diversité des stades phénologiques et la continuité de la production dans le temps. Nous avons ensuite appliqué ces outils sur un indicateur d'intérêt pour les producteurs : le temps entre la floraison des régimes et leur maturité et nous avons démontré en quoi cette nouvelle approche était bénéfique.

### Sommaire

---

<b>4.1</b>	<b>Abstract</b> . . . . .	<b>79</b>
<b>4.2</b>	<b>Introduction</b> . . . . .	<b>79</b>
<b>4.3</b>	<b>Methodological aspects</b> . . . . .	<b>81</b>
4.3.1	Theory . . . . .	81
4.3.2	Description of the studied plant response : Time to reach Fruit Maturity ( <i>t<sub>fm</sub></i> ) . . . . .	82
4.3.3	Modeling of <i>t<sub>fm</sub></i> variability and discretization of the temporal component using cohorts . . . . .	82
4.3.4	Decomposition of the <i>t<sub>fm</sub></i> variability . . . . .	83
<b>4.4</b>	<b>Material and Methods</b> . . . . .	<b>84</b>
4.4.1	Site study and data collection process . . . . .	84
4.4.2	Pre-processing of the data . . . . .	86
4.4.3	Decomposition of the variance . . . . .	86
4.4.4	Spatial analysis . . . . .	87
4.4.5	Data mapping . . . . .	87
4.4.6	Comparison of maps to assess the variance decomposition approach . . . . .	88
4.4.7	Software used . . . . .	88
<b>4.5</b>	<b>Results and Discussion</b> . . . . .	<b>88</b>
4.5.1	General characteristics of the data . . . . .	88



4.5.2	Evaluation of the cohort effect . . . . .	89
4.5.3	Decomposition of the spatial structure . . . . .	91
4.5.4	Effect of the cohorts and of environmental factors on the maps of $t_{fm}$ . . . . .	93
4.5.5	Assessment of the variance decomposition approach . . . . .	93
4.5.6	Interpretation of $\epsilon_{res}$ . . . . .	95
4.5.7	Interpretation of $\epsilon_{res}$ maps . . . . .	96
<b>4.6</b>	<b>Conclusion</b> . . . . .	<b>97</b>
<b>4.7</b>	<b>Acknowledgements</b> . . . . .	<b>98</b>

---

## Spatial analysis and mapping of banana crop properties : issues of the asynchronicity of the banana production and proposition of a statistical method to take it into account

J. Lamour<sup>1,2</sup>, O. Naud<sup>1</sup>, M. Lechaudel<sup>3,4</sup>, G. Le Moguédec<sup>5</sup>, J. Taylor<sup>1</sup> and B. Tisseyre<sup>1</sup>

<sup>1</sup> ITAP, Montpellier SupAgro, Irstea, Univ Montpellier, Montpellier, France

<sup>2</sup> Compagnie Fruitière, Marseille, France

<sup>3</sup> CIRAD, UMR QualiSud, F-97130 Capesterre-Belle-Eau, Guadeloupe, France

<sup>4</sup> QualiSud, Univ Montpellier, CIRAD, Montpellier SupAgro, Univ d'Avignon, Univ de La Réunion, Montpellier, France

<sup>5</sup> AMAP, INRA, Univ. Montpellier, IRD, CNRS, CIRAD, Montpellier, F-34000 France

### 4.1 Abstract

Precision agriculture for banana crops has been little investigated so far. The main difficulty to implement precision agriculture methods lies in the asynchronicity of this crop : after a few cycles, each plant has its own development stage in the field. It results in a diversity of the phenological stages within a field and also a continuous production of bananas over time. Therefore, maps of agronomical interest derived from plant responses are difficult to produce using existing methods. This study proposes a mapping approach that handles the diversity of phenological stages and the temporal continuity of production. It explores the feasibility to apply this general approach to a plant response parameter which is the time between flowering and maturity (time to harvest) of banana denoted  $t_{FH}$ . The  $t_{FH}$  gives an insight into the spatial distribution of vigour. The study was conducted using a large database (more than 395000 observations) generated by two commercial farms in 2015 and 2016 in Cameroon. The temporal variability of  $t_{FH}$ , which is induced by meteorological and operational constraints, and the spatial variability, which is assumed to be due to environmental factors were assessed by decomposing the  $t_{FH}$  variance. This method permitted to map the effect of the temporal variability as well as the effect of environmental factors on  $t_{FH}$  using a block kriging method. Spatial structures highlighted by this decomposition either at the farm level or at the field level, suggest that the map of the effect of environmental factors on  $t_{FH}$  could be used to support agronomic decisions. This idea is reinforced by the identification of factors explaining the environmental variability of  $t_{FH}$  and by the temporal stability of the spatial structures.

### 4.2 Introduction

Many authors have shown the interest of characterizing crop variability as a decision support to optimize either inputs management or important agronomic decisions on annual crops like corn (Zanella et al. 2015), wheat (Diacono et al. 2013; Quebrajo et al. 2015), cotton (Thorp et al. 2015), etc. Such characterizations have also been shown to be of interest in perennial cultivation like vines (Santesteban et al. 2012; Urretavizcaya et al. 2013) or fruit trees (Colaço et al. 2017; Torres-Sánchez et al. 2015). In all these works, the visualisation of plant response as maps at a given time is relevant because all plants are present more or less at the same phenological stage within the field. Therefore, observed differences highlight the incidence of environmental or management factors on the plant response. Some crops, mainly tropical species, however have continuous production cycles that lead to asynchronicity in plant phenology between plants within the same field. Banana production is an example of such a crop. As a result, understanding the differences in the crop response to environmental or management factors is more difficult because of the effect of this asynchronicity. Asynchronicity originates from the banana plants biology as well as its crop system. Banana plants only give one bunch by differentiation of their unique terminal bud, and the following cycle of production is ensured by selecting one sucker

per plant (Tixier et al. 2004). This succession of generations originating from the same corm is called « semi-perennial » : semi because each stem produces only one bunch, and perennial because each corm produces a succession of stems. Floral initiation is day neutral (Turner et al. 2007). It is not triggered by a seasonal factor, as is typical in most plants. Moreover, the growth rate of each stem is variable. It depends on the environmental conditions, but also on human decisions at the time of the selection of the sucker. As a result, the plants in a field become asynchronous with respect to each another. After plantation, the first flowerings are relatively synchronous but as the flowering cycles progress, the flowering peaks in a field become more and more spread out over time and the production becomes continuous. Examples of harvest curves over time can be found in Tixier et al. (2004).

The direct consequence of this asynchronicity is a diversity of phenological stages within fields : some plants are in a vegetative development stage, some others have an inflorescence, and others already have a mature bunch. Moreover, the distribution of the phenological stages over space is not uniform. At the farm level, this distribution depends mainly on the age of the field since planting. At the within-field level, the distribution depends more specifically on management practices and environmental factors. These practices include gradual plantings and replacement of weak or dead plants and environmental factors, including light exposition and local interactions between plants when competing for resources that can explain differences in growth rate of each one.

It follows that the diversity of phenological stages makes mapping in banana crops more difficult than for crops that are season-synchronized. For example, when considering yield mapping in bananas, the weight and number of bunches are measured throughout the year and not at a single date as is usually the case in seasonal crops. As a result, bunch weights at different positions in space are no longer directly comparable because of potential temporal effects. This asynchronicity in production causes problems with any crop observation, not just yield. Optical imagery use is also difficult because of the diversity of phenological stages (stem elongation, inflorescence emergence, flowering, ripening, senescence, etc.) that induce different spectral responses and that can co-exist at any given time.

Despite these methodological difficulties, developing precision agriculture for banana intensive production systems would be interesting in order to improve crop management and provide methods for sustainable production. This is all the more important as banana is one of the most consumed fruits in the world with more than 100 million tons produced in 2013 (FAO 2013).

Few articles present examples of applications of precision agriculture in banana plantations and none of them highlight the effect of asynchronicity in the spatial study because the examples are conducted under particular conditions (observations made at the same time on similar phenological stages). In 2000, Storvoogel proposed a low tech approach to make yield maps in banana plantation by aggregating spatially and temporally the weight of bunches harvested in unit areas of 0.1 to 0.9 ha to make maps. Freitas et al. (2015) studied the intensity and spatial distribution of a disease in a field and avoided the issue of asynchronicity by only sampling banana trees at the same phenological stage. If site-specific banana management is to be achieved, the industry will need robust but relatively user-friendly methods for mapping asynchronous crop data.

To the knowledge of the authors, the issue of mapping asynchronous crop data has not been directly addressed in the literature before. There are some similarities to issues of data harmonisation in soil and crop sensor data that have begun to be addressed over the past decade (Brenning et al. 2008 ; Skøien et al. 2009 ; Baume et al. 2011). These sensors can generate data that contain a bias when surveys are disjointed or performed by multiple sensors. However, these harmonisation methods are focused on situations with typically only one or a limited number of temporal gaps and for situations where neighbouring data are collected for the most part under the same conditions. In the situations studied in these papers, there tends to be a clear divide in the data where the bias occurs. Banana crop data is temporally continuous and the response from each plant (data point) is potentially not spatially auto-correlated with its neighbours. The asynchronous data are completely spatially intermixed. There is no clear divide, which discounts the application of many of the proposed harmonisation methods (Leroux et al. 2018b).

There are also some similarities to issues of non-stationarity of a variable in the case of a spatial trend.

In this case, the variable is the sum of a gradual variation in space and a random auto correlated process. The trend surface can be modelled using a linear or non-linear combination of the spatial coordinates so the prediction of the variable is the sum of the krigged residuals and the trend (Plant 2012; Oliver and Webster 2014). In this paper, the deterministic effects of asynchronicity could be seen as a trend (temporal trend and phenological trend) and the residuals could be seen as a random process which depends on environmental factors. To address the need to map asynchronous crop data, the objectives of the work presented here are : (i) to propose a general model and a methodology to decompose the total variability into two components : the variability due to the asynchronicity and the variability due to the environment and management to improve the mapping process, (ii) to apply this methodology to production data related to a commonly used plant response variable, and (iii) to provide results illustrating its relevance on real commercial banana farm data.

## 4.3 Methodological aspects

### 4.3.1 Theory

The theory of geostatistics has been developed to map and model spatially continuous agri-environmental variables, particularly in the community of soil sciences (Krig 1951; McBratney and Webster 1986). Geostatistical methods have been widely applied in Precision Agriculture (PA) to study and map plant response (Verdugo-Vásquez et al. 2015; Bramley 2005; Acevedo-Opazo et al. 2009) although plant properties are not strictly speaking continuous variables. Indeed, the positions of the measurements are determined by the positions of the plants and therefore their density. The concepts nevertheless apply insofar as each plant can be considered as a biological sensor whose response (yield, leaf area, etc.) is driven by one or several soil properties seen as continuous (soil water availability, soil texture, etc.).

The model commonly used to express this is :

(4.1)

In equation 4.1,  $y_j$  is the  $j^{\text{th}}$  observation of the plant response located on  $\mathbf{s}_j$ ,  $\mu$  the expected plant response in the domain of study and  $\epsilon_j$  a residual, which is assumed to be spatially autocorrelated.

Variogram analysis can be used to quantify the spatially structured variance, called the partial-sill ( $\sigma^2_{\text{ps}}$ ), and the spatially unstructured variance, called the nugget ( $\sigma^2_{\text{n}}$ ) of  $\sigma^2$ . In the absence of management effects, the spatially structured variance is considered to be explained by environmental conditions and predominantly by the soil properties, whose chemical, physical and biological characteristics vary continuously in space (Oliver and Webster 2014). It is also potentially affected by other spatial environmental and crop conditions that affect the development of the crop, e.g. biotic stresses, plant density, etc. The spatially unstructured variance is considered to come from the natural variability of the plants as well as measurement errors. The mapping of  $\sigma^2_{\text{ps}}$  is useful to represent the spatial variability and is usually done by kriging methods. Kriging methods use the spatial structure modelled by the theoretical variogram in order to make unbiased estimations of  $\mu$  at an unobserved location, or the average expected value of  $\mu$  in an area, as well as an estimation of the error of the prediction (the kriging variance).

The model described by equation 4.1 does not take into account the effect of the phenological stage of the plant nor the effect of its temporal component, i.e. past meteorological conditions or applied management that may have modified its response. In crops commonly studied in precision agriculture, this is not a problem because all plants are assumed to have experienced the same temporal conditions and to be at a similar phenological stage at the time of observation. In the case of banana, if the observations are made at the same time on plants at the same phenological stage, then the methodologies developed on seasonal plants apply. This is the case of the two examples of precision agriculture studies mentioned in the introduction (Stoorvogel et al. 2000; Freitas et al. 2015). This is as well the case in a previous work on the pseudo-trunk circumference of a field (Leroux et al. 2018a). In other cases, the effects of the phenological stages and the effects of time difference between plants, which is called in this paper the

temporal component, cannot be ignored. Therefore, it is proposed here to introduce into the equation terms that take these effects into account :

(4.2)

In equation 4.2,  $y_{jt}$  is the  $j^{\text{th}}$  observation of the banana response located on  $t$ ,  $\mu$  the expected plant response in the domain of study,  $\alpha$  the effect of the phenological stage,  $\beta$  the effect of the temporal component. The temporal component ( $\beta$ ) may include effects of weather or farmer's practices, which are likely to be different from one set of plants to another if observations are done at different dates. The  $i$  index denotes a cohort, which is the set of observations made at the same moment within a specified time period. The residual term  $\epsilon_{jt}$  is assumed to be spatially autocorrelated.

This is still a simplistic model that could be enriched with a term related to interaction between  $\alpha$  and  $\beta$ , if relevant. For example, an interaction term would be required if the effect of the weather would depend on the phenological stage.

It is assumed here that in the case of banana the spatial structure identified by the model specified by equation 4.1 will be different from the spatial structure identified by the model specified by equation 4.2 and that only the latter actually represents the effect of environmental factors on the plant response. In other terms, considering equation 4.2, the spatial structure of  $y_{jt}$  is assumed to be different from the spatial structure of  $\mu$  because of  $\alpha$  and  $\beta$ .

In the rest of the article, this hypothesis will be tested by decomposing the variability of the banana plant response.

### 4.3.2 Description of the studied plant response : Time to reach Fruit Maturity (*t<sub>fm</sub>*)

In this study, the plant response of interest ( $t_{fm}$ ) is the Time between the date of Flowering and the date of commercial Maturity of banana bunches ( $t_{fm}$ ). It is an important variable for producers to forecast production and for harvest logistics (Cottin et al. 1987; Bugaud and Lassoudière 2005). It is carefully monitored with the aim to prevent early ripening during shipping (Jullien et al. 2008). While the approach is applied to  $t_{fm}$  here, it could be equally applied to other plant responses. The first hypothesis made in the study is that the spatial variability in  $t_{fm}$  is driven by environmental and management factors (i.e. nutrient availability, light interception, crop density, etc.). The  $t_{fm}$  is also expected to vary seasonally because of weather conditions that modify banana development (Ganry and Sioussaram 1978) and because of adaptation of farmer's practices over the time.

### 4.3.3 Modeling of *t<sub>fm</sub>* variability and discretization of the temporal component using cohorts

The  $t_{fm}$  is a measurement of a fixed time interval made between two fixed phenological stages (flowering and maturity) so  $t_{fm}$  is not relevant in this equation for this study. From equation 2, the  $t_{fm}$  can then be modelled as follows :

(4.3)

In equation 4.3,  $t_{fm}$  is the  $j^{\text{th}}$  observation of the time between flowering and maturity made on  $t$ ,  $\mu$  the expected value of  $t_{fm}$  on the temporal and spatial domain of study,  $\alpha$  the temporal component and  $\epsilon_{jt}$  the residue which is expected to be spatially auto-correlated and which represents the effect of environmental factors. The hypothesis made in the study is that the spatial structure of  $t_{fm}$  is affected by the temporal component. In other words, the spatial structure of  $t_{fm}$  is expected to differ from the spatial structure of  $\mu$ .

The way of handling the temporal component, that is how to evaluate  $t_{fm}$ , needs now to be described. Let  $\Delta t$  be a time interval such that the length  $\Delta t$  is small enough to ensure that the set of

plants flowering during this time interval experienced the same temporal conditions. As a consequence, the differences in  $y_{ij}$  between these plants can be assumed to result only from environmental and crop factors. The plants that flower during  $t_c$  are called a cohort (as in Tixier et al. 2004). The temporal component can therefore also be called the cohort effect.

#### 4.3.4 Decomposition of the tfm variability

Let  $t$  be a period composed of  $c$  distinct and consecutive cohorts, which each contain  $n_c$  observations making a total of  $n$  observations. Thus,  $\bar{y}_{ij}$  is the empirical mean of the observations in a cohort  $c$  and  $\bar{y}_{.j}$  is the empirical mean of all observations in period  $t$ .

The variance decomposition formula is :

(4.4)

Where :

$$c \frac{\sum_{c=1}^c \bar{y}_{ij}^2}{n_c} \quad (4.5)$$

$$\frac{\sum_{c=1}^c \bar{y}_{.j}^2}{n} \quad (4.6)$$

$$c \frac{\sum_{c=1}^c \bar{y}_{ij} \bar{y}_{.j}}{n_c} \quad (4.7)$$

Let  $\sigma_c$ ,  $\sigma_{.j}$  and  $\sigma_{c,j}$  be the following variables :

$$\sigma_c^2 = \frac{\sum_{c=1}^c \bar{y}_{ij}^2}{n_c} - \frac{\sum_{c=1}^c \bar{y}_{.j}^2}{n} \quad (4.8)$$

$$\sigma_{.j}^2 = \frac{\sum_{c=1}^c \bar{y}_{.j}^2}{n} \quad (4.9)$$

$$\sigma_{c,j}^2 = \frac{\sum_{c=1}^c \bar{y}_{ij} \bar{y}_{.j}}{n_c} - \frac{\sum_{c=1}^c \bar{y}_{.j}^2}{n} \quad (4.10)$$

It can be noted that the variances of  $\sigma_c$ ,  $\sigma_{.j}$  and  $\sigma_{c,j}$  correspond to  $\sigma_c^2$ ,  $\sigma_{.j}^2$  and  $\sigma_{c,j}^2$ , respectively.

The empirical variograms of  $\sigma_c$ ,  $\sigma_{.j}$  and  $\sigma_{c,j}$  are denoted  $\gamma_c$ ,  $\gamma_{.j}$  and  $\gamma_{c,j}$  and the generic formula of the variogram is :

$$\gamma_{c,j} = \frac{\sum_{c=1}^c \bar{y}_{ij} \bar{y}_{ij'} - \frac{\sum_{c=1}^c \bar{y}_{ij} \sum_{c=1}^c \bar{y}_{ij'}}{n_c}}{n_c} \quad (4.11)$$

where  $n_c$  is the number of paired comparisons at lag  $h$ .

The cross variogram of  $\sigma_c$  and  $\sigma_{.j}$  is denoted  $\gamma_{c,j}$ . The variogram decomposition formula is then :

(4.12)

Since the variogram of  $\sigma_{c,j}$  corresponds to the variogram of  $\sigma_{c,j}$  this decomposition permits a decomposition of the spatial structure of  $\sigma_{c,j}$  into three components : the spatial structure related to environmental factors ( $\sigma_c$ ), the spatial structure related to the effect of cohorts ( $\sigma_{.j}$ ) and the spatial structure which is due to the correlation between the two precedent structures ( $\sigma_{c,j}$ ). It can be noted that if the cross variogram is not null then  $\sigma_{c,j}^2 > 0$ .

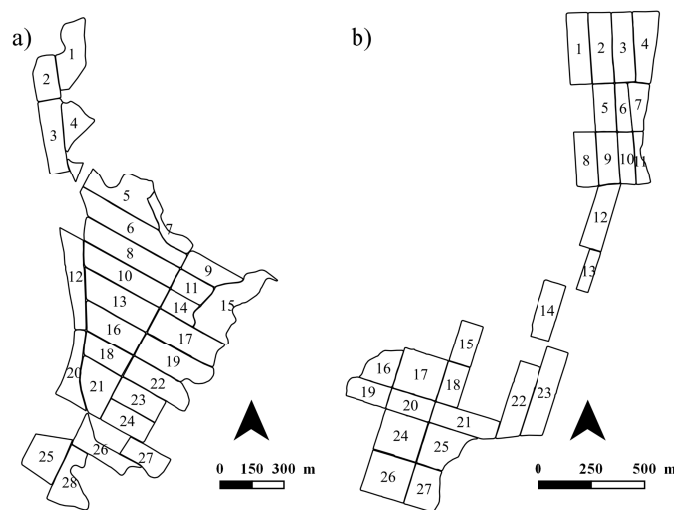


FIGURE 4.1 – Organisation of the fields of Farm A (a-subfield) and Farm B (b-subfield)

## 4.4 Material and Methods

### 4.4.1 Site study and data collection process

This study was conducted in two commercial banana farms, Farm A and Farm B, located in Njombe, Cameroon (WGS84 : E : 4.612, N : 9.639) managed by « Plantation du Haut Penja ».

The two farms are separated by 5.5 km. Farm A and Farm B have respectively 28 and 27 contiguous fields, for a total surface area of 92 and 160 ha respectively (Figure 4.1). Fields have been planted in double rows between 2008 and 2015 in Farm A and between 2007 and 2013 in Farm B. Density of plantation is approximately 1650 plants/ha. Banana plants have been planted from in-vitro multiplication of Cavendish cultivars « Grande Naine » and « Williams ». The climate in this area is tropical and characterized by low variations of mean temperatures in the year and high variations of monthly precipitations. The rainy season is from April to October with a mean monthly precipitation of 330 mm and a mean temperature of 26.8 °C. The dry season is from November to March with a mean monthly precipitation of 72 mm and a mean temperature of 27.9 °C (Figure 4.2). Soils are Andisols (Tchuenteu and Marie 1998), the elevation of Farm A ranges from 115 m to 139 m and the elevation of Farm B ranges from 165 m to 242 m (Figure 4.3). Altitude and slope data were obtained from ASTER Global Digital Elevation Model, a product of NASA and METI with a pixel resolution of 30 30 m. The two farms A and B were managed according to local conventional practices in banana farming, with regards to irrigation, fertilization, pest control, and fruit care. More information on the management of banana plantations can be found in Lassoudière 2007.

In the two farms, each inflorescence (which becomes a bunch) was manually marked with a tag at the flowering stage called « tête de cheval » (Lassoudière 2007). The marking of the inflorescences and maturity assessment was done twice a week. A bunch was mature when a reference fruit in the bunch had reached a specified diameter (determined by the grower). The reference fruit was the central finger of the second hand of each bunch. The diameter determined by the grower could change from a week to another but was spatially the same. An age limit, specified as 1050 degree days, was also used to determine maturity in bunches where the reference fruit did not reach the diameter limit. Overall, the marking of bunches and the method of assessment of bunch maturity were very similar to that commonly used in banana farming (Bugaud and Lassoudière 2005).

For this study, operators were equipped with a smartphone that they used to localize bunches when

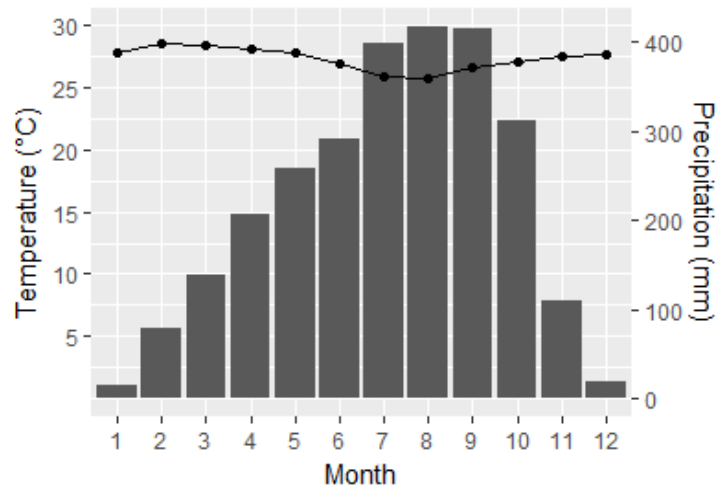


FIGURE 4.2 – Average total precipitation per month and average temperature per month. These averages were calculated from a weather station located on Farm A that has been operating since 2000

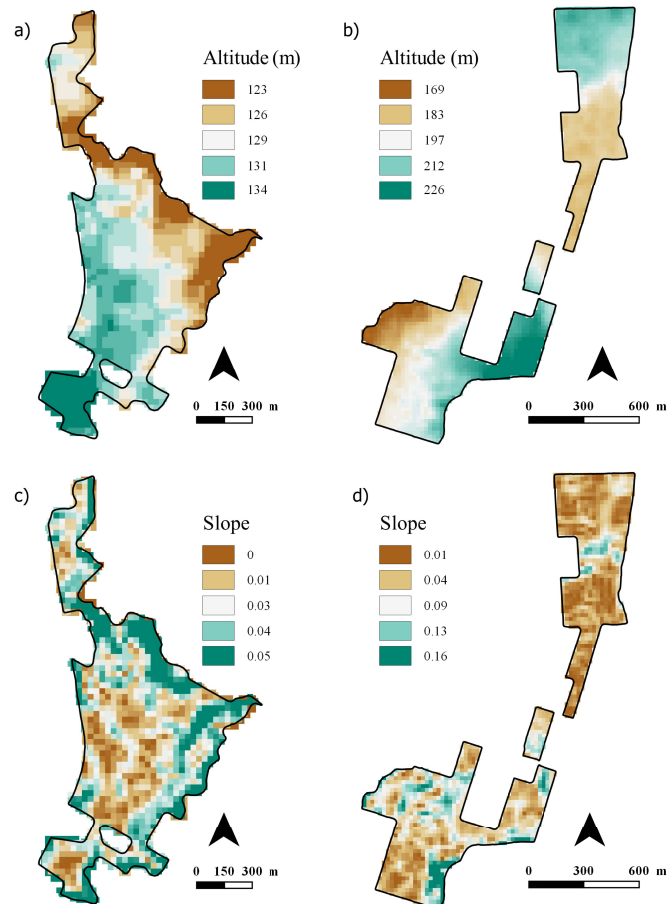


FIGURE 4.3 – Maps of the altitude and slope in Farm A (a and c) and Farm B (b and d)



TABLE 4.1 – Number of observations used for the variance decomposition by period and farm

	Farm A P1	Farm A P2	Farm A P3	Farm A P4	Farm B P3	Farm B P4
Number of observations	46999	54528	41744	38482	98967	114999

marking inflorescences and assessing maturity. The spatial accuracy of GNSS system of smartphones was assessed to be 5 m. The accuracy of the evaluation of the diameter limit was assessed to be around 2 mm. Each tag was identified by a unique bar code and a software (Proprietary software of Plantation du Haut Penja) was used to record the date and position during scanning of tags. Finally, each bunch marked with a tag and a bar code was recorded in a database with the following attributes : date of flowering, date of commercial maturity, spatial coordinates at flowering and spatial coordinates at commercial maturity. The  $\Delta t$  of each bunch was then calculated as the number of days between the date of flowering and the date of commercial maturity.

#### 4.4.2 Pre-processing of the data

Data were acquired by multiple employees for production purposes with time and productivity constraints. The data are massive (730 400 observations) but have characteristics of unreliability and inaccuracy inherent to this type of data. As a result, the first step was to check the quality of the data and delete erroneous data, such as data with unknown attribute inputs or wrong spatial coordinates. Data with wrong spatial coordinates included points that were more than 5 m outside the field borders, or points with a too different position at flowering and maturity. This situation came from a misuse of the smartphone by some operators who scanned the barcode before placing them on the inflorescence. Only 395 719 points were kept for the analysis. The number per period and per farm is given in Table 4.1. For this purpose, the duration  $\Delta t$  which defines the time resolution of a cohort and the duration P during which  $\Delta t$  are studied have been defined. A  $\Delta t$  duration of a week was considered. This was the shortest management period with no change in diameter and age limit criteria in the maturity assessment and where all fields were visited at least once. The P duration was chosen in order to gather around 500 observations per hectare and was set to 20 weeks. This density of points seemed reasonable to highlight the spatial structure of the  $\Delta t$ . This duration permitted the study of four distinct and consecutive periods P1 to P4. Period P1 started on 14 October 2015 and P4 finished on 26 April 2017. In farm B, the data were collected and studied only in periods P3 and P4.

#### 4.4.3 Decomposition of the variance

The second step was to decompose the  $\Delta t$  variance by calculating  $\sigma_{\Delta t}^2$ ,  $\sigma_{\text{env}}^2$  and  $\sigma_{\text{cohort}}^2$  and analysing their variance. The variance of  $\Delta t$  was decomposed for the 6 cases in Table 4.1, i.e. for the 4 periods P1 to P4 for Farm A and for the two periods P3 and P4 of Farm B and gave 6 different sets of  $\sigma_{\Delta t}^2$ ,  $\sigma_{\text{env}}^2$  and  $\sigma_{\text{cohort}}^2$ .

In each case, the effect of the cohort on the  $\Delta t$  was quantified using the ratio :

$$\frac{\sigma_{\text{cohort}}^2}{\sigma_{\Delta t}^2} \quad (4.13)$$

The effects of the cohorts and environment factors on the spatial structure of  $\Delta t$  were assessed using 4 indices : the total proportion of  $\Delta t$  spatially autocorrelated variance  $\rho_{\Delta t}^2$ , the proportion of the spatially autocorrelated variance which is due to environmental factors  $\rho_{\text{env}}^2$ , the proportion of the spatially autocorrelated variance which is due to the cohort denoted  $\rho_{\text{cohort}}^2$  and the proportion of the

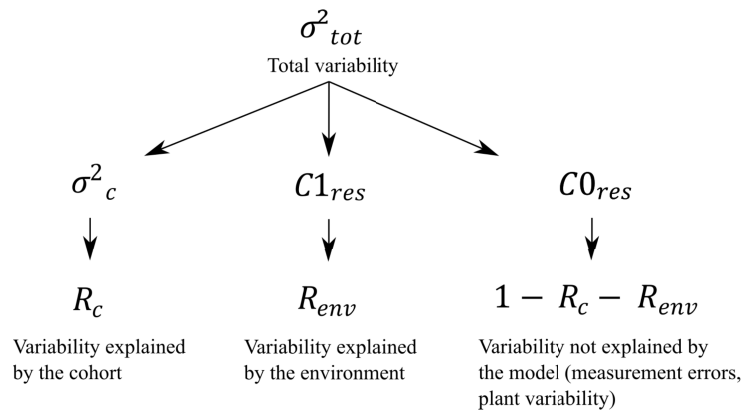


FIGURE 4.4 – Decomposition of the variability. Each arrow that connects a variable V1 to V2 can be read as V2 derives from V1

total variance which is due to environmental factors denoted :

$$\text{-----} \quad (4.14)$$

$$\text{-----} \quad (4.15)$$

$$\text{-----} \quad (4.16)$$

$$\text{-----} \quad (4.17)$$

Details on the signification of C0 and C1 of each variable , and are given in the part « spatial analysis » as they are derived from the variogram of each variable. The schema of the decomposition of the variance is given Figure 4.4.

#### 4.4.4 Spatial analysis

The spatial structure of was studied using an isotropic variogram (Bachmaier and Backes 2008). The components of the spatial structure of i.e., the spatial structure due to the cohort effect and the spatial structure due to environmental factors, were isolated using the variogram decomposition formula (Equation 4.4) by calculating the variogram of and and their cross-variogram . The variograms and cross-variogram were computed using 25 lags ranging from 1 to 400 m. The maximum lag was chosen in order to avoid an undesirable shape effect and directional preference in the variance analysis. The distance between each lag was shorter for small lags ( 50 m) than for larger lags in order to identify short scale variation with a good accuracy.

The nugget variance ( ) which corresponds to the spatially unstructured variability was estimated by the value of the variogram at the minimum lag. The sill variance is the sum of the nugget variance and the partial-sill variance ( ). was estimated by the value at the maximum lag distance (400 m). More details on the parameters can be found in Oliver and Webster (2014).

#### 4.4.5 Data mapping

The maps showing the total variability of , the variability due to cohorts and the variability due to environmental factors were calculated by kriging , and . The kriging was carried out on blocks

of 10 × 10 m centred on a grid of points regularly spaced 10 m apart. Block kriging was chosen rather than punctual kriging because the prediction of one observation is less relevant than the prediction of the mean on an area to represent environmental effects. Indeed, the variance of prediction in the punctual kriging includes the nugget variance whereas this is not the case with block kriging.

Theoretical variogram models (Gaussian, Spherical, Exponential), which are necessary to perform the block kriging, were fitted to the experimental variograms. The best model was chosen as the model that minimized the sum of squares errors. In some cases where the experimental variogram showed several spatial structures, nested models were used to fit the theoretical variogram. More details on the procedure can be found in Oliver and Webster (2014).

#### 4.4.6 Comparison of maps to assess the variance decomposition approach

The map of  $\hat{y}_i$  corresponds to the map without temporal correction, the map of  $\hat{y}_i^*$  corresponds to the unwanted temporal effect on the  $\hat{y}_i$  and the map of  $\hat{y}_i^e$  corresponds to the effect of the environment.

The relevance of the method was assessed using three criteria, (i) the temporal stability of the maps, (ii) their correlation with terrain attributes (slope, altitude) and (iii) their kriging variance. A better temporal stability would be expected after the correction of the temporal effects on the data. A better relation with terrain attributes would also be expected for  $\hat{y}_i^*$  maps than for  $\hat{y}_i$  maps as  $\hat{y}_i^*$  is supposed to better reflect environmental factors than  $\hat{y}_i$ . The slope and altitude are known to affect the response of crops because they govern the soils composition, structure and water retention capability as well as the climatic conditions. So the relation between  $\hat{y}_i^*$  maps and slope and altitude should be more important than with  $\hat{y}_i$ . Finally, the block kriging variance of  $\hat{y}_i^*$  maps should be lower than that of  $\hat{y}_i$  maps because the spatial effect of the temporal component is removed. Note that the nugget variance does not contribute to the block kriging variance (Oliver and Webster 2014). The coefficient used to compare the maps was the correlation coefficient  $r$  obtained by a Pearson test with a modified t-test described by Clifford et al. (1989). This test takes into account the spatial autocorrelation. The same test was used to determine the correlations between  $\hat{y}_i^*$ ,  $\hat{y}_i^e$  and slope and altitude. Altitude and slope maps have a pixel resolution of 30 × 30 m. In order to have the same resolution, the maps of  $\hat{y}_i^*$  and  $\hat{y}_i^e$  were recalculated as previously described with a block size of 30 × 30 m centered on the center of each pixel of the altitude and slope maps.

#### 4.4.7 Software used

All the statistical tests and mapping were implemented using R (R Development Core Team 2008). The package gstat (Pebesma 2004) was used to compute the variograms and cross variograms, to fit the variogram models and to make the block kriging. Package sp (Pebesma and Bivand 2005) was used to make the maps. The package SpatialPack (Vallejos et al. 2018) was used to compute the Pearson test with a modified t-test to compare the maps.

### 4.5 Results and Discussion

#### 4.5.1 General characteristics of the data

By definition, the number of  $\hat{y}_i$  observations in the farms depends on the number of flowering events, which is variable in time and space (Figure 4.5). In recently planted fields, flowering is still synchronous. In these fields, the number of observations for a given cohort can be very high (1000 observations per hectare) or very low, even null, depending on the timing of the cohort. In older fields, the number of observations is less variable from one cohort to another but is relatively low, around 20 observations per hectare. Those results concerning banana plant asynchronicity are extensively described by Tixier et al. (2004). The low density of observations reduces the amount of information available within the fields

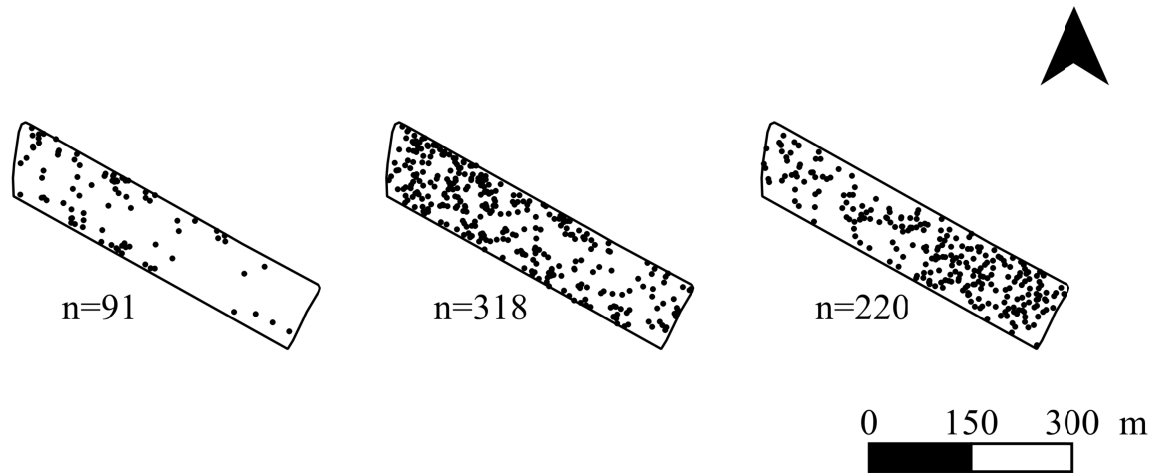


FIGURE 4.5 – Irregular number ( $n$ ) and location of observations for three different cohorts beginning on Mondays 04/18/2016, 05/16/2016 and 07/13/2016, respectively, in field 8 located in Farm A. Banana trees that have bloomed during these cohorts are represented by a dot

and the ability to make relevant maps (Webster and Oliver 1992). At the farm scale, the phenomenon of irregular spatial distributions of flowerings at a given time (Figure 4.5) was also observed because of the different ages of the fields.

If the number of observations and their spatial distribution in a cohort are adapted to characterize the spatial structure, the model generally used in PA and represented by Equation 4.1 may yield to correct results. It is likely to be the case with young plantations, which are still synchronous. However, in order to have a more generic method, whatever the age of the fields and the scale of the study, it is necessary to improve the density of observations and to handle the effect of heterogeneous spatial repartition of observations. Gathering observations of several cohorts over a period  $P$ , as proposed in this study, is a solution to these two issues.

For the period of study, mean values were 82.1 and 86.1 days for Farm A and Farm B, respectively and their standard deviation were 5.75 and 4.71 days, respectively. In the following, the decomposition of the variance aims at explaining the different sources of this variability.

#### 4.5.2 Evaluation of the cohort effect

Figure 4.6 shows that over a period  $P$ , the parameters (mean and percentiles) varied with the cohort and explained a part of the total variability. The importance of the cohort effect is presented in Table 4.2 through the ratio  $\rho$ . This coefficient ranges between 0.10 and 0.52 with the highest being obtained on P2 for Farm A. The cohort effect is high in Period 4 for Farm B ( $\rho = 0.43$ ) but is much lower over the same period for Farm A ( $\rho = 0.16$ ). Mean values were also slightly higher on Farm A than on Farm B (Figure 4.6, black curves). These results confirmed that  $\rho$  is dependent on a temporal component, which is modelled by  $\rho$  in Equation 4.3 and estimated by  $\hat{\rho}$ . In addition, the data presented in Figure 4.6 showed that the temporal component is dependent on the farm. This result confirmed the relevance of differentiating farms for the estimation of the cohort effect ( $\rho$ ). The two farms are not very far away, but they are located at different altitudes and that might explain these differences.

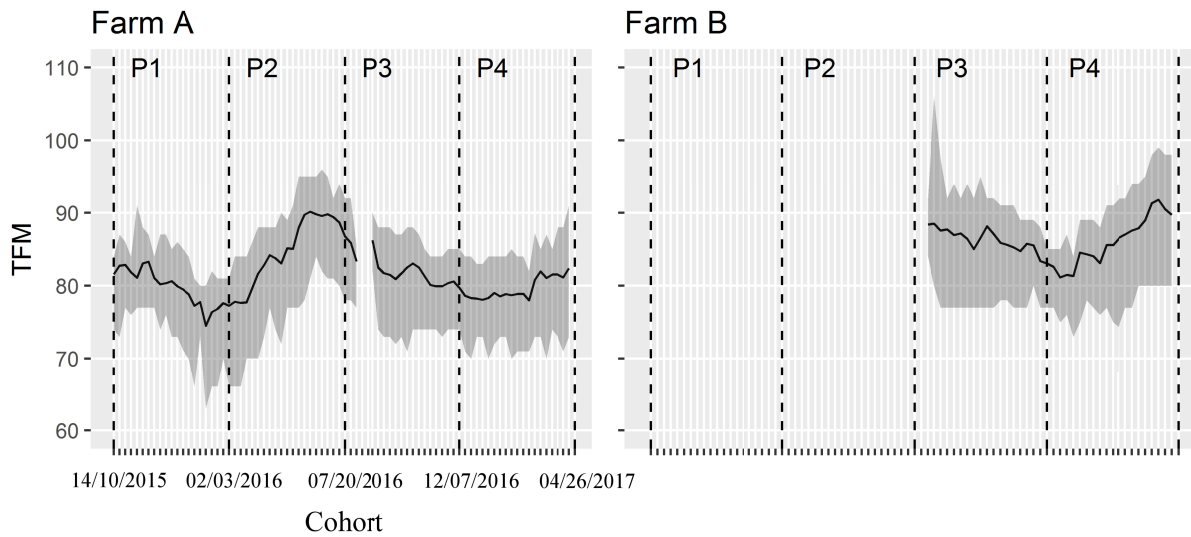


FIGURE 4.6 – Variation of TFM by cohort in Farm A and Farm B. The vertical dotted lines represent the limit between each period P1 to P4. The black curve represents  $\bar{y}$ . The grey area represents intervals encompassing 95 % of the  $y$  values for each cohort

TABLE 4.2 – Decomposition of the variance in two components, the cohort variance and the residual variance.  $\rho$  represents the ratio of cohort variance to the total variance

	Farm A P1	Farm A P2	Farm A P3	Farm A P4	Farm B P3	Farm B P4
(day <sup>2</sup> )	17.56	31.63	13.99	12.03	17.82	25.98
(day <sup>2</sup> )	5.92	16.34	3.48	1.9	1.75	11.13
(day <sup>2</sup> )	11.64	15.29	10.51	10.13	16.07	14.85
	0.34	0.52	0.25	0.16	0.10	0.43

TABLE 4.3 – Decomposition of the spatial structure of  $\sigma^2$ .  $\sigma^2$  and  $\sigma^2_{\text{partial-sill}}$  correspond respectively to the nugget variance and the partial-sill variance estimated from the variograms of each variable  $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$ .

	Farm A P1	Farm A P2	Farm A P3	Farm A P4	Farm B P3	Farm B P4
$\sigma^2_{\text{farm}}$	12,15	25,05	11,40	10,13	16,35	22,86
$\sigma^2_{\text{period}}$	4,79	13,22	3,24	1,81	1,81	10,84
$\sigma^2_{\text{cohort}}$	8,88	12,18	8,29	8,55	14,49	12,41
$\sigma^2_{\text{partial-sill}}$	5,71	6,79	2,18	2,00	1,40	2,36
$\sigma^2_{\text{total}}$	1,55	4,41	0,46	0,25	0,03	0,37
$\sigma^2_{\text{total}}$	2,68	2,85	1,55	1,50	1,51	1,88
$\sigma^2_{\text{total}}$	0,32	0,21	0,16	0,16	0,08	0,09
$\sigma^2_{\text{total}}$	0,37	0,61	0,23	0,14	0,02	0,17
$\sigma^2_{\text{total}}$	0,63	0,39	0,77	0,86	0,98	0,83
$\sigma^2_{\text{total}}$	0,15	0,09	0,11	0,13	0,08	0,07

### 4.5.3 Decomposition of the spatial structure

The decompositions of the spatial structure of  $\sigma^2$  for both farms at different periods are given in Figure 4.7 and the indices summarizing these decompositions are in Table 4.3. Each variogram of the decomposition ( $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$ ) is successively presented. A particular attention was given to the partial-sill ( $\sigma^2_{\text{partial-sill}}$ ) of  $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$  as it represents the spatial autocorrelation of each variable  $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$ .

The variogram  $\sigma^2_{\text{farm}}$  shows that  $\sigma^2_{\text{farm}}$  was spatially autocorrelated in all cases, i.e., for all the periods and the two farms. Indeed, the partial-sill variance of  $\sigma^2_{\text{farm}}$  ranged between 1.4 and 6.8 days (Table 4.3) and explained between 0.08 and 0.31 of  $\sigma^2_{\text{farm}}$  variance ( $\sigma^2_{\text{partial-sill}}$  in table 4.3). This result was expected, since  $\sigma^2_{\text{farm}}$  was supposed to depend on environmental and management factors that varied spatially within the farms.

The variogram  $\sigma^2_{\text{period}}$  shows that  $\sigma^2_{\text{period}}$  was spatially autocorrelated in all the cases except for P3 in Farm B. The autocorrelation validates the hypothesis made in this study that the spatially autocorrelated variance of  $\sigma^2_{\text{period}}$  is not only due to the effect of environmental factors but also to the temporal component. The proportion of the partial-sill of the  $\sigma^2_{\text{period}}$  that is due to the cohort effect ( $\sigma^2_{\text{cohort}}$  in Table 4.3) can even reach 0.61 for farm A at P2. This extreme case will be detailed later.

The variogram  $\sigma^2_{\text{cohort}}$  shows that  $\sigma^2_{\text{cohort}}$  is spatially autocorrelated in all the cases. It can be noted that  $\sigma^2_{\text{cohort}}$  was more stationary than  $\sigma^2_{\text{farm}}$  at P1 and P2 for Farm A. More details on the nugget variance of  $\sigma^2_{\text{cohort}}$  are given on the part « Interpretation of  $\sigma^2_{\text{cohort}}$  ».

The cross-variogram  $\sigma^2_{\text{farm,period}}$  was almost null for high distances in all cases. It was expected by construction according to Equation 4.5 that the covariance between  $\sigma^2_{\text{farm}}$  and  $\sigma^2_{\text{period}}$  was null. However, the value of  $\sigma^2_{\text{farm,period}}$  is not zero for small distances. This explained why  $\sigma^2_{\text{total}}$  was not exactly equal to  $\sigma^2_{\text{farm}} + \sigma^2_{\text{period}} + \sigma^2_{\text{cohort}}$  (Table 4.3). This result implied a spatial correlation between  $\sigma^2_{\text{farm}}$  and  $\sigma^2_{\text{period}}$  for small distances. One plausible reason for this is that the estimation of the cohort effect is not totally independent of the estimation of the environment effect. Indeed, the observations were not regularly distributed within the farm so the estimation of the cohort effect may still have contained some bias. However, the proposed cohort effect correction can still be considered an improved mapping approach.

As expected, the sill of  $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$  (calculated as  $\sigma^2_{\text{partial-sill}}$ ) in Table 4.3 corresponded respectively to the variance  $\sigma^2_{\text{farm}}$ ,  $\sigma^2_{\text{period}}$  and  $\sigma^2_{\text{cohort}}$  of Table 4.2. Therefore, all the results concerning the cohort effect presented in the last section could also be derived from the variograms presented in Figure 4.7.

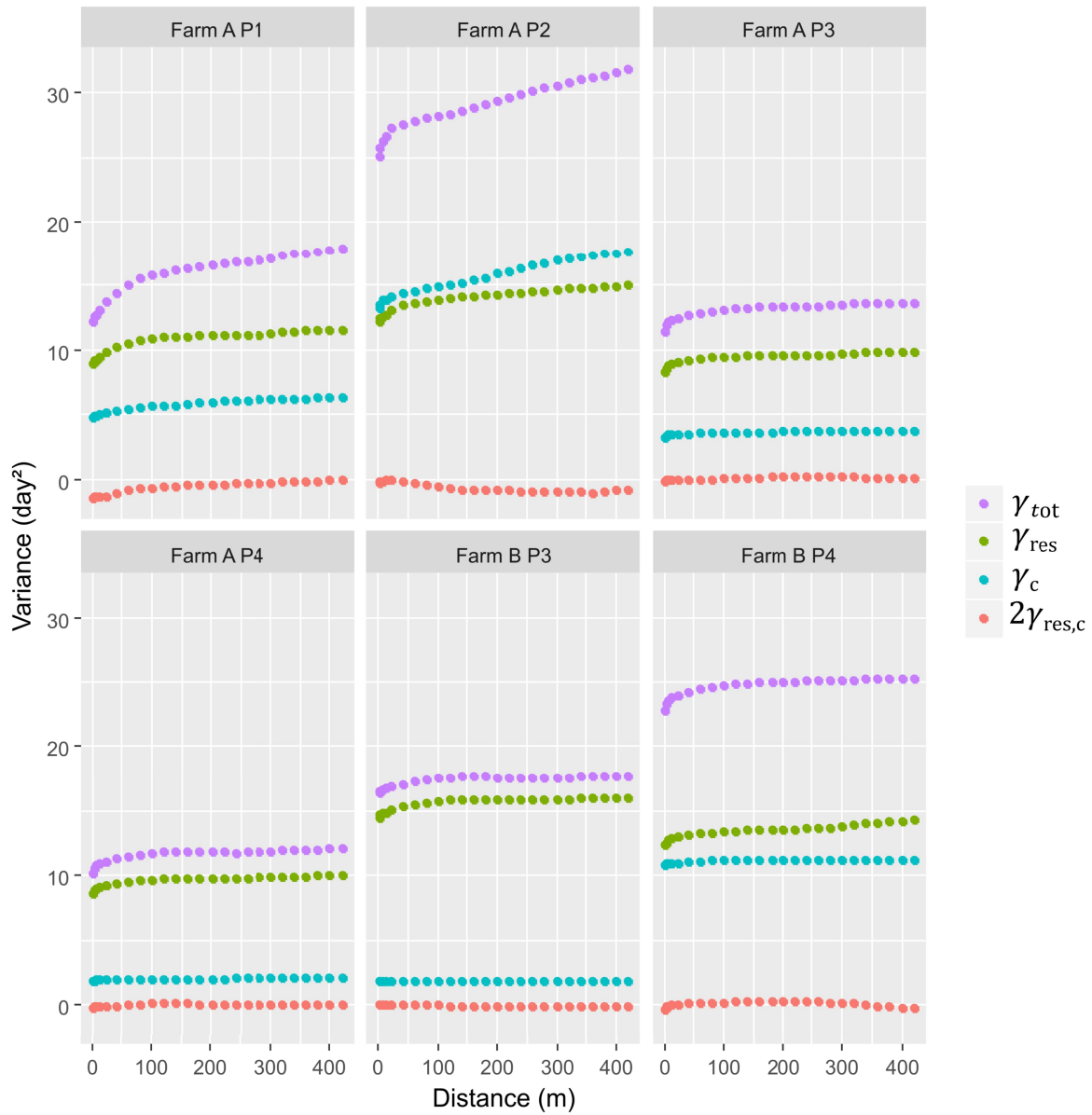


FIGURE 4.7 – Decomposition of the variogram of according to the formula

TABLE 4.4 – Correlation coefficient between maps and maps obtained at each period of each farm.

	Farm A P1	Farm A P2	Farm A P3	Farm A P4	Farm B P3	Farm B P4
	0.92**	0.69**	0.95**	0.93**	0.98**	0.90**

#### 4.5.4 Effect of the cohorts and of environmental factors on the maps of tfm

The map of the three variables resulting from the decomposition are given for Farm A at P2 in Figure 4.8 and for the field 8 of the same farm in Figure 4.9. This farm and period were chosen as they corresponded to the case with the maximum cohort effect ( in Table 4.2) and where the spatial influence of the cohort in was maximum ( in Table 4.3). Field 8 is shown as an example of the map at the field level. Negative and positive values corresponded to places where were respectively shorter and longer than the average. The causes explaining the differences in were different for the analysis of and maps. For the maps, negative values corresponded to areas where environmental factors lead to an acceleration of the growth in diameter and thus a decrease of the time to reach maturity. Conversely, positive values corresponded to places where environmental factors were less favourable to the diameter growth. For the maps, negative values were interpreted as places where there were more measurements made on cohorts that had an accelerated maturity due to temporal conditions (weather, production instructions, etc.). On the contrary, the positive values corresponded to areas where more measurements were made on cohorts where the was longer. The variability within the map was due to the two causes cited above.

For this farm and period, the strong variability in the map Figure 4.8 was due to the irregular recording of observations in the cohorts and to the strong effect of the cohort on ( 0.52, Table 4.2). The irregular repartition of the recordings for field 8 of Farm A was given in Figure 4.5 and may also explain the strong influence of in the spatial structure of (Figure 4.9). Without correction, the growth of the bunches of this field would appear slower than in other parts of the farm (positive values) whereas it is actually faster (negative values in map).

For the other periods and the other farm, the differences between and maps were less important (Table 4.4). This is explained by the combined effect of more uniformity of observations and/or less cohort effect on . For example, Farm B at P3 presented the lowest cohort effect ( 0.1, Table 4.4) and the observations were relatively regular because the plantation is old. This likely explained why in this case and maps were very similar (results not shown) with a correlation coefficient of 0.98 ( in Table 4.4).

#### 4.5.5 Assessment of the variance decomposition approach

The temporal stability of and maps is given in Table 4.5. The maps of were not stable between P1 and P2 and between P2 and P3. On the contrary, the maps of were more stable for these periods. Between the other periods the two farms presented a correlation coefficient which is of the same order of magnitude (Table 4.5). This result was expected because, as previously presented, the difference between the two variables was small at these two periods.

The correlation between the slope and the maps of and is given in Table 4.6. The correlation with the slope and was generally negative and was significant at P1 and P2. For P2, when the difference between and is maximum, the map was correlated to the slope whereas this was not the case for .

The result of the comparison of or with altitude are not shown as the p-value is always higher than 0.05.

The difference between and maps shown in the previous part, the increase of the temporal stability of maps and its better correlation with the slope justifies making a temporal correction to



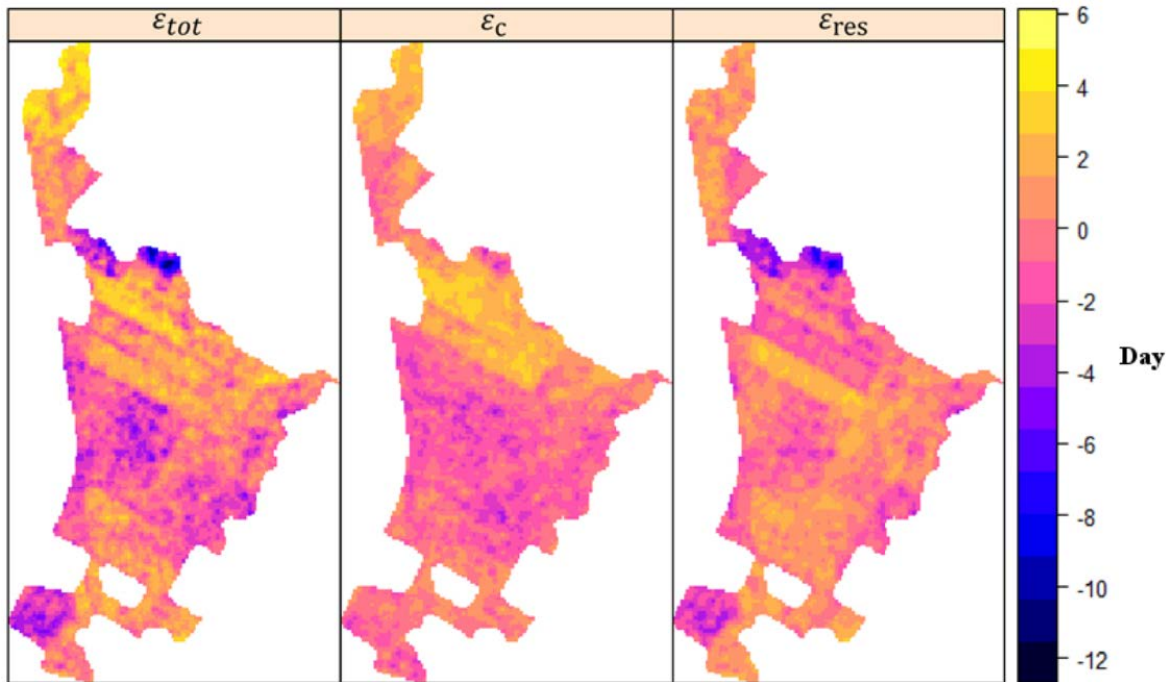


FIGURE 4.8 – Farm A Period 2. Block kriging maps of the  $\epsilon_{tot}$  variability ( ), of the  $\epsilon_c$  variability induced by the cohort ( ), and of the  $\epsilon_{res}$  variability induced by environmental factors ( )

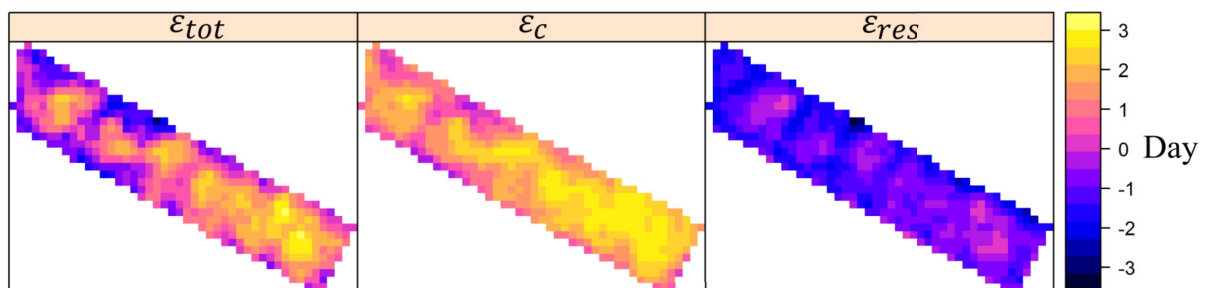


FIGURE 4.9 – Field 8, Farm A. Block kriging maps of the  $\epsilon_{tot}$  variability ( ), of the  $\epsilon_c$  variability induced by the cohort ( ), and of the  $\epsilon_{res}$  variability induced by environmental factors ( ). \*\* represents p-value <0.005

TABLE 4.5 – Correlation coefficient between maps at different periods. · p-value  $\geq 0.05$ , \* p-value  $\geq 0.01$ , \*\* p-value  $\geq 0.001$ 

Variable	Farm	Period 1 - Period 2	Period 2 - Period 3	Period 3 - Period 4
map	A	0.48**	0.55**	0.49*
map	B	-	-	0.48 ·
map	A	0.21 ·	0.04	0.52*
map	B	-	-	0.42 ·

TABLE 4.6 – Correlation coefficient between  $\epsilon_{res}$  or maps and the slope. · p-value  $\geq 0.05$ , \* p-value  $\geq 0.01$ , \*\* p-value  $\geq 0.001$ 

Variable	Farm	Period 1	Period 2	Period 3	Period 4
map	A	-0.29 ·	-0.24 ·	-0.21	-0.06
map	A	-	-	0.02	-0.28
map	A	-0.26 ·	0.03	-0.23	0.00
map	A	-	-	0.01	-0.27

study the effect of environmental factors. This correction decreased the risks of bias due to the cohort effects when analysing the maps of  $\epsilon_{res}$ . Moreover, the mean variance of estimation of  $\epsilon_{res}$  was always lower or equal to that of  $\epsilon_{map}$  which demonstrated a lower uncertainty of the maps of  $\epsilon_{res}$  than that of  $\epsilon_{map}$  (Table 4.7).

#### 4.5.6 Interpretation of $\epsilon_{res}$

More details are now given on the variable  $\epsilon_{res}$  and its properties considering that the effect of environmental factors can be studied from this variable. The partial-sill variance  $\sigma_{res}^2$ , highlighted for the two farms at different periods in Table 4.3, showed that the variable  $\epsilon_{res}$  is spatially autocorrelated. The ranges of the variograms (Figure 4.7) showed that the primary spatial structures had a range of less than 100 m.

Yet, between 76 % and 91 % of  $\sigma_{res}^2$  variance corresponded to a nugget variance (Figure 4.7). This nugget variance was expected given the commercial context of the data acquisition. Several aspects may explain this high variance such as : (i) the inaccuracy of the harvesting decision which is based on the size measurement of a reference banana finger in the bunch, ii) the GNSS inaccuracy, iii) the natural variability between neighbouring and competing plants (Taylor and Bates 2012). The dataset does not allow for differentiating these different sources of imprecision.

Despite this high nugget effect, there was always a spatially auto-correlated component to the variance. It was assumed that the method used for the mapping reduced the noise of the data to highlight the spatial

TABLE 4.7 – Mean Block kriging variance of the estimation of  $\epsilon_{res}$  and  $\epsilon_{map}$  (units : day<sup>2</sup>)

Variable	Farm A P1	Farm A P2	Farm A P3	Farm A P4	Farm B P3	Farm B P4
Mean block kriging variance of $\epsilon_{res}$	0.34	0.41	0.27	0.29	0.30	0.32
Mean block kriging variance of $\epsilon_{map}$	0.53	0.96	0.37	0.34	0.30	0.57

variability. Indeed, the block kriging method makes a surface estimate and reduces the variability related to the imprecision of the measurements. Another advantage of the high number of observations was the possibility to use a small block size for kriging with still a large number of observations within the blocks. This made it possible to make maps with a resolution fine enough to see spatial phenomena even if they have a short spatial range. Other durations and dimensions could have been chosen in order to optimize the information knowing that there is a trade-off between the spatial resolution, the temporal resolution and the ratio information/noise wanted (Tisseyre et al. 2018). It should be pointed out, however, that longer periods of integration might be not relevant as it could mix different spatial structures. For example, it is assumed that in the rainy season or the dry season, the maturity patterns will not be the same because of the interaction between weather and soil. Indeed, the growth limiting factors could be different depending on the season. It was shown that for small distances  $\rho$  was not null so that the estimation of  $\rho$  and  $\sigma$  were not totally independent. It would be possible to make a sampling of the measurements in order to have a more regular number of the observations in space to estimate the cohort effect. This implies that it may be preferable to use a subset of the data to ensure the same density of observations per field across the entire study area. This possibility wasn't chosen as some fields and cohorts have limited data and reducing the density of points to this number would imply removing a lot of measurements.

#### 4.5.7 Interpretation of $\epsilon_{res}$ maps

Figures 4.10 and 4.11 show the maps of  $\epsilon_{res}$  obtained with a block kriging method for Farm A and Farm B, respectively. They show differences in growing duration ranging from 4 to 12 days on each farm. The spatial pattern involved different causes for Farm A and Farm B, so each Farm is considered independently.

On Farm A (Figure 4.10), environmental factors on some fields were systematically unfavourable for the diameter growth. For example, Field 10, had positive values at all times. Other fields, such as Field 6, had systematically favourable conditions. These effects might explain why there was an overall stability of 0.5 between each period (Table 4.5). Conversely, some areas presented an instability between periods. For example, values in field 25 were positive at P1 and negative at P2. The edges of Fields 27 and 28 showed very low values at P2. These places corresponded to areas where banana trees have been replanted and where they were in their first flowering cycle. Their growth was faster regardless of local factors. Lassoudière (2007) explains that the first production cycle is a special case of banana. It is only from the second cycle that banana is physiologically stable. Flowers from the first cycle disturb the spatial analysis by increasing the growth rate and locally decreasing the temporal stability.

On Farm B (Figure 4.11), the effect of fields was less evident than on Farm A, but the effect of field edges was clear. The  $\epsilon_{res}$  values were lower on the edges, which gives the impression of a grid on the figure. The border of the roads effect was cited by Lassoudière (2007) to explain a part of the variability of physiological maturity in banana (Green-Life). Its effect on the response of other crops has also been highlighted (Gomez and Datta 1971; MacKenzie et al. 2003). The correlation coefficient between the two periods was of the same order of magnitude as for Farm A ( $\rho = 0.48$ , Table 4.5). The stable effect of the borders between periods should explain this overall stability. On both farms, the correlation coefficient between slope derived from the digital elevation model and  $\epsilon_{res}$  was generally negative (Table 4.6). The steeper the slope, the shorter the time it took for the bunches to be harvested. It was statistically significant at periods P1 and P2 of Farm B. The correlation with terrain attributes (slope, borders) validated the hypothesis that the spatial structure was, at least partially, explained by permanent environmental factors (Kaspar et al. 2003). The stability of those factors may explain the temporal stability of the maps. This stability is a guarantee of the quality and relevance of their use for site specific management (Whelan and McBratney 2000). Indeed, some authors discourage site specific management from yield maps when there is a lack of temporal stability (Bakhsh et al. 2000; Diacono et al. 2013).

Some phenomena reduced the overall stability and made the coefficient of stability lower than its maximum value of 1. The effect of the first flowering peak was identified. Other transient factors, such as pests that are spatially organized could be involved like, e.g. the black Sigatoka (Freitas et al. 2015)

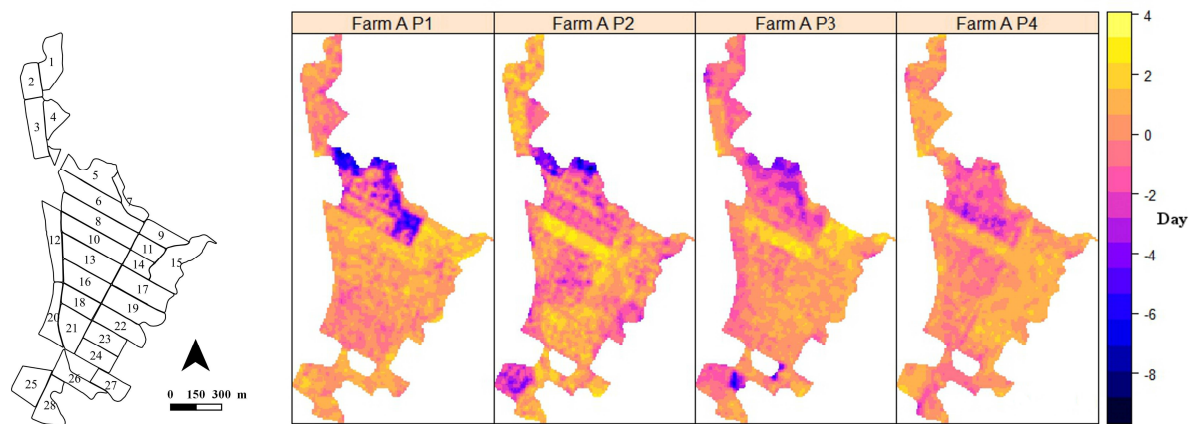


FIGURE 4.10 – Maps of farm A at four different periods P1 to P4. The colour scale unit is in days. Negative values correspond to locations where the maturity is reached faster than in other locations and, on the contrary, positive values correspond to locations where the maturity is reached slower

or weevils (Vinatier et al. 2009) which damage the banana crops. Operations made by the planters may also affect the stability (Kaspar et al. 2003). In the absence of additional data to identify precisely the factors causing the differences of maturity in the farms, this shows the importance of producers' expertise in interpreting these maps. It is to be noted that these factors may influence the development of the fruit directly or indirectly by affecting the vigour of the plant before bloom (Turner et al. 2007; Hotegni et al. 2014).

## 4.6 Conclusion

Banana production is asynchronous. As a consequence, the fields are made of plants at different phenological stages and the production is continuous throughout the year. In order to provide maps of the effect that environmental factors (soil, topography, etc.) have on plants, a model and a method were proposed. The method was applied on a maturity property obtained from commercial data and involved the processing of data resulting from observations done at a specific phenological stage (flowering) and at different dates of observation. The aim of the method was to account for the temporal variability induced by the gathering of measurements made at different dates and to create a new time-independent indicator.

The temporal correction using cohort permitted to estimate more precisely the effect of environmental factors with lower risk of bias in the analyses. The findings from the maps were consistent with terrain condition and local expertise. The cohort effect correction method would be relevant for other types of measurements made at a specific phenological stage. Such measurements could include the foliar nutrient diagnosis of banana plants at bloom, the assessment of Sigatoka on plants before bloom, the weight of the bunch at harvest. The approach showed that a diversity of cohorts leads to a higher uncertainty and bias in the maps. Some of the findings of this paper may also contribute to develop robust methods for remote sensing and other measures that need to be made at the same time on fields that mix a diversity of phenological stages.

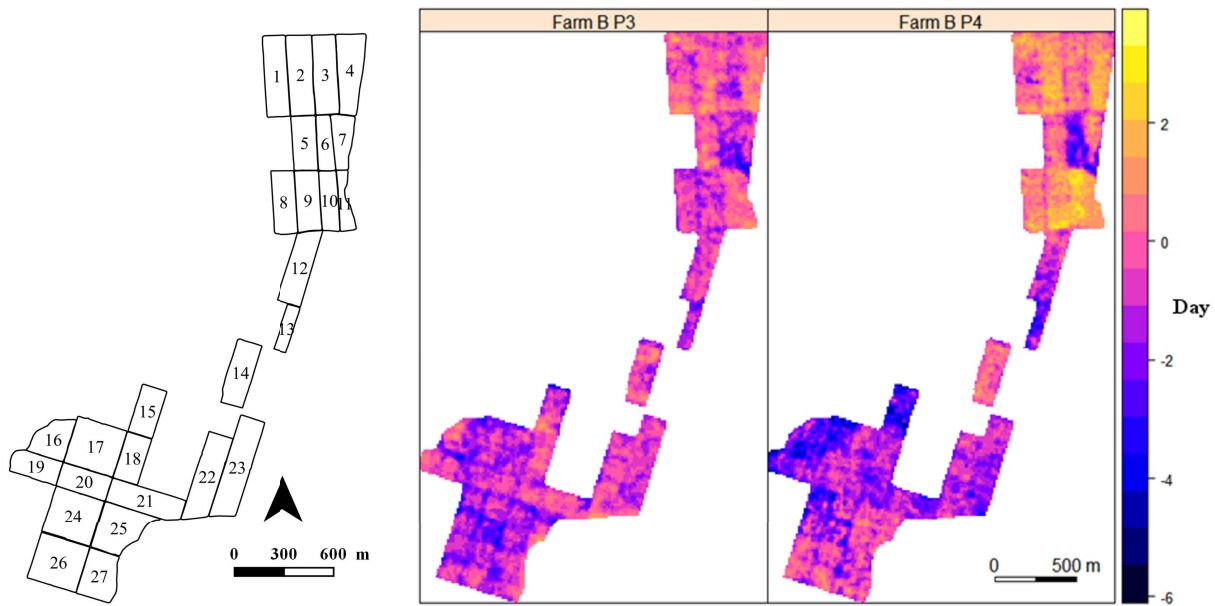


FIGURE 4.11 – Maps of farm B at two different periods P3 and P4. The colour scale unit is in days. Negative values correspond to locations where the maturity is reached faster than in other locations and, on the contrary, positive values correspond to locations where the maturity is reached slower

## 4.7 Acknowledgements

The authors would like to thank the Compagnie Fruitière and Plantation du Haut Penja, in particular Alain Normand, Gilbert Bonnefont, Aurélien Pugeaux and Patrick Vieil who provided the data and supported the study. They also would like to thank all the field workers who were involved in the data collection.

## Chapitre 5

# Étude de faisabilité du diagnostic agronomique des bananeraies par proxydétection et télédétection

Dans ce chapitre, nous évaluons l'intérêt de deux capteurs optiques en proxydétection et télédétection pour le diagnostic agronomique des bananeraies. Les producteurs de la Compagnie Fruitière évaluent l'état général des parcelles en intégrant de nombreuses sources d'information. Parmi elles, l'évaluation de la couleur des bananiers, de la dimension du pseudo-tronc, de la taille du rejet, du nombre de mains à la floraison, du type de sol ou encore du système d'irrigation et de drainage. Les producteurs peuvent compléter ces observations par des mesures chimiques de la composition minérale des feuilles ou du sol, par l'évaluation de l'infestation en maladies, etc. Le plus souvent, l'ensemble de ces observations associées à leurs connaissances permet de faire un diagnostic, c'est-à-dire d'identifier les causes des éventuels problèmes mis en évidence.

La proxydétection et la télédétection pourraient être des outils complémentaires permettant d'améliorer l'évaluation de l'état des parcelles et le diagnostic, notamment en diminuant les coûts d'analyses, en augmentant le nombre de mesures ou en permettant de mesurer de nouvelles caractéristiques. Cependant, le manque de recul sur leur intérêt dans les bananeraies reste un frein à leur adoption. Il est donc nécessaire d'établir des références sur ces nouvelles sources d'information pour juger de leur intérêt dans le processus de diagnostic. Il est aussi nécessaire d'évaluer l'effet de la mixité des stades phénologiques sur ces sources d'information, particulièrement en télédétection lorsque les stades phénologiques ne sont pas connus.

Au cours de la thèse, deux outils ont été évalués : un capteur piéton qui permet d'estimer le taux de chlorophylle des feuilles (Dualix) et un imageur multispectral (Sequoia) embarqué sur drone.

Je présente dans un premier temps le plan d'expérience qui a été mis en place pour recueillir les données permettant de caractériser ces nouvelles technologies par rapport à l'expertise traditionnelle. Cette partie est rédigée sous la forme d'un article de données Section 5.1 (non soumis).

Dans un second temps, nous présentons une étude qui a permis de vérifier si la teneur en chlorophylle des feuilles pouvait être un indicateur agronomique pertinent pour faire un diagnostic en bananeraie. Nous avons donc vérifié, en première approche, si la teneur en chlorophylle dépendait des conditions agronomiques des parcelles telles qu'elles étaient perçues par les producteurs. Nous avons aussi évalué les sources de variabilité de la mesure (stade phénologique, rang de la feuille, position sur la feuille). Cette étude a fait l'objet d'un article accepté à la conférence « European Conference on Precision Agriculture » ECPA dont la dernière version est présentée Section 5.2.

Nous présentons ensuite une seconde étude qui a été mise en place pour cartographier la teneur en chlorophylle des parcelles. La chaîne de traitement nécessaire pour prédire la teneur en chlorophylle a été réalisée par Gilles Rabatel. J'ai pu ensuite comparer les cartes produites aux cartes de circonférence des

pseudo-troncs qui sont classiquement utilisées par les producteurs pour analyser la vigueur des plantations. Cette étude a fait l'objet d'un second article accepté à la conférence ECPA et est présentée Section 5.3.

Enfin nous présentons les perspectives d'utilisation de ce travail dans les bananeraies Section 5.4.

## Sommaire

---

<b>5.1</b>	<b>Mise en place du plan d'expérience . . . . .</b>	<b>101</b>
5.1.1	Abstract . . . . .	101
5.1.2	Value of the data . . . . .	101
5.1.3	Data . . . . .	101
5.1.4	Acknowledgements . . . . .	108
<b>5.2</b>	<b>Analyse de plusieurs sources de variabilité du taux de chlorophylle à l'échelle du bananier . . . . .</b>	<b>108</b>
5.2.1	Abstract . . . . .	108
5.2.2	Introduction . . . . .	108
5.2.3	Material and methods . . . . .	109
5.2.4	Resultats and discussion . . . . .	112
5.2.5	Conclusion . . . . .	114
5.2.6	Acknowledgments . . . . .	115
<b>5.3</b>	<b>Conception d'une chaîne de traitement pour la prédiction par drone du taux de chlorophylle dans les bananeraies . . . . .</b>	<b>115</b>
5.3.1	Abstract . . . . .	115
5.3.2	Introduction . . . . .	115
5.3.3	Material and methods . . . . .	116
5.3.4	Results and discussion . . . . .	119
5.3.5	Conclusion . . . . .	120
<b>5.4</b>	<b>Perspectives d'utilisation de ces deux capteurs dans les bananeraies . . . . .</b>	<b>121</b>
5.4.1	Utilisation du Dualex . . . . .	121
5.4.2	Utilisation de la télédétection et du capteur Sequoia . . . . .	122

---

## 5.1 Mise en place du plan d'expérience

### Hyperspectral and multispectral datasets of banana leaves at different ranks, stages and agronomic conditions for diagnosis of the nutritional state of banana plantations by remote sensing

Julien Lamour<sup>1,2</sup>, Daniel Moura<sup>2</sup>, Olivier Naud<sup>2</sup>, Gilles Rabatel<sup>2</sup>

<sup>1</sup> Compagnie Fruitière, 13000 Marseille, France

<sup>2</sup> ITAP (Information–Technologies–Environmental Analysis–Agricultural Processes, French Denomination), Montpellier SupAgro, Irstea, University of Montpellier, 34000 Montpellier, France

#### 5.1.1 Abstract

Nowadays, more and more UAV companies are offering multi-spectral acquisitions but the link between the spectral signal of the canopy and its agronomic properties is sometimes difficult to interpret. This is no exception for banana production, especially since this crop has one particularity : asynchronism. Each plant develops at its own rhythm, which means that a field is composed of banana plants at different phenological stages. It is assumed that the spectral response of the leaves will vary according to phenological stages, which will complicate the interpretation of multispectral images.

This data set consists of information and observations made on 280 leaves. They were sampled from 100 banana plants spread over 3000 ha of plantations. Several types of data are presented : spectra measured with an integrating sphere, multispectral images made on flat leaves with Sequoia sensor, the composition of the leaves, and finally the characteristics of the leaves : leaf rank, phenological stage, field, etc. For 35 leaves from the same field, additional measurements were made to characterise the sampling field. These are the multispectral images, vigour mapping and leaf identification in these layers.

#### 5.1.2 Value of the data

The dataset could be used by plant physiologists to assess the effect of leaf rank, phenological stage, and field on the banana leaves compounds linked to the nutritional status ;

The data provide the composition of the macro-nutriments of the leaves and a spectral characterisation (from visible to near infra-red wavelengths) of sample points taken at different positions on the leaves. Chemometric models of leaf composition may be derived ;

Multispectral images of leaves for which point spectra, Dualex data and macro-nutriments data are also available make it possible to investigate potential of high resolution multispectral imagery for assessing leaf nutritional status ;

The geospatialized observations obtained at the within field level can be of interest for the precision agriculture community ;

Besides research purposes, the data may support development of numerical services for banana producers by companies or institutions.

#### 5.1.3 Data

##### Background

The banana plant is a giant monocotyledon with an underground stem (corm) that produces several pseudo-trunks (Turner et al. 2007). Each pseudo-trunk produces a succession of leaves at its top. The rank of leaves along the banana pseudo-trunk's axis indicates the order in which they appeared. The first leaf in apical position of the pseudo-trunk is the last leaf formed and the first basal leaf is the oldest



leaf of the banana plant. Flowering stops the appearance of new leaves and produces an inflorescence that will produce a bunch of bananas. After the harvest, the pseudo-trunk wilts. In many commercial plantations one sucker, called the « child » is selected before flowering of the « mother » and gives the following bunch which makes the production continue after each harvest. The flowering is aperiodic and can occur all year long, it only depends on the state of development of the pseudo-trunk. This characteristic leads to asynchronism within fields (Tixier et al. 2004). Indeed, each sucker develops at its own rhythm, which means that after a few flowering cycles the plants in a field can be at any stage of development. Consequently, the cover of a banana field can be characterised as mixed as opposed to uniform cover as is the case for seasonal crops such as wheat.

The composition of the banana leaves depends on their age, but also on the phenological stage of the banana plant. The nutritional state of the banana plant also affects the composition of the leaves (Martin-Prével 1977). Therefore, it is assumed that the phenological variability induced by asynchronism will modify the spectral response of the cover which will disturb the detection of the nutritional state by remote sensing.

The data set presented in this article was acquired to evaluate the effect of asynchronism and nutritional status on leaf spectral response. The composition of leaves was measured using a portable optical sensor (punctual data acquisition) and also by dosage of macro-nutrients. Spectral data were acquired as follows : (i) under optimal conditions, using an integrative sphere with controlled light emission and Zeiss MMS1 spectrometer, (ii) under controlled 3D positions, by positioning the leaves flat on the ground and acquiring the spectral data with a multi-spectral drone sensor that requires calibration according to current natural light, and finally (iii) in field conditions using the same multi-spectral sensor.

The data come from a measurement campaign carried out in a banana plantation located near Njombe (WGS84, E : 4.612, N : 9.639) in Cameroon. The fields where the measurements were made belong to the Plantation du Haut Penja (PHP). These are commercial fields and not experimental fields. The data were collected during a period of several weeks. The measurement campaign was carried out in two phases. The first phase consisted in creating a learning data set and the second resulted in a validation data set, although the data may be used differently.

### Dataset description

Different types of data are given. The « learning data set » is composed of :

- reflectance spectra of the leaves (Figure 5.1) ;

- tables of measurements made with Dualex Scientific+ ;

- Sequoia images acquired on leaves laid flat on the ground (Figure 5.2) with their corresponding annotated images giving their position and identifier (Figure 5.3) ;

- tables with the mineral content of the leaves.

Note that each leaf has an identifier that is used for each type of measurement to link the different data and metadata.

The « validation data set » was acquired on one field of the plantation and is composed of :

- a map of circumference of the pseudo-trunk (Figure 5.4), raw data not provided) ;

- RGB images, 3D photogrammetric model and multispectral Sequoia images of this field ;

- the position (polygon shapefile) of 35 leaves in this field (Figure 5.5) ;

- the same data as in the « learning data set » for those 35 leaves.

The diagram of the data structure is described in Figure 5.6.

### Experimental design

The experiments were carried out in the PHP plantations located near Njombe in Cameroon (WGS84, E : 4.612, N : 9.639). Fifteen fields were selected by the producers to represent the existing vigour range

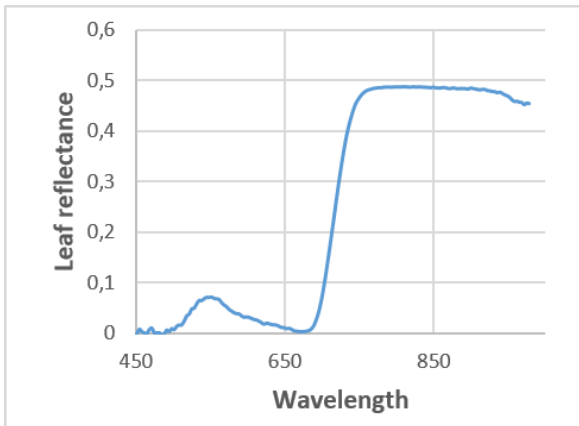


FIGURE 5.1 – Example of banana leaf reflectance spectrum

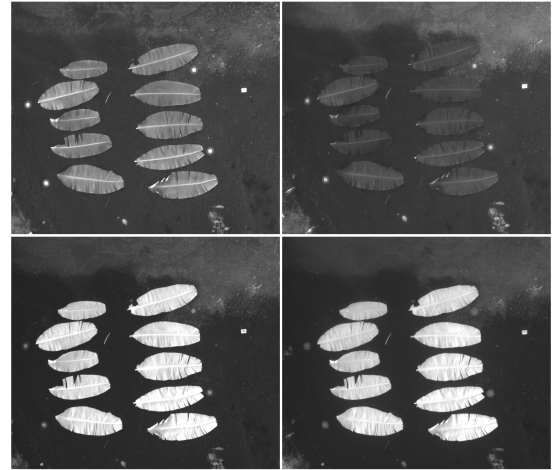


FIGURE 5.2 – Example of band images of banana leaves obtained from Sequoia acquisition at 10 m above the ground

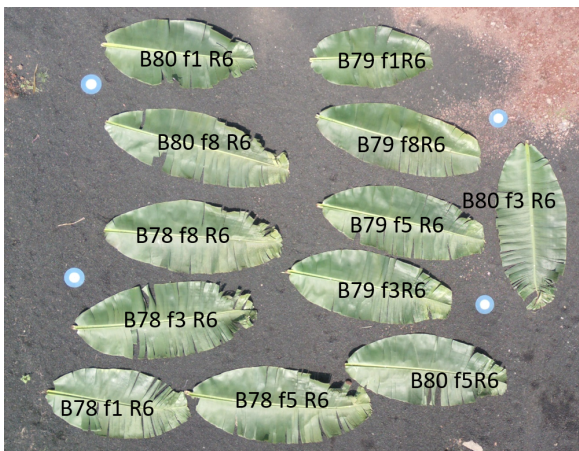


FIGURE 5.3 – Example of annotated RGB image giving the position and identification of the leaves

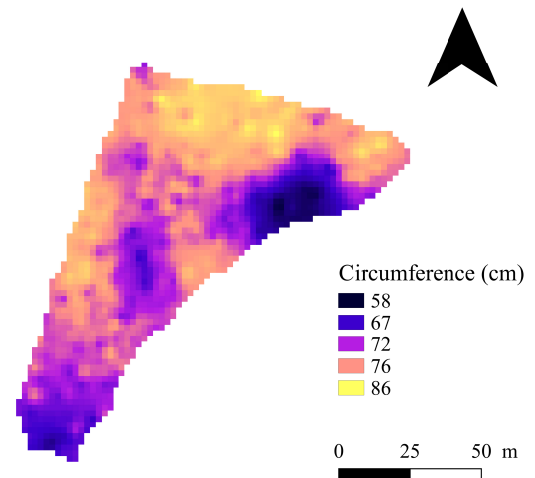


FIGURE 5.4 – Vigour map of the field where 35 leaves were sampled for the validation dataset

among the 3000 ha of banana plantation according to their expertise. From these 15 fields, 8 were selected for further experiments based on the average chlorophyll composition measured with Dualex (Figure 5.7) on leaf 3 of approximately 10 flowering banana plants.

In these eight fields, banana plants were selected at several phenological stages : before the apparent flower (BF), at flowering stage (F) or after flowering (AF2, AF6 and AF10 ; 2, 6 and 10 are the age of the bunches in weeks on the banana plant). The flowering date of each bunch was identified by producers at the time of flowering as it is usually done in this plantation and generally in banana plantations.

Leaf ranks were calculated as described in Lassoudière (2007). The last leaf to appear, i.e. the youngest



FIGURE 5.5 – Example of the RGB layer and of a polygon used to identify the 35 leaves of sampled in the field for the validation dataset

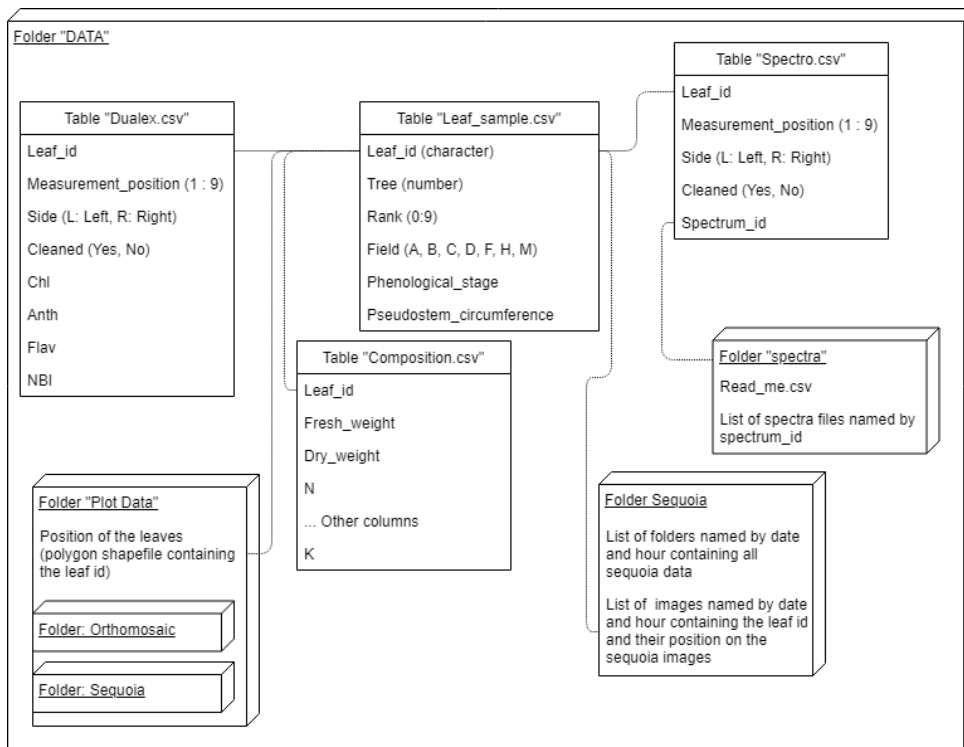


FIGURE 5.6 – Database structure

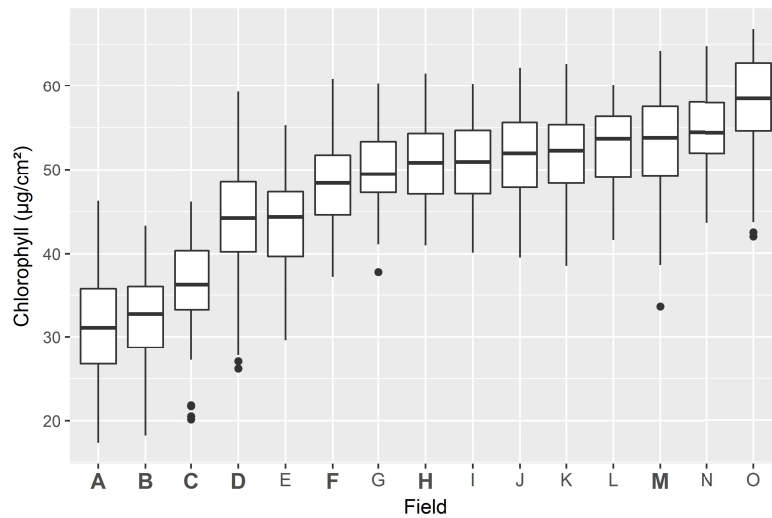


FIGURE 5.7 – Boxplot of the chlorophyll content of leaf 3 of flowering banana tree located in different fields. The fields used for the experiments are highlighted in bold

TABLE 5.1 – Number of samples by leaf rank and phenological stage. BF : Before Flowering, F : At flowering, AFn : n number of weeks After Flowering

	<b>BF stage</b>	<b>F stage</b>	<b>AF2 stage</b>	<b>AF6 stage</b>	<b>AF10 stage</b>
Rank 0	0	2	0	0	0
Rank 1	9	23	9	9	9
Rank 3	23	24	14	14	14
Rank 5	9	23	9	10	4
Rank 8	9	9	8	6	2

leaf completely unwound, located at the top of the banana plant is the leaf number 1. The leaf that appeared just before the leaf 1 and located at his armpit is called the leaf 2 and so on. In rare cases, the false leaf that appears at the same time as the inflorescence and which is smaller than the vegetative leaves has been sampled and numbered 0.

The sampled leaves were sent to the measurement laboratory in the plantation. Precautions have been taken to protect the leaves from the sun and prevent them from drying out. The number of samples per phenological stage and per leaf is described in Table 5.1.

## Leaf data acquisition

### Measuring chain

The measuring chain described in this section consists of 6 steps : (1) spectra acquisition with the integrative sphere, (2) Dualex measurement, (3) Sequoia flat leaf multispectral imagery, (4) rib separation and blade weighing, (5) drying, (6) mineral measurements. The sampling steps up to point 4 took about 2 hours, which limited the evolution of the chemical and spectral composition of the leaf once separated from the plant. Steps 4, 5 and 6 were only performed on the 150 leaves whose mineral composition was measured.

The measurements (1) and (2) were made on 18 positions in the leaves (Figure 5.8) : 9 on the left side and 9 on the right side. One side of each leaf (left or right depending on the leaf) was cleaned beforehand

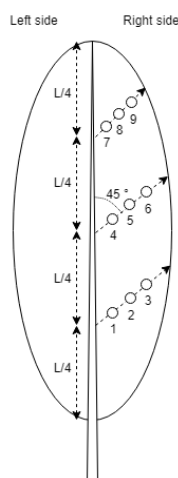


FIGURE 5.8 – Locations of the Dualex measurements and spectral measurement with the integrative sphere on the leaf. 9 locations are represented on the right part of the leaf. 9 other measurements were made at symmetrical positions on the left part of the leaf

to remove deposits (dust, residues of phytosanitary treatments) and assess the effect of the deposits on the measurements. Measurements (2) and (3) were made on the upper face of the leaves, i-e, on the face which is exposed to the sun when on the banana plant.

#### Spectral measurements with the integrative sphere

The measurements were conducted using a 50 mm diameter reflectance sphere (AvaSphere-50, Avantes) fibered with 0.91 mm core optical fibers (Ocean Optics). Illumination was provided by a tungsten halogen light source (HL – 2000, Ocean Optics) and the spectral acquisition was carried out using a spectrometer (MMS1, Zeiss). The temperature fluctuations of the light source were limited by turning on the lamp at least one hour before beginning the measurements : this lapse of time was observed to be sufficient for the light source to be as stable as possible. The spectral region of the spectrometer ranges between 380 and 1020 nm with 10 nm resolution intervals and 3 nm step intervals.

A white reference using a Spectralon was recorded between the set of 9 measurements made on the same side of each leaf.

An average dark current value for all wavelengths (mean : 141.5 standard deviation : 3.6), previously measured with the source light off, was applied on collected spectra, leading to the following correction equation :

$$\text{Corrected Reflectance} = \frac{\text{Raw Leaf Spectrum} - \text{Dark Current}}{\text{Raw White Spectrum} - \text{Dark Current}} \quad (5.1)$$

With  $R(\lambda)$  the reflectance spectrum ;  $I_L(\lambda)$  the raw leaf spectrum ;  $I_W(\lambda)$  the white raw spectrum and  $\lambda$  the wavelength. The data given are the raw intensity spectra  $I_L(\lambda)$  of each 18 points of each leaf, with their corresponding white reference  $I_W(\lambda)$  , as well as the corrected reflectance spectra  $R(\lambda)$  .

#### Dualex measurements

Dualex measurements were made with the emitter part of the clamp on the upper face of the leaves. A table of the 18 Dualex measurements made on each leaf is given in the dataset.

### Flat leaves Sequoia acquisitions

The multi-spectral sensor and the sunshine sensor of the Parrot Sequoia sensor were mounted on a quadrirotor drone (Quad Mikrokopter). The elevation of the drone was controlled by the pilot and stabilised at around 10 meters above the soil (Altimeter based on a pressure sensor). The Sequoia acquisition was then triggered manually.

All Sequoia measurements were made outside on a gravel floor composed of black pouzollane. Measurements were done at different hours of the day with different sky conditions (cloudy, clear). However, signals from the sunshine sensor that comes with the Sequoia sensor were recorded in the image Exif data (Exchangeable Image File Format), allowing reflectance correction. Four targets were positioned with a measured and fixed position during all Sequoia acquisitions. The leaves were placed flat in the space between the targets, just before the acquisition of the drone to prevent the leaves from drying out. Just after each acquisition, a schema of the position of the leaves on the scene was drawn and the identifier of each leaves was reported on it. The four narrow-band imagers of the Sequoia are separated by 15 mm. Normally, the acquisition has to be done at a minimum altitude of 30 m to ensure a correct coregistration of the 4 bands to make a multi-spectral image. At 10 meters, the offset between the imagers creates a significant spatial shift between bands. The coregistration method proposed by (Rabatel and Labbé 2016), combined with reflectance correction using the sunshine sensor data was used to build a multispectral image. The raw images for the four bands as well as multispectral reflectance images in ENVI format are given in this dataset. The positions of the target are also given as well as the schema of the leaf position with their identifier.

### Mineral analysis

The mineral composition (%N, %P, %K, %Ca, %Mg, %Na, %B, %Cu, %Fe, %Mn and %Zn) of 150 leaves were analysed by SADEF. These leaves were first cut to separate the blade from the central rib. The blade was then weighed, cleaned with water and cut in small parts. All the parts of the blade was then dried to limit their chemical evolution and then sent to the SADEF laboratory.

The composition of each leaf is given in a table with its identifier. The surface area, the fresh weight and dry weight of the leaves are given to express the mineral composition in weight by surface if necessary.

## **Supplementary data of the validation dataset**

### Leaves collection

A total of 35 leaves were sampled in a new field. This field of 1 ha was chosen for its heterogeneity in pseudo-stem circumference. Among them, 25 leaves were chosen on banana plants at various phenological stages. A paper (orange or blue without importance) was stapled on those leaves in order to be visible from the sky. Once drone measurements were made, those leaves were cut and brought to the laboratory where the same measurement as for the other leaves were made. The remaining 10 leaves were chosen because they had contrasting values of NDVI (data not given). The contour of the sampled leaves were manually drawn on the RGB and multispectral images (see the part « Field Sequoia Acquisition ») using QGIS and given in the dataset as polygon shapefiles (Figure 5.5).

### Vigour map of the field

The pseudo-stem circumference is a good indicator of the banana plant vigour (Fernandez Caldas and Garcia 1972). The pseudo-stem circumference was measured on plants for which the vegetative growth was over, i.e. plants that were either at the flowering stage or at a later phenological stage. A total of 551 measurements was taken using a tape measure at 1 m height and georeferenced with a trail type hand held GPS (raw data not given). These measures were made one month before the field Sequoia Acquisition (see next part). The circumference was interpolated as described in Leroux et al. (2018a) using variogram modelling and kriging. The interpolated map is given as a raster.

### Field Sequoia acquisition

The same drone and Sequoia sensor were used during the flat leaf acquisition and in the field. The altitude of the flight was approximately 30 m above the banana plants and the flight plan was designed to have an overlapping of the images of 80 % in the direction of travel of the UAV and 60 % between travels. Eight targets were positioned at the edge of the field. The distance between each target was measured with a tacheometer. Their absolute position (WGS 84) was georeferenced with a trail type handheld GPS (Garmin Etrex 30x).

The set of raw images obtained from the flight are given in the dataset as well as the RGB orthomosaic and the position of the targets.

#### **5.1.4 Acknowledgements**

The experiences were made in Plantation du Haut Penja with the support of its agronomy department. We are grateful for all the staff of the plantation. The mineral analysis were made by SADEF – pole Aspach.

## **5.2 Analyse de plusieurs sources de variabilité du taux de chlorophylle à l'échelle du bananier**

### **Disentangling the sources of chlorophyll-content variability in banana crops : a case-study in an African commercial plantation using an optical chlorophyll meter**

J. Lamour<sup>1,2</sup>, C.<sup>1,3</sup> Leroux, O. Naud<sup>1</sup>, M. Lechaudel<sup>4,5</sup>, G. Le Moguédec<sup>5</sup> and B. Tisseyre<sup>1</sup>

<sup>1</sup> ITAP, Montpellier SupAgro, Irstea, Univ Montpellier, Montpellier, France

<sup>2</sup> Compagnie Fruitière, Marseille, France

<sup>3</sup> SMAG, Montpellier 34960, France

<sup>4</sup> CIRAD, UMR QualiSud, F-97130 Capesterre-Belle-Eau, Guadeloupe, France

<sup>5</sup> QualiSud, Univ Montpellier, CIRAD, Montpellier SupAgro, Univ d'Avignon, Univ de La Réunion, Montpellier, France

<sup>5</sup> AMAP, INRA, Univ. Montpellier, IRD, CNRS, CIRAD, Montpellier, F-34000 France

#### **5.2.1 Abstract**

The objective of this study was to evaluate the effect of banana phenological stages on the chlorophyll content of its leaves in order to assess possible restrictions on the potential use of remote sensing for nutritional diagnoses when phenological stages cannot be discriminated on images. Chlorophyll measurements were made with a Dualix sensor at different positions on banana leaves with different ranks on the plant, on plants at different phenological stages and located in different fields of a commercial farm. The effect of these factors was studied using mixed models with nested factors and the results showed that they have a significant effect on the chlorophyll content. It is therefore suggested to take into account the diversity of phenological stages in banana fields before considering nutritional diagnoses services based on remote sensing.

#### **5.2.2 Introduction**

Chlorophyll is the pigment responsible for the leaf light interception during photosynthesis. As such, there is an interest in measuring the chlorophyll-content to evaluate plant health status and productivity.

Amongst all the leaf constituents, a major advantage of chlorophyll is that its concentration can be estimated through optical measurements. This property has led to the development of multiple proxy (SPAD<sup>®</sup>, Dualex<sup>®</sup>, N-sensor<sup>®</sup>) and remote (multi and hyper-spectral images) sensors. An empirical relation has also been shown between the chlorophyll-content and the leaf nitrogen content. Some studies have therefore tried to estimate nitrogen nutritional status from chlorophyll optical measurements to avoid time-consuming and costly nitrogen laboratory measurements. These studies were made by proxy detection (Debaeke et al. 2006; Prost and Jeuffroy 2007) or remote sensing studies (Haboudane et al. 2002; Gitelson et al. 2005). Operational services, like Farmstar<sup>®</sup> that rely on this principle, exist to manage crop fertilization in arable crops such as wheat or barley (Coquil and Bordes 2005).

Applying such approaches on banana crops is not trivial. Contrary to most of the crops for which vegetation cover can be considered uniform (same phenological stage), banana fields contain banana plants at different phenological stages, e.g., before flowering and after flowering. The biological and cropping causes that lead to this phenomenon are detailed in Tixier et al. (2004). In simple words, it must be understood that each banana plant blooms only once and that each sucker (the new stem) produced by the banana corm ensures the following blooming (Figure 5.9). Suckers have different growth rates and flowering dates, leading to a diversity of phenological stages in the population (field, field group). This is called asynchronism and it makes it more difficult to evaluate the banana nutritional status given the nitrogen evolution dynamics within the plants. Indeed, the leaf nitrogen content varies with the phenological stage, especially because of dilution phenomena (the nitrogen content diminishes when the biomass increases, Lemaire (1997)) but also because of nitrogen reallocation phenomena during fruit filling. Thus, in the case of a banana crop, it is assumed that the chlorophyll content in a population is not only due to the nutritional status of the banana plants but also to the different phenological stages caused by asynchronism. The main objective of this study was to quantify and compare the influence of these two factors on the banana leaf chlorophyll-content in a commercial plantation. The interest being, in the long-term scale, to evaluate the reliability of remote sensing to make spatialized nutritional diagnoses in banana plantations when information on phenological stages are missing.

### 5.2.3 Material and methods

#### Definitions and notations

Within a banana plantation, a field contains banana plants at different phenological stages (Figure 5.9). Each banana plant has different leaves that are characterized by their leaf rank on the pseudo-trunk. The youngest fully unfolded leaf is located at the top of the banana plant and is the 1-rank leaf. The leaf that appeared just before the 1-rank leaf and situated just below it is the 2-rank leaf, and so forth until the last leaf. The central rib splits the leaf's limb in two sides (Figure 5.9). The left side is the one on the left when one is looking at the upper face and when the peduncle is the nearest to the operator.

#### Study site and experiments

The two experiments were carried out on a 3000 ha commercial banana plantation located near Njombe, Cameroun (WGS84 : E :4.612, N : 9.639). Chlorophyll measurements were performed on the leaves of banana plants using the Dualex scientific sensor (FORCE-A, Orsay, France) that estimates it through light transmittance ( g.cm ).

**Experiment 1** : 15 out of the 1500 fields of the plantation were selected according to the farmers' expertise to obtain fields that they considered representative of the diversity in growing conditions in the plantation, e.g., soil fertility, climate, water status. These fields are represented by a letter from A to O. Approximately 10 banana plants at flowering were chosen in the diagonals of these rectangular fields. For each of these banana plants, approximately 10 chlorophyll measurements were performed on one of the two sides of the 3-rank leaf that was previously cleaned. Measurements per leaf were made following a X shape (Figure 5.9 b) on the leaf side of interest. The main objective of this preliminary experiment was



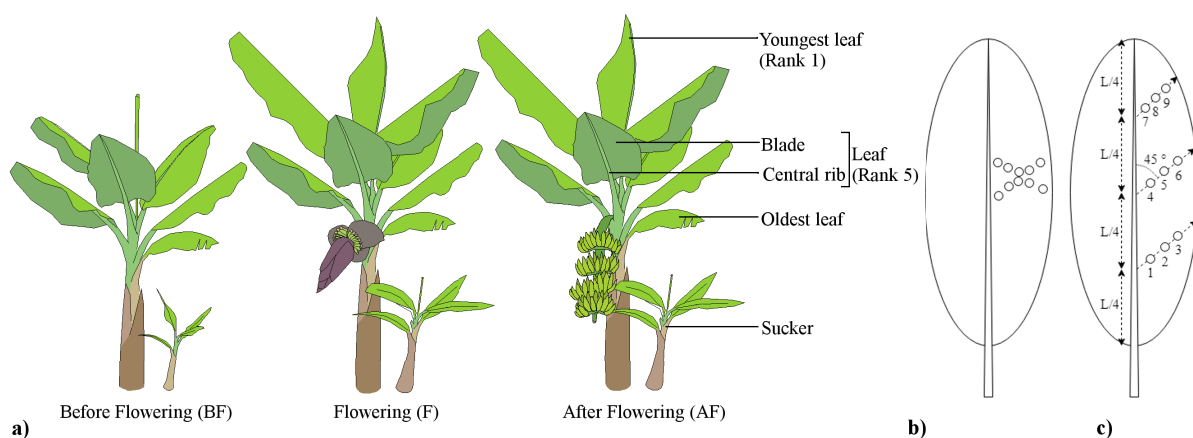


FIGURE 5.9 – Banana plants at different phenological stages (a) Location of chlorophyll measurements on a leaf for the first (b) and second (c) experiments

TABLE 5.2 – Number of samples by phenological stage and leaf rank in experiment 2

	BF stage	F stage	AF2 stage	AF6 stage	AF10 stage
Rank 1	9	23	9	9	9
Rank 3	23	24	14	14	14
Rank 5	9	23	9	10	4
Rank 8	9	9	8	6	2

two-fold : (i) make sure that the growing conditions had an influence on the field chlorophyll content, and (ii) compare the inter-field, inter-plant, and intra-leaf chlorophyll-content. This sample selection protocol is very similar to the one used for foliar analysis in banana crops (Lassoudière 2007).

**Experiment 2** : 8 out of these 15 fields (A, B, C, E, F, G, H, M) were used to evaluate the sources of chlorophyll-content variability due to banana plant asynchronism and architecture. Table 5.2 sums up the number of leaves that were studied for each leaf rank and phenological stage of interest. Five different phenological stages were considered : before flowering (BF), at flowering (F), and after flowering (AF2, AF6, AF10). For the last three stages (AF2, AF6, AF10), the number corresponds to the age of the bunches in weeks since flowering. More measurements were performed on the 3-rank leaf and for flowering banana plants because those are the rank and stage that are used by farmers to make leaf nutritional diagnoses (Table 5.2). In this experiment, uncommon banana plants or those for which some leaves were missing were avoided. For practical reasons and to reduce the within-plant variability, banana plants were selected within a restricted perimeter in the fields.

For all the leaves of this experiment 2, 9 measurements were made at different locations on each side of the leaves (18 measurements per leaf) following two axes x and y (Figure 5.9 c). The y-axis corresponds to the main rib of the leaf. The x-axis is a fictive axis with a 45° angle with respect to the y-axis. The main objective of these measurements was to quantify the leaf chlorophyll-content heterogeneity. Working with a 45° angle was interesting to make sure that measurements were made in different secondary ribs to better estimate the within-leaf chlorophyll-content variability. One of the two sides of each leaf was cleaned to remove dust and phytosanitary treatment residues. This was done because deposits are likely to be more important on aged leaves, i.e., with advanced ranks and phenological stages. To avoid any confusion with a possible leaf lateralization effect (the left side of the leaf always unfolds first), the side of the leaf that was cleaned was not always the same. In total, four sources of chlorophyll-content variability were investigated, i.e., the variabilities due to (i) the banana plant growing conditions (that are assumed

to be different amongst the investigated fields), (ii) the banana plant phenological stages, (iii) the banana plant's architecture (location of the measurements on the leaf, side and rank of the leaf), and (iv) the leaf cleanliness.

### Data analysis

The influence of the sources of chlorophyll-content variability was characterized using a mixed model with nested factors. These models are an extension of the fixed models to which random effects are added to the model residuals (Equation 5.2). One of the main limits of fixed models is that they assume that the response variable is independent amongst the models individuals, e.g., between the chlorophyll-content of different leaves on the same banana plant in this case study. Often, this assumption of independence cannot be accepted, especially when measurements are nested, i.e, made at different levels, on the same individual. The main interest of adding random effects, and therefore to work with mixed models, is to not only simply represent a population effect but also to quantify the variability that is due to each individual. Modelling these sources of variability also enables to model the relationship between observations made at the same hierarchical level and to dispose of unbiased statistical tests on fixed effects.

For the first experiment, the inter-field, inter-plant, and intra-leaf variabilities were evaluated with a mixed linear model (Equation 5.2). The « fields » effect was modelled by a fixed effect as fields were selected by the farmers. The « banana plants » effect was modelled by a random effect nested to field, as they were randomly selected in each field :

$$Chl_{b,l,i} = \mu + \mu_{field_b} + \alpha_{stage_b} + \beta_1 * rank_l + \gamma_1 * stage_b * rank_l + \beta_2 * rank_l^2 + \gamma_2 * stage_b * rank_l^2 + a_b + m_b * rank_l + c_l + \varepsilon_{b,l,i} \quad (5.2)$$

Where  $Chl_{b,l,i}$  is the  $i^{th}$  chlorophyll observation that was made on the banana plant  $b$ ,  $\mu$  is the population's reference,  $\mu_{field_b}$  is the constant of the field in which  $b$  can be found,  $\alpha_{stage_b}$  is the random effect of the  $b$  that is the difference between the mean of observations on  $b$  and the mean of the observations in the field where  $b$  is located,  $\varepsilon_{b,l,i}$  are the model's residuals.

It is assumed that the variables  $a_b$  and  $m_b$  are independent and identically distributed and that  $c_l$  is normally distributed. Similarly,  $\alpha_{stage_b}$  and  $\varepsilon_{b,l,i}$  are independent and it is assumed that  $\beta_1$  and  $\beta_2$  are independent.

For the second experiment, the effects of the banana phenological stage, leaf rank, within-leaf measurement location, leaf cleanliness and leaf side were modelled by fixed effects. The effects of the banana plant nested within the field and the leaf nested within the banana plant are represented by nested random effects.

$$Chl_{b,l,i} = \mu + \mu_{field_b} + \alpha_{stage_b} + \beta_1 * rank_l + \gamma_1 * stage_b * rank_l + \beta_2 * rank_l^2 + \gamma_2 * stage_b * rank_l^2 + \mu_{side_l} + \mu_{clean_l} + \delta_1 * e^{x_i} + \delta_2 * e^{y_i} + \delta_3 * e^{x_i} * e^{y_i} + a_b + m_b * rank_l + c_l + \varepsilon_{b,l,i} \quad (5.3)$$

} Fixed effects  
} Random effects  
} Residuals

Where  $Chl_{b,l,i}$  is the  $i^{th}$  chlorophyll measurement made on leaf  $l$  on banana plant  $b$ .

Fixed effects :  $\mu$  is the population's reference ;  $\mu_{field_b}$  is the constant of the field in which  $b$  can be found ;  $\alpha_{stage_b}$  is the effect of the phenological stage of  $b$  ;  $\beta_1$  and  $\beta_2$  represent the linear and quadratic coefficients between the chlorophyll-content and leaf rank ;  $\gamma_1$  and  $\gamma_2$  represent the interaction terms between the phenological stage and leaf rank ;  $\delta_1$  and  $\delta_2$  are constants related to the cleanliness and side of leaf ;  $\delta_3$ ,  $a_b$  and  $m_b$  are the coefficients of the relationship between the chlorophyll and the location of the observations on the  $x$  and  $y$  axes of leaf  $l$ .

Random effects :  $a_b$  and  $m_b$  represent the random intercept and slope associated with  $b$ ,  $\varepsilon_{b,l,i}$  is the random effect of  $b$ . Variables  $\alpha_{stage_b}$  and  $\varepsilon_{b,l,i}$  model the fact that the relationship between the leaf rank and leaf chlorophyll-content varies across banana plants around the mean relationship represented by the fixed

effects. Variable  $\epsilon_{ijk}$  models the fact that the mean chlorophyll content of  $i$  nested within  $j$  can deviate from the mean linear relationship between the chlorophyll content and the rank of the leaves of  $k$ .

Model's residuals :  $\epsilon_{ijk}$  are the residuals of the model. It is assumed that  $\epsilon_{ijk} \sim N(0, \sigma^2)$ ,  $\epsilon_{ijk} \perp \mu_i$ , and that  $\mu_i$ ,  $\sigma^2$ , and  $\epsilon_{ijk}$  are mutually independent.

Note that the relationships between the chlorophyll-content and the leaf rank, observations locations on axes  $x$  /  $y$  and phenological stages are not linear (Equation 5.3). This choice was made because the assumptions of linearity were not met when looking at the structure of residuals. The effects of the fixed factors were tested with a top-down approach using the Bayesian information criterion (BIC). Once the structure of the fixed effects was determined, the estimation parameters was done using the restricted maximum likelihood (REML). Model were fitted with the nlme package (Pinheiro et al. 2018) in R (R Core Team 2018).

## 5.2.4 Resultats and discussion

### Analysis of model 1

Figure 5.10 presents the chlorophyll-content variability of the 3-rank leaf for flowering banana plants on the 15 fields under study. Boxplots highlight variations in the leaf chlorophyll-content at multiple scales, i.e. between fields (different colors), within fields (same color), and in the 3-rank leaf of each banana plant (same boxplot). In model 1, the fixed effect « field » is significant meaning that the growing conditions set out by farmers have an influence on the leaf chlorophyll-content. The variance estimations  $\sigma^2$  and  $\sigma^2_{\epsilon_{ijk}}$  are 18.41 g /cm<sup>2</sup> and 12.43 g /cm<sup>2</sup>, respectively. The  $\sigma^2_{\epsilon_{ijk}}$  estimation shows a banana plant effect and also corresponds to the covariance between observations made on the same banana plant and justifies the use of a mixed model. Finally,  $\sigma^2_{\epsilon_{ijk}}$  shows the residual variability. It can be due to the leaf heterogeneity (leaf thickness, rib proportion) that can affect the Dualex measurements, to the chlorophyll gradients within the leaf, and to the acquisition condition (operator, temperature). It must be noted that the variability related to the repeatability of the measurements is weak if one refers to the standard deviation obtained by Cerovic et al. (2012) i.e., 0.13 g/cm<sup>2</sup>. These results demonstrate the impact of the banana growing conditions on the leaf chlorophyll content in a commercial plantation and the possible interest of non-destructive measurements to better assess these sources of variability. However, this experiment was done with plants at the same phenological stage. The objective of the following experiment is to better characterize the sources of variability related to the banana architecture and phenological stages within a field.

### Analysis of model 2

The analysis of the fixed effects of model 2 (Equation 5.3) is presented in Table 5.3 and Figure 5.11. Apart from  $\mu_i$ , the remaining fixed factors of model 2 had a significant influence on the leaf chlorophyll-content. This result shows that in banana plants, cleaning of the leaves is not mandatory in order to obtain reliable measurements. The importance of the « field » effect that was already highlighted in the first experiment was here again confirmed (Table 5.3). However, the classification of fields according to their average chlorophyll content is different between Experiment 1 and Experiment 2 (  $\mu_i$  in Table 5.3 and mean levels in Figure 5.10). The sampling design for Experiment 1 was built to ensure that the measurements are representative of the mean chlorophyll level of the fields. This was not the case for experiment 2, whose objective was to see the effect of phenological stages on banana plants located in the same environment. These differences in the sample design may explain the different classification between the two experiments.

Whatever the banana phenological stage, the chlorophyll-content varied with the leaf rank (Figure 5.11). The quadratic relationship between the chlorophyll-content and leaf rank was found to be dependent

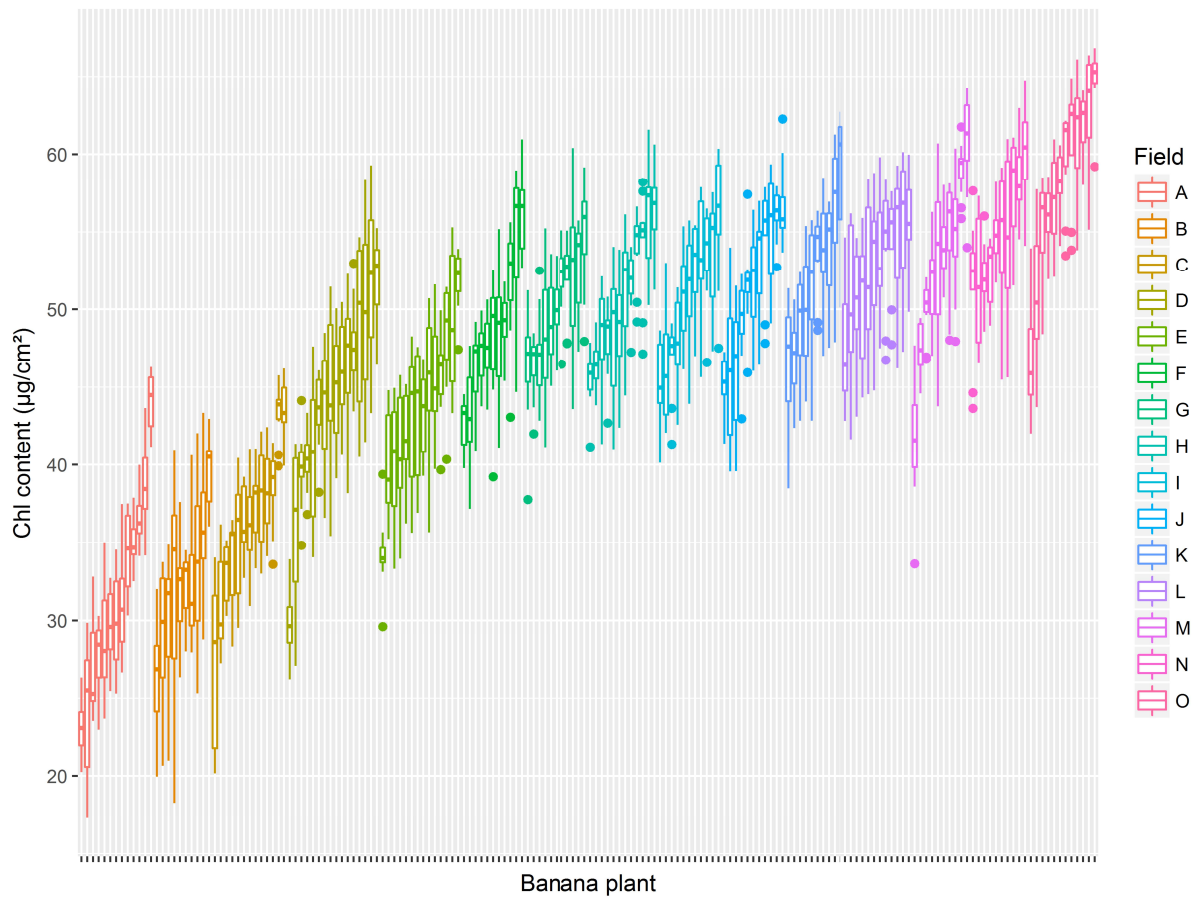


FIGURE 5.10 – Chlorophyll-content variability in the 3-rank leaf for flowering banana plants in the 15 fields under study

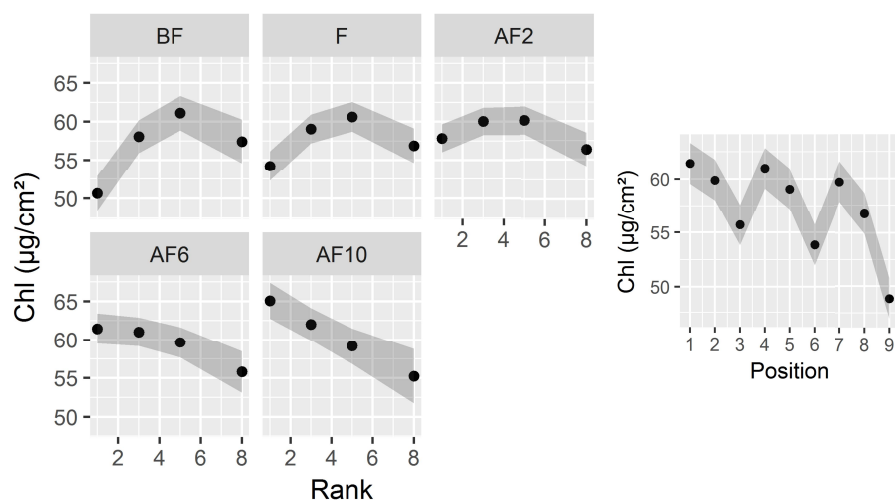


FIGURE 5.11 – a) Mean effect of the leaf rank and phenological stage on the chlorophyll content b) Mean effect of the within-leaf observation's location on the chlorophyll content. The hull represents the confidence interval at 95 %

on the phenological stage. When looking at the 1-rank leaf, chlorophyll-contents were weak for stages BF, F and AF2, whereas they were much stronger for stages AF6 and AF10. The chlorophyll-content also varied spatially within a leaf. The chlorophyll-content decreased as measurements moved away from the central rib (Figure 5.11). The left side of the leaf had a systematically higher chlorophyll-content than the right side. The chlorophyll-content also decreased along the y-axis, but more slightly. It must be noted that the mean difference between measurements made at positions 4 and 9 was  $12 \text{ g}/\text{cm}^2$ . This is in the same range as the difference between the chlorophyll-contents of the 1-rank and 3-rank leaves at the BF stage ( $11 \text{ g}/\text{cm}^2$ ). The effect of the phenological stage (in Table 5.3) is not interpreted here. This is because banana plants at different phenological stages appeared at different dates and certainly experienced different growing conditions. The experiment did not allow discrimination between the effect of the past growing conditions and the effect of the banana plant phenological stage. The variance associated with the random effects in model 2 was not compared to the one in model 1 because sampling was different in both experiments (within-leaf measurements' positions, choice of the banana plants within the fields).

### 5.2.5 Conclusion

This study highlighted that the growing conditions of the fields, the asynchronism in the banana development cycles and the banana plant architecture had an effect on the chlorophyll content of the leaves. When measuring leaf-chlorophyll content on banana crops with a handheld chlorophyll meter, the phenological stage, rank and position can be known and controlled during the sampling. Using remote sensing, the interpretation of the maps for diagnoses purposes should be done with greater caution than for those obtained with ground-based measurements because of all the aforementioned sources of chlorophyll-content variability. Further research is needed before considering nutritional diagnoses services for banana crops based on remote sensing.

TABLE 5.3 – Parameter estimates of Model 2 through REML

Parameter	Estimation	Standard Error	Parameter	Estimation	Standard Error
	29.14	1.85		29.14	1.85
	-2.70	1.77		-1.88	0.22
	4.00	1.90		0.14	0.22
	7.05	1.77		1.33	0.12
	7.91	1.60		-0.28	0.014
	7.00	1.89		-0.05	0.014
	4.89	1.77		-0.017	0.0011
	12.55	1.57		9.47	3.07
	5.37	0.48		0.20	0.44
	7.78	0.64		4.58	2.14
	-0.69	0.07		15.18	3.90

### 5.2.6 Acknowledgments

The experiments were made in Plantation du Haut Penja with the support of its agronomy department. We are grateful to all the plantation staff and particularly E. Andjengo, P. Dontsop, P. Esnault, R. Nkongo, A. Pugeaux, J. Simo and R. Solano for their help with the experiments.

## 5.3 Conception d'une chaîne de traitement pour la prédiction par drone du taux de chlorophylle dans les bananeraies

### A multispectral processing chain for chlorophyll content assessment in banana fields by UAV imagery

Gilles Rabatel<sup>1</sup>, Julien Lamour<sup>1,2</sup>, Daniel Moura<sup>1</sup>, Olivier Naud<sup>1</sup>

<sup>1</sup> ITAP, Montpellier SupAgro, Irstea, Univ Montpellier, Montpellier, France

<sup>2</sup> Compagnie Fruitière, 13000 Marseille, France

#### 5.3.1 Abstract

A specific processing chain has been designed to build leaf chlorophyll content ( ) prediction maps in banana fields using multispectral UAV imagery. A supervised regression model was first calibrated out-of-field using multispectral images of collected leaves laid on the ground, and then applied to build a ortho-mosaic for a complete field using the same multispectral sensor. Though the quality of the resulting map was limited mainly due to light scattering inside the canopy, a significant correlation with within-field vigour heterogeneity has been found, making this approach promising for further banana production management.

#### 5.3.2 Introduction

Banana production for export is an input intensive agricultural system organised in large area plantations (several hundred hectares). To optimise the use of agricultural inputs with a precision farming approach, characterisation of the spatial variability induced by growing conditions is necessary. This is presently done by means of in-field data measurements, representing a very labour intensive task. Moreover,

due to natural asynchronism of banana plants, various phenological stages usually occur simultaneously in the same field, inducing specific sampling difficulties.

On the other hand, Unmanned Aerial Vehicle (UAV) imagery is more and more used for the spatial characterisation of various crops, thanks to its low cost and its ability to measure vegetation characteristics over large fields. Such advantages would obviously be of great interest for banana production but feasibility studies are still required. The objective of the present study was to evaluate the feasibility of using an UAV equipped with a multispectral imaging sensor as an alternative to ground measurements in banana fields. As a first step, the leaf chlorophyll content ( ) has been chosen as variable of interest, due to its strong relationship with the plant vigor and nitrogen status and its ability to be accessed through optical measurements (Cerovic et al. 2012).

### 5.3.3 Material and methods

In order to build maps from multispectral data, a supervised prediction model based on partial least square regression (PLS-R) has been chosen (Wold et al. 2001). In contrast to spectral based vegetation indices like MCARI (Modified Chlorophyll Absorption in Reflectance Index) (Daughtry et al. 2000), supervised models require a learning dataset. In counterpart, they are more likely to be adapted to specific crops and available spectral bands of a given image sensor. In the present case, the learning dataset must combine multispectral imager data with reference values. Due to the practical difficulty in collecting a sufficient number of learning samples in banana fields, a specific procedure has been designed : the PLS-R model was built out-of-field by cutting and collecting individual banana leaves in various growing conditions.

Such a learning procedure presents several advantages : i) reference measurements can be made in better conditions, allowing measurement all over the leaf area, ii) multispectral image data are not biased by leaf inclination and multiple reflection of light inside the canopy (Makdessi et al. 2017), iii) correspondence between imaged leaves and reference measurements is straight forward (instead of very tedious or uncertain in canopy images). In counterpart, it requires a specific data processing chain to overcome technical issues like band co-registration and proper application of the resulting out-of-field model on canopy images acquired by UAV.

#### Calibration of a prediction model

##### Leaf collection and reference measurements

The experiments were carried out in commercial fields of the Plantation du Haut Penja (PHP), located near Njombe in Cameroon (WGS84, E : 4.612, N : 9.639). Eight fields were selected to provide a set of leaves representative of the diversity that can be found in a banana plantation : three fields with a low average reference chlorophyll content (around 35 g/cm<sup>2</sup>), two fields with a medium content (around 45 g/cm<sup>2</sup>) and three fields with a high content (around 50 g/cm<sup>2</sup>). In addition, the leaves were collected at various leaf order numbers on the banana plant (1 to 8, mostly 1, 3 and 5) and at various phenological stages : before flowering, at flowering and 2, 6 or 10 weeks after flowering. In total, 234 leaves were collected and transported to PHP facilities where chlorophyll measurements were made with a Dualex Scientific<sup>+</sup> (Force-A, Orsay, France). Average values per leaf ( stands for Dualex) were estimated from distributed measurements all over the leaves (9 on each leaf side).

##### Multispectral image acquisition of laid leaves

After Dualex measurements, leaves were laid outdoors on a gravel floor made of black pouzollane, and imaged using a multi-spectral sensor Sequoia<sup>R</sup> (Parrot, Paris, France) mounted on a quadrotor UAV (Quad Mikrokopter<sup>R</sup>, Moormerland, Germany). The elevation of the UAV was 10 m above the soil, providing a sufficient resolution (1 cm/pixel) to limit pixel mixing at the strip edges of banana leaves (Figure 5.12).

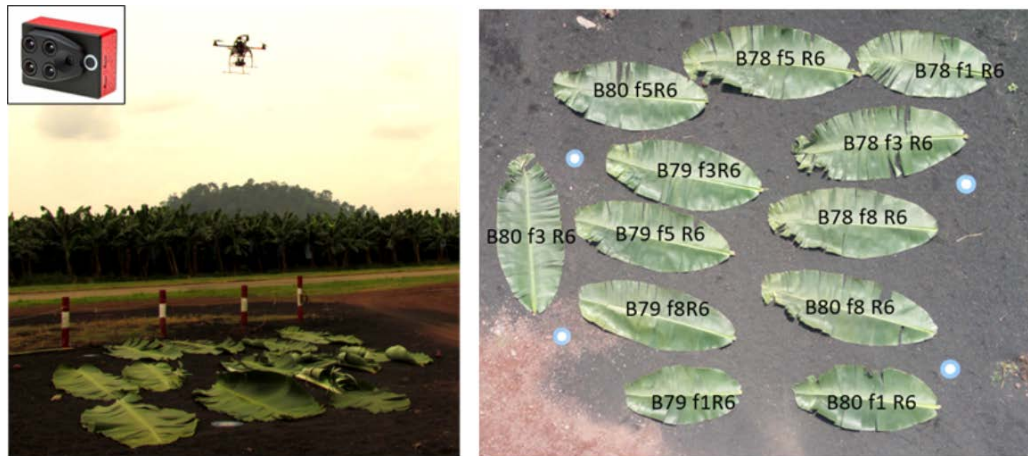


FIGURE 5.12 – Image acquisition of laid leaves. Left : image acquisition (Sequoia sensor shown top-left). Right : example of Sequoia image (RGB channel) with labelled leaves

#### Radiometric correction of multispectral bands

The Sequoia sensor includes a standard RGB camera (image example Figure 5.12 right) as well as four monochromatic cameras providing pixel images in four spectral bands. These bands, named GRE, RED, REG and NIR, have central wavelengths and FWHM (full width at half maximum) of respectively 550 nm and 40 nm (green), 660 nm and 40 nm (red), 735 nm and 10 nm (vegetation red-edge), 790 nm and 40 nm (near infrared). These band images require radiometric correction in order to get a reflectance signal independent of the lighting conditions. For this purpose, the Sequoia sensor comes with a sunshine sensor that has to be mounted on the top of the UAV to measure ambient light during acquisition. This sunshine sensor includes four photoelectric cells with respectively the same spectral sensitivities as the GRE, RED, REG and NIR cameras. Its data are recorded and stored in the Exif header (Exchangeable image file format) of each image file. Camera-specific parameters are also available for raw signal to radiance conversion and image vignetting and distortion corrections, as detailed in Sequoia (2017). All these corrections have been systematically applied in order to obtain a reflectance image for every band. Simultaneously, pixels with saturated raw signals were marked to be eliminated in further processing.

#### Co-registration of multispectral bands

The GRE, RED, REG and NIR camera lenses of the Sequoia are set at the corners of a 15x15 mm square (see Figure 5.12 left). As a consequence, at 10 m elevation, parallax effects create a significant spatial shift between multispectral components (Figure 5.13 left). The co-registration method proposed by Rabatel and Labbé (2016), with the GRE band as master band, was used to correct this shift and build a usable multispectral image (Figure 5.13 right).

#### PLS-R model calibration

PLS-R is usually devoted to spectrometric data involving hundreds of wavelengths, due to its ability to optimize the reduction of the variable space dimensionality according to the output variable of interest. However, it can be used with multispectral data as well, with no dimensionality reduction.

For every collected leaf, its position in the corresponding multispectral image was manually recovered and the average value of (GRE, RED, REG, NIR) pixel components computed. The resulting 234x4



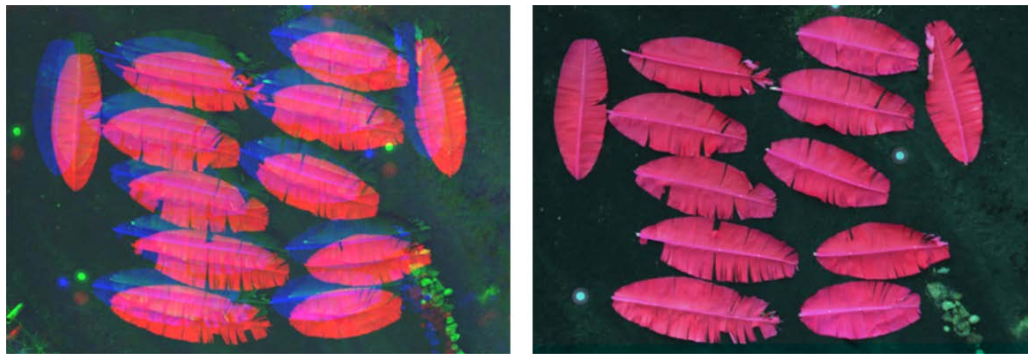


FIGURE 5.13 – Multispectral image before (left) and after (right) co-registration. Images shown in false colors with  $\lambda_{1}$ ,  $\lambda_{2}$  and  $\lambda_{3}$

spectral values were associated with corresponding  $\lambda_{1}$  leaf average values, and then randomly split into a learning set and a test set of respectively 161 and 73 samples (  $\lambda_{1}$  69 % and 31 %).

A PLS-R model was calibrated using the learning dataset to predict  $\lambda_{1}$  ( stands for Sequoia) from the spectral values. The prediction quality was evaluated by calculating the coefficient of determination (  $\lambda_{1}$  ) and the root mean square error (RMSE).

### In-field assessment of the model

#### Field image acquisition by UAV

A particular field among the eight fields where leaves were collected was chosen for UAV image acquisition. This field was chosen because of the heterogeneity of its growing conditions, highlighted by a previous mapping of plant's circumferences. Details on this circumference mapping are given in (Leroux et al. 2018a).

The UAV with the Sequoia sensor was flown over the whole field at an elevation of 30 m above the banana plants. The flight was designed to have an image overlapping of 80 % in the direction of travel and 60 % between travels, giving a total of 230 RGB and 4 band images all over the field (after removing take-off and landing phases).

#### Field UAV image processing

UAV image processing usually relies on the use of specialized mosaicking software to produce a geo-referenced ortho-mosaic from the set of acquired images. The bulk of such software consists in searching correspondence points between overlapping images, and then determining camera 3D-positions and a digital surface model (DSM) of the covered scene. Finally, an ortho-mosaic is built from these data. In the present study, a slightly different approach has been used to ensure the correct implementation of the PLS-R model designed above :

Radiometric and band co-registration procedures (see above) were applied to every original set of four band images, and the PLS-R model was applied on the resulting multispectral images, leading to a set of 230  $\lambda_{1}$  images.

RGB Sequoia images were mosaicked using the commercial software package Photoscan (Agisoft, St. Petersburg, Russia).

Finally,  $\lambda_{1}$  images were assembled to build a  $\lambda_{1}$  map by re-using camera 3D positions and DSM issued from the Photoscan RGB mosaicking process.

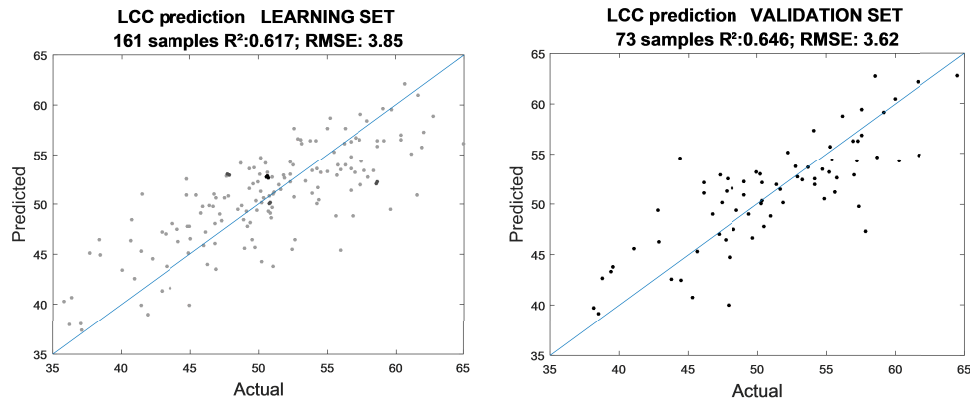


FIGURE 5.14 – estimation by PLS-R on learning and test sets

quality assessment on selected leaves

Before UAV acquisition, 25 leaves were chosen in the field at various phenological stages, and colored papers were stapled on them in order to be visible from the sky. After the UAV acquisition, the 25 marked leaves were collected, brought to the PHP facilities and included in the PLS-R calibration dataset previously described.

Thanks to the colored papers, those leaves were retrieved in the RGB ortho-mosaic, and their average values in the map were calculated and compared with their reference values by means of a linear regression.

Filtering and relevance assessment of the map

The map resolution was much higher than the plant circumference map resolution (resp. 3.3 cm and 1 m per pixel). In addition, the map showed short-scale variations within leaves or within plants. In order to filter out these short scale variations and only highlight inter-plant variations, which were assumed to reflect different growing conditions, the map was averaged on a grid of blocks of 4 x 4 m. The variogram was then computed, modelled and used for interpolating values at the same locations as the circumference map by ordinary kriging (see Oliver and Webster 2014). This final map was then compared to the circumference map.

### 5.3.4 Results and discussion

#### PLS-R model for LCC estimation

The range of values for all collected leaves was about 35-65 g/cm . The PLS-R model calibrated on the learning set gave the following prediction equation :

$$(5.4)$$

Its application on learning and test sets gave respectively coefficients 0.62 and 0.65, with a RMSE of respectively 3.85 and 3.62 g/cm , i.e. 12 % of the total range (Figure 5.14).

The interpretation of Equation 5.4 is not straight forward. While optical prediction of in chlorophyll-meters usually relies only on red-edge and near infrared bands (Cerovic et al. 2012), Equation 5.4 also involves GRE and RED spectral components, in an unexpected way (the more the RED component, the more the predicted ).

Indeed, measurement conditions are slightly different : image reflectance is more sensitive to the shininess of leaf cuticles than transmittance signal of the Dualex. Moreover, the numerous veins of the

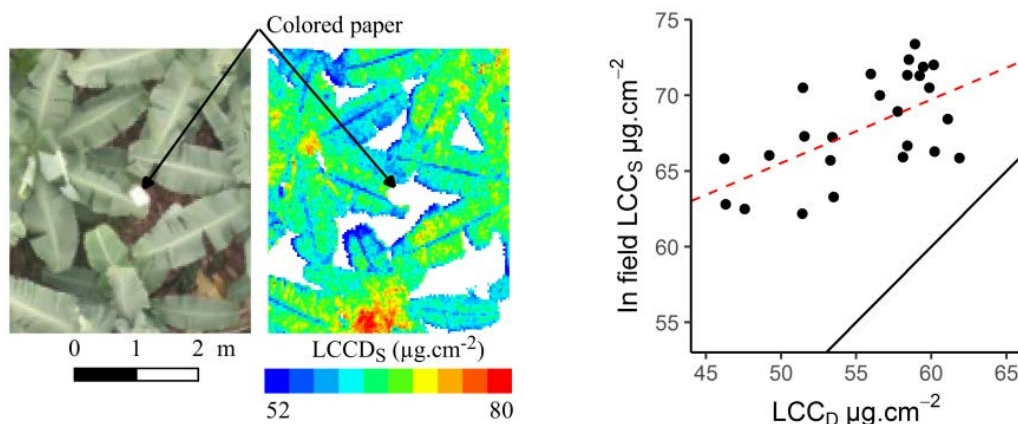


FIGURE 5.15 – Leaf LCC estimation on selected leaves in canopy. Left : RGB and LCCD<sub>S</sub> map detail. Right : In field LCC<sub>S</sub> - LCC<sub>D</sub> relationship

banana leaves are taken into account, but were not included in Dualex measurements. The PLS-R model probably compensates for such variability factors by combining all available spectral bands.

### In-field LCC estimation

Figure 5.15 shows an example of marked leaf retrieved in the RGB and LCCD<sub>S</sub> maps, and the linear regression between LCC<sub>D</sub> and mapped LCC<sub>S</sub> values for all marked leaves. The regression coefficient has been drastically reduced compared to the one obtained when leaves were laid on the soil (0.8 versus more than 0.6), and LCC<sub>S</sub> values are significantly over-estimated.

The low regression coefficient can be explained by the difficulty in correctly estimating the average LCC<sub>S</sub> for marked leaves in the map due to their positions and orientations in the scene, and by multiple light reflections in canopy, which modify the leaf spectra. This last cause is also likely to explain the over-estimation of LCC<sub>S</sub> (Makdessi et al. 2017). Unfortunately, the correction method proposed by the same author (Makdessi et al. 2018) cannot be used due to an insufficient number of spectral bands.

### Assessment of the LCC<sub>S</sub> map to identify agronomic zones

The previous results were acquired at the leaf level, but the agronomic diagnosis of fields is generally carried out at a larger scale in order to identify spatial trends. The map obtained after filtering and kriging is presented in Figure 5.16 (right). Indeed, different LCC<sub>S</sub> zones are visible within the field. Pseudo-trunk circumference maps for adult plants and for all plants are also given in Figure 5.16 (left and middle). Zones of small circumferences are very similar to those of low LCC<sub>S</sub> (Figure 5.16), showing that valuable agronomic information can be obtained from LCC<sub>S</sub> map at this scale. This result must be moderated. In this field, pseudo-trunk circumference maps obtained with adult plants and with all plants are very similar, indicating a uniform mixing of the phenological stages, due to plant asynchronism effect over a long period. In younger fields with more organized phenological stages, biases could be expected.

### 5.3.5 Conclusion

In this study, an original approach for LCC<sub>S</sub> monitoring in banana fields by UAV has been developed, using a multispectral sensor. It relies on a supervised PLS-R model. The learning phase has been made possible by combining reference measurements and image acquisition in optimal out-of-field conditions. As a counterpart, specific image processing developments were required to properly manage multispectral

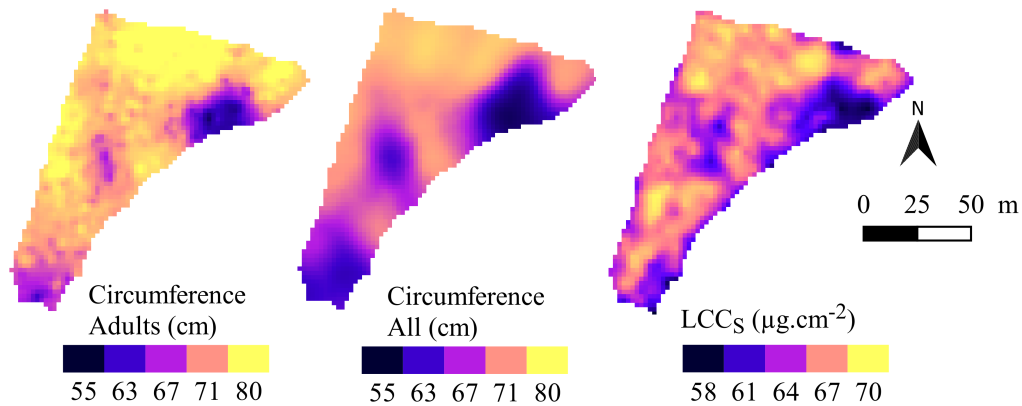


FIGURE 5.16 – Comparison between and vigor maps. Left and middle : circumference maps. Right : map after kriging

image radiometry both in model calibration and model application to aerial canopy images. In terms of results, prediction performance was limited when applied to field images, probably due to canopy scattering effects. More spectral information (e.g. hyperspectral imagery) would be required both to improve the prediction model and investigate canopy effect corrections. However, this UAV approach remains a promising tool for within-field heterogeneity monitoring at large spatial scale, at least for fields where plant phenological stages are uniformly distributed.

## 5.4 Perspectives d'utilisation de ces deux capteurs dans les bananeraies

### 5.4.1 Utilisation du Dualex

D'autres résultats issus de l'analyse des données montrent qu'il y a une corrélation faible entre la teneur en chlorophylle et la teneur en azote pour la feuille 3 des bananiers à floraison (Figure 5.17). La relation est améliorée lorsque la teneur en azote est exprimée en masse surfacique (Figure 5.17) mais reste modérée. Si toutes les mesures sont utilisées, quel que soit le stade et la feuille, cette relation a un coefficient de 0.27 (non montré).

Dans certaines parcelles, il y a une corrélation faible entre la teneur en chlorophylle mesurée sur la feuille 3 des bananiers à la floraison et la circonférence des pseudo-troncs mais cette relation n'est pas générale (résultat non montré). La teneur en chlorophylle est donc une nouvelle source d'information pour les producteurs qui ne peut pas remplacer directement les mesures qu'ils font traditionnellement. Cette mesure n'est pas non plus utilisable directement pour faire un diagnostic, c'est-à-dire pour identifier la cause d'un défaut sur lequel les producteurs pourraient agir. En effet, la teneur en chlorophylle dépend de la nutrition minérale, hydrique mais aussi de l'exposition lumineuse (Israeli et al. (1995)) et de l'infestation en cercosporiose (Morel et al. (2018)). Il est donc nécessaire de faire d'autres observations pour identifier la source de carence lorsqu'il y en a une.

Finalement, l'intérêt de ce capteur est de permettre un nombre important de mesures à moindre coût et avec un résultat immédiat ce qui permet facilement de classer les parcelles. Il pourrait donc être utilisé en production en suivant le même protocole que celui que nous avons proposé pour identifier les parcelles carencées en chlorophylle. Pour cela, nous proposons le seuil de 50  $\mu\text{g}/\text{cm}^2$  comme seuil de déficit de chlorophylle (Figure 5.7). Dans les parcelles avec un taux de chlorophylle inférieur à ce seuil, la Compagnie Fruitière pourrait garder ses protocoles actuels de diagnostic et notamment les analyses

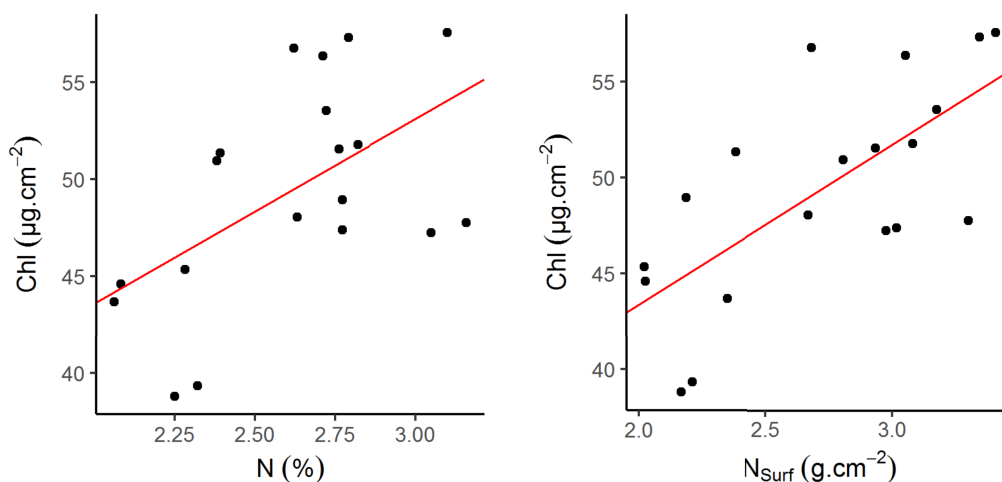


FIGURE 5.17 – Relation entre la teneur en chlorophylle mesurée par le Dualex et la teneur en azote exprimée en pourcentage massique (Figure de gauche, ) ou en masse surfacique (Figure de droite, )

minérales pour identifier les causes de carence. Cela aurait pour intérêt de limiter le nombre de parcelles où des analyses chimiques sont réalisées. Le seuil que nous proposons devrait être évalué à d'autres saisons et sur d'autres cultivars pour que cette démarche soit applicable dans un plus grand nombre de situations.

## 5.4.2 Utilisation de la télédétection et du capteur Sequoia

### Télédétection

Des sociétés proposent régulièrement des services de télédétection à la Compagnie Fruitière. Parmi les fonctionnalités proposées on retrouve essentiellement : le dénombrement des pieds, la mesure de la surface de feuilles et de sol, la mesure d'indices de vigueur (NDVI), l'évaluation du stress hydrique ou du degré d'infestation des maladies. Les chaînes de traitement ne posent généralement pas de problème pour créer des cartes référencées et à forte résolution (cm). Leur avantage est de pouvoir proposer ces indices sur des grandes surfaces ce qui serait pertinent pour la Compagnie Fruitière. Cependant, il est difficile de connaître la performance de leur services notamment sur des indices qui ont été développées sur d'autres cultures ou pour lesquels il n'y a pas de référence dans les bananeraies (vigueur, stress hydrique, détection de maladies). En dehors de la faisabilité technique, il est aussi difficile d'identifier l'utilité de ces services et notamment les actions correctives qui pourraient découler des informations fournies.

Le travail que nous avons réalisé pourrait servir de référence à la Compagnie Fruitière pour tester les services proposés par ces sociétés. Les tests consisteraient à comparer les prédictions réalisées par ces sociétés avec des mesures terrains, soit dans des conditions optimales (feuilles à plat) soit dans des conditions terrains (feuilles dans la canopée, comparaison de cartes).

Le protocole que nous avons proposé dans un champ devrait être testé sur une surface plus grande qui comprendrait plusieurs parcelles. Dans ce cas, la difficulté serait que les parcelles ne seraient pas au même stade de développement en terme de mixité des stades phénologiques et de stade phénologique moyen. Pour s'assurer que les données obtenues soient comparables il faudrait être en mesure de ne comparer que les parcelles dans les mêmes conditions. On suppose que leur stade de développement peut être évalué par les producteurs qui connaissent leur date de plantation. Les indices d'hétérogénéité des stades phénologiques ( ) et d'avancement moyen ( ) présentés dans la Section 3.3 pourraient aussi être utilisés.

## Chapitre 6

# Conclusion

Suite à la mécanisation des cultures, au remembrement des parcelles et à l'agrandissement des exploitations, les petites parcelles qui étaient gérées individuellement ont été regroupées et cultivées de manière uniforme. L'agriculture de précision s'est développée pour répondre à l'enjeu de l'uniformité des pratiques agricoles dans des contextes pédo-climatiques variés. Cette approche a permis des innovations techniques et les exploitations agricoles sont aujourd'hui de plus en plus équipées de capteurs qui permettent de mesurer la variabilité intra-parcellaire. Les équipements sont aussi adaptés pour moduler leur action dans les parcelles. Le contexte de production de la culture bananière est différent. En effet, les exploitations s'appuient sur une présence humaine à l'hectare très élevée et largement supérieure à celle des grandes cultures. De plus, les parcelles sont relativement petites, quelques hectares, et la culture est très peu mécanisée. C'est aussi une production continue dans l'année et asynchrone ce qui la différencie fortement des cultures sur lesquelles l'Agriculture de Précision s'est développée. Les enjeux sont aussi un peu différents, car l'objectif n'est pas seulement d'augmenter les rendements mais aussi de limiter les risques liés à une maturité trop avancée qui entraîne des pertes commerciales. Malgré ces différences structurelles, l'objectif de réduire l'échelle spatiale de décision et de moduler les pratiques est commun avec les cultures « synchrones ». Actuellement, les règles de décisions utilisées pour gérer les récoltes sont homogènes et les pratiques sont uniformes sur de très grandes surfaces. L'objectif de valoriser les données de production pour identifier les sources de variabilité et *in fine* adapter les pratiques est donc tout aussi pertinent.

Pour répondre à cet enjeu opérationnel, nous avons analysé les données recueillies par la Compagnie Fruitière. Cela nous a permis de montrer qu'elles pouvaient être utilisées pour faire un état des lieux des conditions de production et dans certains cas un diagnostic (nous y reviendrons). Cela a abouti à la création d'un package R appelé « PreviFlo » qui permet aussi de faire des prévisions, ce qui n'a pas été montré dans ce mémoire. Des méthodes permettant de faire des cartographies de l'effet des conditions de croissance sur la culture ont aussi été mises en place. Ces outils devront faire l'objet d'une étude particulière pour être intégrés au mieux dans le système d'information de la Compagnie Fruitière et répondre à ses besoins en termes d'interfaces et d'utilisations. Pour répondre aux questions de la Compagnie Fruitière sur l'intérêt des nouvelles technologies nous avons aussi mené des expérimentations dans les bananeraies au Cameroun avec deux capteurs optiques (Chapitre 5). Ces expérimentations ont surtout permis de montrer des méthodes qui permettent de caractériser ces technologies par rapport à l'expertise traditionnelle. Dans le cas de l'usage des drones, je pense que le protocole mis en place peut servir de « banc d'essai » pour d'autres capteurs ou directement pour tester les services de sociétés.

La thèse devait aussi répondre à des questions scientifiques que nous rappelons ici : Est-il possible d'appliquer une approche d'agriculture de précision à la culture bananière malgré l'asynchronisme ? Est-ce que les données recueillies par les producteurs pour des besoins de gestion peuvent être utilisées pour faire un diagnostic de la production, c'est-à-dire pour identifier les facteurs qui en affectent le développement ? Ces problématiques ont été évaluées en répondant à des sous-questions sur l'asynchronisme :

Comment estimer la diversité des stades phénologiques et le stade phénologique moyen des parcelles ?

A quelle vitesse se met en place la diversité des stades phénologiques dans les parcelles ?

Est ce que la vitesse dépend des conditions de culture de la parcelle ?

D'autres questions sur la structure spatiale du développement des bananeraies ont aussi été abordées :

Les propriétés végétatives et reproductives des bananiers sont elles variables dans l'espace ?

La mesure de la structure spatiale est elle biaisée par l'asynchronisme ?

Est-il possible de corriger les effets de l'asynchronisme et comment ?

Nous faisons une synthèse des réponses à ces questions dans les parties suivantes.

## 6.1 Méthodes permettant de caractériser l'asynchronisme d'un peuplement de bananiers

L'asynchronisme des bananiers correspond à une variabilité de la vitesse de développement des plants et a été décrit par plusieurs auteurs dont Tixier et al. (2004). Cela s'explique par l'absence de deux phénomènes de synchronisation qui sont souvent présents sur d'autres cultures :

la phénologie du bananier n'est pas synchronisée par des variations saisonnières du milieu (photopériode, température) ce qui fait que les floraisons puis les récoltes sont réparties dans le temps ;

le début des cycles culturaux suivant chaque récolte n'est pas synchrone, car l'apparition des rejets sur les rhizomes est continue et parce que la sélection du rejet fils par les producteurs n'est pas synchrone.

Cela a deux conséquences :

la production est continue toute l'année ;

les stades phénologiques sont différents entre les parcelles et au sein des parcelles.

Le chapitre 3 de la thèse propose plusieurs méthodes pour caractériser l'asynchronisme d'une parcelle de bananiers ou d'une plantation. Ces méthodes sont utilisables à partir des données de floraison des producteurs qui sont mesurées systématiquement pour gérer les récoltes. Elles permettent d'estimer plusieurs indicateurs :

la durée du Cycle de Développement des bananiers ;

la variabilité de au sein des parcelles ou d'un ensemble de parcelles ;

le stade phénologique moyen ;

la diversité des stades phénologiques .

Ces indicateurs ont été étudiés sur une plantation commerciale au Cameroun à partir des données du système d'information de la production et ont permis de montrer que :

La durée est variable et auto-corrélée spatialement à l'échelle inter-parcellaire. Nous avons aussi montré que cette variable était auto-corrélée au niveau intra-parcellaire pour une parcelle. Ce dernier résultat, même s'il n'a pas été montré sur d'autres parcelles, est *a priori* généralisable car d'autres propriétés comme la circonférence des pseudo-troncs ou le temps de maturité sont structurées spatialement au niveau intra-parcellaire.

La durée dépend des caractéristiques du milieu ainsi que des choix des producteurs (altitude, accès à la lumière, irrigation, cultivar)

La diversité des stades phénologiques au sein d'une parcelle augmente au cours du temps jusqu'à atteindre une limite maximale. Nous avons vu que la vitesse de désynchronisation des bananiers dépend des parcelles. Les deux facteurs qui agissent sur la vitesse de désynchronisation sont la durée moyenne de et l'hétérogénéité des conditions de croissance locales ( ). Plus la durée des cycles de développement est courte et plus les conditions de croissance sont hétérogènes plus les bananiers d'une parcelle se désynchronisent vite.

Les stades phénologiques sont auto-corrélés spatialement mais cette auto-corrélation diminue au fur et à mesure des cycles de développement de la parcelle même si elle est auto-corrélée spatialement. Cela a pu être montré par simulation sur une parcelle et confirme l'expertise des producteurs.

## 6.2 Prise en compte de l'asynchronisme pour analyser les conditions de croissance dans les bananeraies

Dans le cadre de l'agriculture de précision, la production d'une parcelle est étudiée comme le résultat d'une expérience scientifique d'homogénéité. Les conditions de croissance, dont on cherche à déterminer si elles sont variables, sont mises en évidence par la comparaison spatialisée des caractéristiques de la culture. Classiquement, l'étude de la structure spatiale des conditions de croissance se fait à partir d'un variogramme. Dans les bananeraies, le premier cycle de développement des bananiers peut être étudié comme le résultat d'une expérience d'homogénéité. S'il y a des différences de développement des bananiers, elles seront liées à des variabilités du milieu ou de l'itinéraire technique ou à la variabilité naturelle du cultivar. Dès les cycles de production suivants, les stades phénologiques seront variables et leurs différences ne seront plus uniquement dues aux conditions de croissance mais aussi au mode de culture (mort du bananier parent et sélection d'un rejet). Dans ce cas, si les propriétés de la culture sont analysés comme le résultat d'une expérience d'homogénéité, on s'attend à ce que le variogramme présente deux défauts :

- une augmentation de la variance pépîte

- une modification du modèle de structure spatiale (portée  $\downarrow$ , modèle théorique, palier partiel  $\downarrow$ )

En conséquence, les cartes produites par Krigeage ponctuel ou Krigeage par bloc à partir du variogramme vont présenter une variance d'estimation plus élevée (bruit) et un motif différent (biais spatial) que dans le cas d'une parcelle en premier cycle.

L'augmentation de l'effet pépîte a des conséquences sur l'échantillonnage : un nombre plus élevé d'observations sera nécessaire pour mettre en évidence la structure spatiale liée aux conditions variables de croissance. Par contre, cette condition réalisée, l'augmentation de l'effet pépîte n'est pas très problématique pour la carte finale et le diagnostic car le krigeage permet de représenter les tendances spatiales en diminuant l'effet du bruit.

La question de la modification de la structure spatiale est plus problématique. En effet, la carte obtenue par Krigeage reflètera à la fois les conditions de croissance et les différences structurées de stades phénologiques (biais). Dans ce cas, il est nécessaire de corriger l'effet des stades phénologiques pour isoler les conditions de croissance dans le signal mesuré sur le couvert.

Il est possible de définir des règles générales qui définissent à quels moments l'analyse spatiale présentera plutôt une augmentation de  $\sigma^2$  ou une modification de la structure spatiale. Comme nous l'avons vu, le premier cycle de développement des bananiers peut être analysé sans bruit et sans biais. Pour les cycles de développement suivants, on s'attend à ce qu'il y ait une diversité des stades phénologiques et qu'elle soit structurée spatialement. Dans ce cas, il y aura donc une augmentation de  $\sigma^2$  et une modification de la structure spatiale. Lorsque la parcelle est établie, les stades phénologiques sont uniformément répartis, on s'attend donc à ce qu'il y ait seulement une augmentation de  $\sigma^2$ .

Le Chapitre 4 propose une méthode de correction des données qui permet de diminuer l'effet des deux défauts. Cette méthode s'appuie sur la connaissance des stades phénologiques et du moment des mesures. Elle permet de présenter un variogramme corrigé et une carte corrigée.

Dans le cas de la télédétection et quand les plants ne sont pas géolocalisés, les stades phénologiques ne sont pas connus. En conséquence, la méthode de correction proposée ne peut pas être utilisée. En reprenant les règles générales qui viennent d'être définies on suppose que seul le signal mesuré sur des parcelles établies sera interprétable. Les différences entre ces parcelles ne pourront être dues qu'à des différences de condition de croissance car ces parcelles présenteront la même diversité des stades phénologiques et le même stade phénologique moyen. Les parcelles plus jeunes présenteront des stades phénologiques



différents ce qui pourra modifier le signal. Il serait possible d'utiliser le modèle que nous avons proposé Chapitre 3 pour caractériser l'asynchronisme des parcelles afin de comparer uniquement le signal des parcelles comparables (même diversité des stades phénologiques et même stade phénologique moyen).

### 6.3 Exploitation de la variabilité des conditions de croissance en bananeraies

Mettre en place une démarche d'agriculture de précision dans les bananeraies aurait pour objectif de moduler l'itinéraire technique par zone agronomique afin d'optimiser l'utilisation des intrants et les coûts de production, augmenter les rendements et enfin diminuer les impacts environnementaux négatifs de la culture. Pour que ce type d'approche soit pertinente, il est nécessaire d'identifier une variabilité spatiale structurée du fonctionnement agronomique des bananeraies et de pouvoir la caractériser par des mesures. Nous avons vu que les propriétés mesurées suivantes étaient souvent auto-corrélées spatialement ce qui valide cette condition :

La durée des cycles de floraison (Chapitre 3) à l'échelle inter-parcellaire (Section 3.1) et à l'échelle intra-parcellaire sur une parcelle (Section 3.2) ;

La durée entre la floraison et la maturité ( ) des régimes à l'échelle inter-parcellaire (Chapitre 4). Ce même indicateur présentait aussi une auto-corrélation spatiale à l'échelle intra-parcellaire et cela a été montré pour la Ferme A lors d'une communication à ECPA 2017 qui n'a pas été reprise dans le mémoire (Lamour et al. 2017) ;

La circonférence des pseudo-troncs à l'échelle intra-parcellaire (Section 5.3). Ce résultat a aussi été obtenu sur d'autres parcelles lors d'expériences préliminaires, une parcelle en Guadeloupe et 5 autres parcelles au Cameroun (résultats non montrés) ;

La teneur en chlorophylle de la parcelle présentée Section 5.3.

La plantation sur laquelle nous avons fait nos analyses est très vallonnée (Figure 3.2). De manière générale, sur des exploitations de grande taille comme celle étudiée dans ce mémoire, l'environnement est rarement homogène spatialement. Nous nous attendions donc à pouvoir mettre évidence des variabilités spatiales. Les méthodes développées et mises en œuvre dans la thèse ont permis de quantifier objectivement ces variabilités. Les données facilement accessibles comme les données de pente ou d'altitude, ainsi que les données sur l'état des parcelles (cultivar, irrigation) ont permis de comprendre les causes responsables d'une partie de la variabilité observée.

### 6.4 Intérêt et limites de l'agriculture de précision dans les bananeraies

L'Agriculture de Précision est apparue dans les grandes cultures avec les capteurs de rendements qui permettent d'estimer le poids moyen de la récolte par surface (Chapitre 2). Je n'ai pas pu exploiter ce type de donnée dans la thèse car l'automatisation de la mesure des poids dans les stations de conditionnement n'était pas finalisée. C'est une donnée importante pour faire un état des lieux de la production car le poids des régimes résulte à la fois des conditions de croissance mais aussi de la qualité de l'itinéraire technique. L'absence de cette donnée limite donc pour l'instant l'application d'une agriculture de précision classique basée sur le rendement. Les poids des régimes sont des données similaires aux données de étudiées dans le Chapitre 4 dans le sens où ils sont mesurés à un stade phénologique constant (maturité de récolte) et en continu toute l'année. Comme pour les , il y a un effet temporel sur le poids des régimes qui est lié aux modifications de pratiques des producteurs (choix des dates de récolte, choix du nombre d'ablations de mains sur les régimes, etc.) mais aussi aux modifications saisonnières climatiques qui agissent sur la croissance. La méthode proposée Chapitre 4 pour supprimer le biais temporel lié à

la production continue dans le temps pourrait être directement appliquée pour corriger et interpréter les données brutes de poids et établir des cartes exploitables. Nous rappelons que le poids des régimes, la durée des cycles de développement CD et la durée entre la floraison et la récolte sont trois composantes du rendement complémentaires avec des relations complexes entre elles. La possibilité de les mesurer sans biais toutes ensemble et de les représenter sur des cartes offre donc des perspectives opérationnelles intéressantes.

Nous avons vu que les données enregistrées par la Compagnie Fruitière pour sa gestion de production peuvent être représentées sur des cartes. Celles-ci synthétisent le résultat de la production sur de grandes surfaces (l'étude présentée Chapitre 3 sur les      a été réalisée sur 724 parcelles représentant une surface d'environ 3200 ha, l'étude réalisée sur les      et présentée Chapitre 4 s'appuie sur les données récoltées sur 55 parcelles représentant une surface de 252 ha). Dans les bananeraies industrielles qui occupent de grandes surfaces et dont la gestion est assurée par un grand nombre de personnes, la cartographie des propriétés des cultures permet d'obtenir une vue d'ensemble de la production et d'en faire un état des lieux. Ceci est la première étape d'une démarche d'agriculture de précision.

Dans certains cas, les données permettent aussi de faire un diagnostic à caractère agronomique, c'est-à-dire d'identifier les sources de variabilités. Nous avons par exemple identifié l'effet de la topographie et de l'accès à la lumière (bordure de parcelle) sur le      et      , ainsi que l'effet de l'irrigation et du cultivar sur      . Ce diagnostic est partiel car de nombreuses informations sont manquantes pour identifier l'origine de la variabilité observée de manière plus exhaustive. Les données de productions sont enregistrées à des fins de gestion ou de supervision. Elles sont souvent parcimonieuses en termes d'attributs et un grand nombre de variables permettant de déterminer l'état de la culture ne sont pas enregistrées. Notamment, les données sur les types de sol, sur les événements subis par les parcelles (tornades, mortalité, décisions de gestion, etc.) étaient manquantes au moment de ces études. Une grande partie de la préconisation et de la décision est donc pour l'instant laissée à l'expertise des producteurs pour à la fois comprendre et choisir les actions correctives à mettre en place (correction de la fertilisation, modification de la densité de plants, décision de mise en jachère). La qualité de ces nouvelles informations pour améliorer la décision des producteurs n'a pas été analysée dans la thèse mais devra l'être dans une logique de transfert des méthodes. Dans l'avenir, avec la plus grande accessibilité des appareils mobiles qui permettent la saisie d'information au champ et du fait de l'informatisation de la production, les données seront de plus en plus accessibles et permettront d'améliorer le diagnostic et la préconisation. Malgré l'intérêt d'exploiter les données de production pour le diagnostic agronomique, je pense que ces données ne remplaceront pas le rôle des expérimentations spécifiques.

Des expérimentations limitées dans le temps ou l'espace, éventuellement avec un plan d'expérience simplifié, permettraient de mieux estimer la fiabilité des observations et des estimations réalisées sur la base des données de production. Par exemple, dans le cas du modèle probabiliste de floraison que nous avons proposé, ces expérimentations permettraient de mieux choisir les lois de probabilités qui régissent les phénomènes modélisés (par exemple le temps entre l'apparition de deux rejets successifs, le temps entre l'apparition du rejet et la floraison du plant, etc.). Par ailleurs, en reprenant la classification des approches agricoles de Hill and MacRae (1996) citée dans le Chapitre 2 pour situer l'agriculture de précision, nous pouvons considérer que l'agriculture de précision se situe avant tout dans une approche « Efficiency ». L'analyse des données de production permet en effet de comprendre le système de production en place afin d'aider à la décision. Par contre, ces données ne permettent pas de voir « au delà » du système de production en place. Pour évaluer l'intérêt de substituer des étapes de l'itinéraire technique par d'autres ou pour repenser certaines étapes (approches « Substitution » et « Re-design »), il serait nécessaire de faire des expérimentations spécifiques. Au vu des contraintes qui pèsent sur le système de culture actuel (nouvelle maladie TR4, objectif 0 produit phytosanitaire, etc.) cela semble d'autant plus important. Par exemple, dans une approche de production durable, il serait nécessaire de tester de nouveaux cultivars, de nouvelles démarches de gestion des maladies, etc. Les données de gestion recueillies par les producteurs peuvent accompagner le choix des parcelles expérimentales car elles permettent de bâtir un référentiel des conditions de culture et donc de choisir les sites les plus représentatifs pour mener les essais.



# Bibliographie

- Acevedo-Opazo, C., Tisseyre, B., Ojeda, H., and Guillaume, S. (2009). Spatial extrapolation of the vine (*vitis vinifera* l.) water status : a first step towards a spatial prediction model. Irrigation Science, 28(2) :143–155.
- Antonelli, M. G., Zobel, P. B., Durante, F., and Raparelli, T. (2017). Development of an automated system for the selective harvesting of radicchio. International Journal of Automation Technology, 11(3) :415–424.
- Arantes, A. M., Donato, S. L. R., Silva, T. S., Filho, V. A. R., and Amorim, E. P. (2017). Agronomic evaluation of banana plants in three production cycles in southwestern state of bahia. Revista Brasileira de Fruticultura, 39(1).
- Armstrong, R. D., Fitzpatrick, J., Rab, M. A., Abuzar, M., Fisher, P. D., and O’Leary, G. J. (2009). Advances in precision agriculture in south-eastern australia. III. interactions between soil properties and water use help explain spatial variability of crop production in the victorian mallee. Crop and Pasture Science, 60(9) :870.
- Bachmaier, M. and Backes, M. (2008). Variogram or semivariogram ? understanding the variances in a variogram. Precision Agriculture, 9(3) :173–175.
- Bakache, A., Douzals, J.-P., Bonicelli, B., Cotteux, E., de Lapeyre de Bellaire, L., and Sinfort, C. (2018). Development of a rapid methodology for biological efficacy assessment in banana plantations : application to reduced dosages of contact fungicide for black leaf streak disease (BLS) control. Pest Management Science, 75(4) :1081–1090.
- Bakhsh, A., Jaynes, D. B., Colvin, T. S., and Kanwar, R. S. (2000). Spatio-temporal analysis of yield variability for a corn-soybean field in iowa. Transactions of the ASAE, 43(1) :31–38.
- Balafoutis, A. T., Beck, B., Fountas, S., Tsiropoulos, Z., Vangeyte, J., van der Wal, T., Soto-Embodas, I., Gómez-Barbero, M., and Pedersen, S. M. (2017). Smart farming technologies – description, taxonomy and economic impact. In Progress in Precision Agriculture, pages 21–77. Springer International Publishing.
- Barnes, A., Soto, I., Eory, V., Beck, B., Balafoutis, A., Sánchez, B., Vangeyte, J., Fountas, S., van der Wal, T., and Gómez-Barbero, M. (2019). Exploring the adoption of precision agricultural technologies : A cross regional study of EU farmers. Land Use Policy, 80 :163–174.
- Basso, B., Dumont, B., Cammarano, D., Pezzuolo, A., Marinello, F., and Sartori, L. (2016). Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. Science of The Total Environment, 545-546 :227–235.
- Basso, B., Ritchie, J. T., Cammarano, D., and Sartori, L. (2011). A strategic and tactical management approach to select optimal n fertilizer rates for wheat in a spatially variable field. European Journal of Agronomy, 35(4) :215–222.

- Bates, T., Dresser, J., Eckstrom, R., Badr, G., Betts, T., and Taylor, J. (2018). Variable-rate mechanical crop adjustment for crop load balance in 'concord' vineyards. In 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany (IOT Tuscany). IEEE.
- Baume, O., Skøien, J., Heuvelink, G., Pebesma, E., and Melles, S. (2011). A geostatistical approach to data harmonization – application to radioactivity exposure data. International Journal of Applied Earth Observation and Geoinformation, 13(3) :409–419.
- Bégué, A., Todoroff, P., and Pater, J. (2008). Multi-time scale analysis of sugarcane within-field variability : improved crop diagnosis using satellite time series ? Precision Agriculture, 9(3) :161–171.
- Bellon-Maurel, V. and Huyghe, C. (2016). L'innovation technologique dans l'agriculture. Geoeconomie, (3) :159–180.
- Bendiab, A. T., Bathily, M., Vigreux, C., Escalier, R., Pradel, A., Kribich, R. K., and Bendoula, R. (2018). Chalcogenide rib waveguides for the characterization of spray deposits. Optical Materials, 86 :298–303.
- Blackmore, S., Godwin, R. J., and Fountas, S. (2003). The analysis of spatial and temporal trends in yield map data over six years. Biosystems Engineering, 84(4) :455–466.
- Blasch, G. and Taylor, J. A. (2018). Multi-temporal yield pattern analysis–adaption of pattern recognition to agronomic data. In Proceedings of the 14th International Conference on Precision Agriculture. International Society of Precision Agriculture.
- Blazy, J.-M., Dorel, M., Salmon, F., Ozier-Lafontaine, H., Wery, J., and Tixier, P. (2009). Model-based assessment of technological innovation in banana cropping systems contextualized by farm types in guadeloupe. European Journal of Agronomy, 31(1) :10–19.
- Bramley, R. (2005). Understanding variability in winegrape production systems 2. within vineyard variation in quality over several vintages. Australian Journal of Grape and Wine Research, 11(1) :33–42.
- Bramley, R., Ouzman, J., and Thornton, C. (2011). Selective harvesting is a feasible and profitable strategy even when grape and wine production is geared towards large fermentation volumes. Australian Journal of Grape and Wine Research, 17(3) :298–305.
- Brenning, A., Koszinski, S., and Sommer, M. (2008). Geostatistical homogenization of soil conductivity across field boundaries. Geoderma, 143(3-4) :254–260.
- Brisson, N., Ozier-Lafontaine, H., and Dorel, M. (1998). Effects of soil management and water regime on banana growth between planting and flowering - simulation using the stics model. Acta Horticulturae, (490) :229–238.
- Bronson, K. F., Booker, J. D., Bordovsky, J. P., Keeling, J. W., Wheeler, T. A., Boman, R. K., Parajulee, M. N., Segarra, E., and Nichols, R. L. (2006). Site-specific irrigation and nitrogen management for cotton production in the southern high plains. Agronomy Journal, 98(1) :212.
- Bugaud, C., Belleil, T., and Tixier, P. (2011). Modelling the effect of source/sink modifications on pulp dry weight of banana 'cavendish'. In IX International Symposium on Modelling in Fruit Research and Orchard Management 1068, pages 261–265.
- Bugaud, C. and Lassoudière, A. (2005). Variabilité de la durée de vie verte des bananes en conditions réelles de production. Fruits, 60(4) :227–236.
- Bullock, D. G., Bullock, D. S., Nafziger, E. D., Doerge, T. A., Paszkiewicz, S. R., Carter, P. R., and Peterson, T. A. (1998). Does variable rate seeding of corn pay ? Agronomy Journal, 90(6) :830.

- Castelan, F., Saraiva, L., Lange, F., de Lapeyre de Bellaire, L., Cordenunsi, B., and Chillet, M. (2012). Effects of black leaf streak disease and sigatoka disease on fruit quality and maturation process of bananas produced in the subtropical conditions of southern brazil. Crop Protection, 35 :127–131.
- Cerovic, Z. G., Masdoumier, G., Ghazlen, N. B., and Latouche, G. (2012). A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. Physiologia Plantarum, 146(3) :251–260.
- Chabrier, C. and Queneherve, P. (2003). Control of the burrowing nematode (*Radopholus similis* Cobb) on banana : impact of the banana field destruction method on the efficiency of the following fallow. Crop Protection, 22(1) :121–127.
- Champion, J. (1963). Le bananier. G.P. Maisonneuve & Larose, Paris.
- Clifford, P., Richardson, S., and Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. Biometrics, 45(1) :123.
- Colaço, A., Trevisan, R., Molin, J., Rosell-Polo, J., and Escolà, A. (2017). A method to obtain orange crop geometry information using a mobile terrestrial laser scanner and 3d modeling. Remote Sensing, 9(8) :763.
- Cook, S. E. and Bramley, R. G. V. (1998). Precision agriculture — opportunities, benefits and pitfalls of site-specific crop management in australia. Australian Journal of Experimental Agriculture, 38(7) :753.
- Coquil, B. and Bordes, J. P. (2005). Farmstar : an efficient decision support tool for near real time crop management from satellite images. In Proceedings of the 5th European Conference on Precision Agriculture, pages 873–880. Wageningen Academic Publishers.
- Cottin, R., Melin, P., and Ganry, J. (1987). Modélisation de la production bananière. influence de quelques paramètres en martinique. Fruits, 42(12) :691–701.
- Daughtry, C., Walthall, C., Kim, M., De Colstoun, E. B., and McMurtrey Iii, J. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote sensing of Environment, 74(2) :229–239.
- Debaeke, P., Rouet, P., and Justes, E. (2006). Relationship between the normalized SPAD index and the nitrogen nutrition index : Application to durum wheat. Journal of Plant Nutrition, 29(1) :75–92.
- Dens, K., Swennen, R., Romero, R., and Turner, D. (2008). Simulation of banana plant growth (*Musa* spp. *aaa*, cavendish sub-group) based on temperature and ontogeny. Acta Horticulturae, (803) :181–186.
- Diacono, M., Rubino, P., and Montemurro, F. (2013). Precision nitrogen management of wheat. a review. Agronomy for Sustainable Development, 33(1) :219–241.
- Diker, K., Heermann, D., and Brodahl, M. (2004). Frequency analysis of yield for delineating yield response zones. Precision Agriculture, 5(5) :435–444.
- Doré, T., Makowski, D., Malézieux, E., Munier-Jolain, N., Tchamitchian, M., and Tittone, P. (2011). Facing up to the paradigm of ecological intensification in agronomy : Revisiting methods, concepts and knowledge. European Journal of Agronomy, 34(4) :197–210.
- Dorel, M., Damour, G., Leclerc, N., Lakhia, S., Ricci, S., Vingadassalon, F., and Salmon, F. (2016). Parent plant vs sucker - how can competition for photoassimilate allocation and light acquisition be managed in new banana hybrids? Field Crops Research, 198 :70–79.
- Dupuis, E. M. and Whalen, J. K. (2007). Soil properties related to the spatial pattern of microbial biomass and respiration in agroecosystems. Canadian Journal of Soil Science, 87(5) :479–484.

- Eckstein, K. and Robinson, J. C. (1995). Physiological responses of banana (musa AAA ; cavendish subgroup) in the subtropics. i. influence of internal plant factors on gas exchange of banana leaves. Journal of Horticultural Science, 70(1) :147–156.
- Eckstein, K. and Robinson, J. C. (2000). Influence of mother plant on sucker growth, development and photosynthesis in banana musa aaa. Acta Horticulturae, (531) :161–168.
- Famoye, F., Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). Continuous univariate distributions, volume 1. Technometrics, 37(4) :466.
- FAOSTAT (2016). Bananas. production quantities.
- Fernandez Caldas, E. and Garcia, V. (1972). Etude sur la nutrition du bananier aux iles canaries i-effet de la nutrition azotée sur la circonférence du pseudo-tronc. Fruits, 27(7-8) :509–512.
- Freitas, A. S., Pozza, E. A., Alves, M. C., Coelho, G., Rocha, H. S., and Pozza, A. A. A. (2015). Spatial distribution of yellow sigatoka leaf spot correlated with soil fertility and plant nutrition. Precision Agriculture, 17(1) :93–107.
- Ganry, J. and Chillet, M. (2008). Methodology to forecast the harvest date of banana bunches. Fruits, 63(6) :371–373.
- Ganry, J. and Sioussaram, D. (1978). Recherche d'une méthode d'estimation de la date de récolte du bananier à partir de données climatiques dans les conditions des antilles. Fruits, 33(10) :669–680.
- Geoprospectors (2019). Top soil mapper. <http://www.geoprospectors.com/gb/products-services/agriculture/>. [Online; accessed January-2019].
- Gitelson, A. A., Vina, A., Ciganda, V., Rundquist, D. C., and Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters, 32(8).
- Goenaga, R. and Irizarry, H. (1995). Yield performance of banana irrigated with fractions of class a pan evaporation in a semiarid environment. Agronomy Journal, 87(2) :172.
- Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O., and Ravid, G. (2017). Applying machine learning on sensor data for irrigation recommendations : revealing the agronomist's tacit knowledge. Precision Agriculture, 19(3) :421–444.
- Gomez, K. A. and Datta, S. K. D. (1971). Border effects in rice experimental plots i. unplanted borders. Experimental Agriculture, 7(01) :87.
- Gonçalves, A. L. and Kernaghan, J. R. (2014). Banana production methods - a comparative study.
- Grisso, R. D., Alley, M. M., Thomason, W. E., Holshouser, D. L., and Roberson, G. T. (2011). Precision farming tools : variable-rate application.
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sensing of Environment, 81(2-3) :416–426.
- Hatfield, J. L. (2012). Spatial patterns of water and nitrogen response within corn production fields. In Agricultural science. Intech.
- Hatfield, J. L. and Kitchen, N. R. (2013). The role of precision agriculture in food production and security. In Precision Agriculture for Sustainability and Environmental Protection, pages 38–51. Routledge.

- Heslop-Harrison, J. S. and Schwarzacher, T. (2007). Domestication, genomics and the future for banana. Annals of Botany, 100(5) :1073–1084.
- Hill, S. B. and MacRae, R. J. (1996). Conceptual framework for the transition from conventional to sustainable agriculture. Journal of Sustainable Agriculture, 7(1) :81–87.
- Hotegni, V. N. F., Lommen, W. J. M., Agbossou, E. K., and Struik, P. C. (2014). Heterogeneity in pineapple fruit quality results from plant heterogeneity at flower induction. Frontiers in Plant Science, 5.
- Israeli, Y., Plaut, Z., and Schwartz, A. (1995). Effect of shade on banana morphology, growth and production. Scientia Horticulturae, 62(1-2) :45–56.
- Jullien, A., Chillet, M., and Malézieux, E. (2008). Pre-harvest growth and development, measured as accumulated degree days, determine the post-harvest green life of banana fruit. The Journal of Horticultural Science and Biotechnology, 83(4) :506–512.
- Kaspar, T. C., Colvin, T. S., Jaynes, D. B., Karlen, D. L., James, D. E., Meek, D. W., Pulido, D., and Butler, H. (2003). Relationship between six years of corn yields and terrain attributes. Precision Agriculture, 4(1) :87–101.
- Keskin, S., Khalilian, A., Han, Y., and Dodd, R. (2011). Variable-depth tillage based on geo-referenced soil compaction data in coastal plain soils. International Journal of Applied Science and Technology, 1(2).
- Koman, V. B., Lew, T. T. S., Wong, M. H., Kwak, S.-Y., Giraldo, J. P., and Strano, M. S. (2017). Persistent drought monitoring using a microfluidic-printed electro-mechanical sensor of stomata in planta. Lab on a Chip, 17(23) :4015–4024.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, 52(6) :119–139.
- Lamour, J., Naud, O., Lechaudel, M., and Tisseyre, B. (2017). Mapping properties of an asynchronous crop : the example of time interval between flowering and maturity of banana. Advances in Animal Biosciences, 8(02) :481–486.
- Lassoudière, A. (2007). Le bananier et sa culture. editions Quae.
- Lawless, J. F. (2002). Statistical Models and Methods for Lifetime Data. John Wiley & Sons, Inc.
- Le Moguédec, G. (2019). Elaboration d'un modèle pour la floraison du bananier.
- Lemaire, G., editor (1997). Diagnosis of the Nitrogen Status in Crops. Springer Berlin Heidelberg.
- Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., Naud, O., Crestey, T., Lablee, J.-L., and Tisseyre, B. (2018a). GeoFIS : An open source, decision-support tool for precision agriculture data. Agriculture, 8(6) :73.
- Leroux, C., Jones, H., pichon, L., Taylor, J., and Tisseyre, B. (submitted 2018b). Aspatial and spatial approaches to automatic harmonization of heterogeneous agronomical and environmental spatial datasets. Submitted to Precision Agriculture.
- Leroux, C., Jones, H., Taylor, J., Clenet, A., and Tisseyre, B. (2018c). A zone-based approach for processing and interpreting variability in multi-temporal yield data sets. Computers and Electronics in Agriculture, 148 :299–308.



- Machado, S., Bynum, E., Archer, T., Lascano, R., Wilson, L., Bordovsky, J., Segarra, E., Bronson, K., Nesmith, D., and Xu, W. (2000). Spatial and temporal variability of corn grain yield : Site-specific relationships of biotic and abiotic factors. Precision Agriculture, 2(4) :359–376.
- Machovina, B. L., Feeley, K. J., and Machovina, B. J. (2017). Uav remote sensing of spatial variation in banana production. Crop and Pasture Science, 67(12) :1281–1287.
- MacKenzie, S. J., Xiao, C. L., Mertely, J. C., Chandler, C. K., Martin, F. G., and Legard, D. E. (2003). Uniformity of strawberry yield and incidence of botrytis fruit rot in an annual production system. Plant Disease, 87(8) :991–998.
- Maestrini, B. and Basso, B. (2018). Drivers of within-field spatial and temporal variability of crop yield across the US midwest. Scientific Reports, 8(1).
- Makdessi, N. A., Ecartot, M., Roumet, P., and Rabatel, G. (2018). A spectral correction method for multi-scattering effects in close range hyperspectral imagery of vegetation scenes : application to nitrogen content assessment in wheat. Precision Agriculture.
- Makdessi, N. A., Jean, P.-A., Ecartot, M., Gorretta, N., Rabatel, G., and Roumet, P. (2017). How plant structure impacts the biochemical leaf traits assessment from in-field hyperspectral images : A simulation study based on light propagation modeling in 3d virtual wheat scenes. Field Crops Research, 205 :95–105.
- Martin-Prével, P. (1977). Echantillonnage du bananier pour l'analyse foliaire : Conséquences des différences de techniques. Fruits, 32(3) :151–166.
- Matheron, G. (1966). Présentation des variables régionalisées. Journal de la société française de statistique, 107 :263–275.
- McBratney, A. and Pringle, M. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture, 1(2) :125–152.
- McBratney, A. B. and Taylor, J. A. (2000). Pv or not pv. In Proceedings of the 5th international symposium on cool climate viticulture and œnology. Melbourne, Australia.
- McBratney, A. B. and Webster, R. (1986). Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. Journal of Soil Science, 37(4) :617–639.
- Mehra, L. K., Cowger, C., Gross, K., and Ojiambo, P. S. (2016). Predicting pre-planting risk of stagonospora nodorum blotch in winter wheat using machine learning models. Frontiers in Plant Science, 7.
- Mihail, J. D. (1989). Macrophomina phaseolina : Spatio-temporal dynamics of inoculum and of disease in a highly susceptible crop. Phytopathology, 79(8) :848.
- Miller, K. A., Luck, J. D., Heeren, D. M., Lo, T., Martin, D. L., and Barker, J. B. (2017). A geospatial variable rate irrigation control scenario evaluation methodology based on mining root zone available water capacity. Precision Agriculture, 19(4) :666–683.
- Morel, J., Jay, S., Féret, J.-B., Bakache, A., Bendoula, R., Carreel, F., and Gorretta, N. (2018). Exploring the potential of PROCOSINE and close-range hyperspectral imaging to study the effects of fungal diseases on leaf physiology. Scientific Reports, 8(1).
- Morphet, W. J. (2009). Simulation, Kriging, and visualization of circular-spatial data. PhD thesis, Utah State University.

- Multigner, L., Ndong, J. R., Giusti, A., Romana, M., Delacroix-Maillard, H., Cordier, S., Jégou, B., Thome, J. P., and Blanchet, P. (2010). Chlordecone exposure and risk of prostate cancer. Journal of Clinical Oncology, 28(21) :3457–3462.
- Öhlmér, B., Olson, K., and Brehmer, B. (1998). Understanding farmers' decision making processes and improving managerial assistance. Agricultural economics, 18(3) :273–290.
- Oliver, M. (2013a). An overview of precision agriculture. In Precision Agriculture for sustainability and environmental protection, pages 21–37. Routledge.
- Oliver, M. and Webster, R. (2014). A tutorial guide to geostatistics : Computing and modelling variograms and kriging. CATENA, 113 :56–69.
- Oliver, M. A. (2013b). Precision agriculture and geostatistics : How to manage agriculture more exactly. Significance, 10(2) :17–22.
- Oren, S., Ceylan, H., Schnable, P. S., and Dong, L. (2017). High-resolution patterning and transferring of graphene-based nanomaterials onto tape toward roll-to-roll production of tape-based wearable sensors. Advanced Materials Technologies, 2(12) :1700223.
- Parton, W. J. (1996). The CENTURY model. In Evaluation of Soil Organic Matter Models, pages 283–291. Springer Berlin Heidelberg.
- Paustian, M. and Theuvsen, L. (2016). Adoption of precision agriculture technologies by german crop farmers. Precision Agriculture, 18(5) :701–716.
- Pebesma, E. J. (2004). Multivariable geostatistics in s : the gstat package. Computers & Geosciences, 30(7) :683–691.
- Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. R News, 5(2) :9–13.
- Pervanchon, F. and Blouet, A. (2002). Lexique des qualificatifs de l'agriculture. Le Courrier de l'environnement de l'INRA, (45) :117–136.
- Ping, J. L. and Dobermann, A. (2005). Processing of yield map data. Precision Agriculture, 6(2) :193–212.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2018). nlme : Linear and Nonlinear Mixed Effects Models. R package version 3.1-137.
- Pinheiro, J. C. and Bates, D. M. (2000). Mixed-Effects Models in Sand S-PLUS. Springer New York.
- Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S., and Upchurch, D. R. (2003). Remote sensing for crop management. Photogrammetric Engineering & Remote Sensing, 69(6) :647–664.
- Plant, R. E. (2012). Spatial data analysis in ecology and agriculture using r.
- Ploetz, R. C. and Evans, E. A. (2015). The future of global banana production. In Horticultural Reviews : Volume 43, pages 311–352. John Wiley & Sons, Inc.
- Pringle, M., McBratney, A., Whelan, B., and Taylor, J. (2003). A preliminary approach to assessing the opportunity for site-specific crop management in a field, using yield monitor data. Agricultural Systems, 76(1) :273–292.
- Prost, L. and Jeuffroy, M.-H. (2007). Replacing the nitrogen nutrition index by the chlorophyll meter to assess wheat n status. Agronomy for Sustainable Development, 27(4) :321–330.

- Quebrajo, L., Pérez-Ruiz, M., Rodríguez-Lizana, A., and Agüera, J. (2015). An approach to precise nitrogen management using hand-held crop sensor measurements and winter wheat yield mapping in a mediterranean environment. *Sensors*, 15(3) :5504–5517.
- R Core Team (2018). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rabatel, G. and Labbé, S. (2016). Registration of visible and near infrared unmanned aerial vehicle images based on fourier-mellin transform. *Precision Agriculture*, 17(5) :564–587.
- Robert, P. (1993). Characterization of soil conditions at the field level for soil specific management. *Geoderma*, 60(1-4) :57–72.
- Rubin, D. B. (2006). The design versus the analysis of observational studies for causal effects : parallels with the design of randomized trials. *Statistics in Medicine*, 26(1) :20–36.
- Rudd, J. D. and Roberson, G. T. (2018). Using unmanned aircraft systems to develop variable rate prescription maps for cotton defoliant. In *2018 Detroit, Michigan July 29 - August 1, 2018*. American Society of Agricultural and Biological Engineers.
- Saddem, R., Naud, O., Cazenave, P., Dejean, K. G., and Crestani, D. (2017). Precision spraying : from map to sprayer control using model-checking. *Journal of Agricultural Informatics*, 8(3).
- Santesteban, L. G., Guillaume, S., Royo, J. B., and Tisseyre, B. (2012). Are precision agriculture tools and methods relevant at the whole-vineyard scale? *Precision Agriculture*, 14(1) :2–17.
- Sebillotte, M. (1969). Le « tour de plaine ». facteur de rentabilité dans l'entreprise agricole. *Oléagineux, Corps gras, Lipides*, 18(3) :105–112.
- Sequoia (2017). Parrot Announcement - Release of application notes.
- Shaffer, T. L. and Johnson, D. H. (2008). Ways of learning : Observational studies versus experiments. *Journal of Wildlife Management*, 72(1) :4–13.
- Shanahan, J. F., Doerge, T. A., Johnson, J. J., and Vigil, M. F. (2004). Feasibility of site-specific management of corn hybrids and plant densities in the great plains. *Precision Agriculture*, 5(3) :207–225.
- Shannon, K. (2018). Yield mapping - a useful tool in a challenging year too.
- Shaukat, S. S. and Siddiqui, I. A. (2004). Spatial pattern analysis of seeds of an arable soil seed bank and its relationship with above-ground vegetation in an arid region. *Journal of Arid Environments*, 57(3) :311–327.
- Siles, P., Bustamante, O., Valdivia, E., Burkhardt, J., and Staver, C. (2011). Photosynthetic performance of banana (gros michel, aaa) under a natural shade gradient. In *VII International Symposium on Banana : ISHS-ProMusa Symposium on Bananas and Plantains : Towards Sustainable Global Production 986*, pages 71–77.
- Singh, I., Srivastava, A. K., Chandna, P., and Gupta, R. K. (2006). Crop sensors for efficient nitrogen management in sugarcane : potential and constraints. *Sugar Tech*, 8(4) :299–302.
- Skøien, J. O., Baume, O. P., Pebesma, E. J., and Heuvelink, G. B. M. (2009). Identifying and removing heterogeneities between monitoring networks. *Environmetrics*, pages n/a–n/a.
- Smith, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *The Journal of Agricultural Science*, 28(01) :1.

- Stoorvogel, J. (1998). Banman : a decision support system for banana management. Information technology as a tool to assess land use options in space and time. Quantitative Approaches in Systems Analysis, 16 :13–22.
- Stoorvogel, J., Orlich, R., et al. (2000). An integrated system for precision agriculture in bananas. In Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA, 16-19 July, 2000, pages 1–14. American Society of Agronomy.
- Stoorvogel, J. J., Kooistra, L., and Bouma, J. (2015). Managing soil variability at different spatial scales as a basis for precision agriculture. Soil-Specific Farming : Precision Agriculture ; Lal, R., Stewart, BA, Eds, pages 37–72.
- Taylor, J. A., Praat, J.-P., and Bollen, A. F. (2007). Spatial variability of kiwifruit quality in orchards and its implications for sampling and mapping. HortScience, 42(2) :246–250.
- Taylor, J. A., Tisseyre, B., and Leroux, C. (2018). A simple index to determine if within-field spatial production variation exhibits potential management effects : application in vineyards using yield monitor data. Precision Agriculture.
- Tchuenteu, F. and Marie, P. (1998). Cartographie des sols de bananeraies du groupe spnp-sbm-php.
- Thomas, D., Turner, D., and Eamus, D. (1998). Independent effects of the environment on the leaf gas exchange of three banana (*musa sp.*) cultivars of different genomic constitution. Scientia Horticulturae, 75(1-2) :41–57.
- Thorp, K. R., Hunsaker, D. J., French, A. N., Bautista, E., and Bronson, K. F. (2015). Integrating geospatial data and cropping system simulation within a geographic information system to analyze spatial seed cotton yield, water use, and irrigation requirements. Precision Agriculture, 16(5) :532–557.
- Tilman, D. (1989). Ecological experimentation : Strengths and conceptual problems. In Long-Term Studies in Ecology, pages 136–157. Springer New York.
- Tilman, D. (1999). Global environmental impacts of agricultural expansion : The need for sustainable and efficient practices. Proceedings of the National Academy of Sciences, 96(11) :5995–6000.
- Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., and Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? Precision Agriculture, 19(5) :957–971.
- Tisseyre, B., Mazzoni, C., and Fonta, H. (2008). Within-field temporal stability of some parameters in viticulture : potential toward a site specific management. OENO One, 42(1) :27.
- Tixier, P., Malezieux, E., and Dorel, M. (2004). SIMBA-POP : a cohort population model for long-term simulation of banana crop harvest. Ecological Modelling, 180(2-3) :407–417.
- Tixier, P., Malézieux, E., Dorel, M., and Wery, J. (2008). SIMBA, a model for designing sustainable banana-based cropping systems. Agricultural Systems, 97(3) :139–150.
- Torres-Sánchez, J., López-Granados, F., and Peña, J. (2015). Mapping olive-tree geometric features from 3d models generated with an unmanned aerial vehicle. In Precision agriculture 2015, pages 321–328. Wageningen Academic Publishers.
- Turner, D. W., Fortescue, J. A., and Thomas, D. S. (2007). Environmental physiology of the bananas (*musa spp.*). Brazilian Journal of Plant Physiology, 19(4) :463–484.
- Urretavizcaya, I., Santesteban, L. G., Tisseyre, B., Guillaume, S., Miranda, C., and Royo, J. B. (2013). Oenological significance of vineyard management zones delineated using early grape sampling. Precision Agriculture, 15(1) :111–129.

- USDA (1999). Sustainable agriculture : Definitions and terms. <https://www.nal.usda.gov/afsic/sustainable-agriculture-definitions-and-terms-related-terms#term23>. [Online; accessed January-2019].
- Vallejos, R., Osorio, F., and Bevilacqua, M. (2018). Spatial Relationships Between Two Georeferenced Variables : with Applications in R. Springer, New York.
- Vargas, A., Araya, M., Guzmán, M., and Murillo, G. (2009). Effect of leaf pruning at flower emergence of banana plants (musa aaa) on fruit yield and black sigatoka (mycosphaerella fijiensis) disease. International Journal of Pest Management, 55(1) :19–25.
- Vasconez, J. P., Kantor, G. A., and Cheein, F. A. A. (2019). Human–robot interaction in agriculture : A survey and current challenges. Biosystems Engineering, 179 :35–48.
- Verdugo-Vásquez, N., Acevedo-Opazo, C., Valdés-Gómez, H., Araya-Alman, M., Ingram, B., de Cortázar-Atauri, I. G., and Tisseyre, B. (2015). Spatial variability of phenology in two irrigated grapevine cultivar growing under semi-arid conditions. Precision Agriculture, 17(2) :218–245.
- Vinatier, F., Tixier, P., Page, C. L., Duyck, P.-F., and Lescourret, F. (2009). COSMOS, a spatially explicit model to simulate the epidemiology of cosmopolites sordidus in banana fields. Ecological Modelling, 220(18) :2244–2254.
- von Gadow, K., Zhao, X. H., Tewari, V. P., Zhang, C. Y., Kumar, A., Rivas, J. J. C., and Kumar, R. (2016). Forest observational studies : an alternative to designed experiments. European Journal of Forest Research, 135(3) :417–431.
- Weber, O. B., dos Santos Garruti, D., Norões, N. P., and de Oliveira e Silva, S. (2017). Performance of banana genotypes with resistance to black leaf streak disease in northeastern brazil. Pesquisa Agropecuária Brasileira, 52(3) :161–169.
- Webster, R. and Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. Journal of Soil Science, 43(1) :177–192.
- Whelan, B. and McBratney, A. (2000). The “null hypothesis” of precision agriculture management. Precision Agriculture, 2(3) :265–279.
- Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression : a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2) :109–130.
- Woomer, P. L., Bekunda, M., and Bwamiiki, D. (1998). Modelling banana growth and soil organic matter dynamics with the CENTURY model. African Crop Science Journal, 6(2).
- Wyse-Pester, D. Y., Wiles, L. J., and Westra, P. (2002). Infestation and spatial dependence of weed seedling and mature weed populations in corn. Weed science, 50(1) :54–63.
- Zanella, V., Ortiz, B., Thorp, K., Morari, F., Mosca, G., and Hoogenboom, G. (2015). Combining crop sensing and simulation modeling to assess within-field corn nitrogen stress. In Precision agriculture 2015, pages 391–398. Wageningen Academic Publishers.
- Zwaenepoel, P. and Le Bars, J.-M. (1997). L’agriculture de précision. Ingénieries-EAT, (12) :p-67.

# Table des figures

1.1	Diversité des bananes vendues sur un marché en Inde. (Source : Heslop-Harrison and T. Schwarzacher, 2007)	8
1.2	Schéma d'un bananier et de son rejet. (Source : Champion 1963)	10
1.3	Schéma d'une inflorescence de bananier (Photo : Paul Esnault)	11
1.4	Inflorescence au stade « Tête de cheval » (Photo : Paul Esnault)	11
1.5	Schéma de développement d'un bananier dans une plantation	12
1.6	Exemple de courbe de floraison d'une parcelle de bananiers étudiée au cours de la thèse. La première floraison est relativement synchrone dans le temps, puis les floraisons suivantes sont de plus en plus étendues jusqu'à ce que les pics ne soient plus discernables	14
1.7	Représentation de la diversité des stades phénologiques au sein d'une parcelle	15
1.8	Représentation de l'augmentation de la diversité des stades phénologiques au cours du temps. Au moment de la plantation (en 2008) les stades phénologiques sont identiques, puis au fur et à mesure de l'âge de la parcelle et du fait de l'asynchronisme, la diversité des stades phénologiques augmente	15
1.9	Quantités de bananes produites par pays en 2016 en millions de tonnes (source : FAOSTAT 2016)	18
1.10	Quantités de bananes exportées par pays en 2016 en millions de tonnes (source : FAOSTAT 2016)	18
1.11	Quantités de bananes importées par pays en 2016 en millions de tonnes (source : FAOSTAT 2016)	19
2.1	Exemple de variabilité spatiale du rendement (exprimé en Bushels/Acre) dans un champ de maïs. Les facteurs qui expliquent les variations locales sont présentés par des flèches. (Source : Shannon 2018)	25
2.2	Représentation de la variable régionalisée $Z$ , de son espérance (surface plane) et de	28
2.3	Exemple de variogramme empirique (points) et de variogramme théorique (ligne). La variance pépité correspond à la variance non structurée spatialement, le palier partiel correspond à la variance structurée spatialement et le palier correspond à la variance totale. La portée correspond à la distance à partir de laquelle il n'y a plus d'auto-corrélation spatiale entre les données.	29
2.4	Exemples de variogrammes théoriques. Chaque variogramme caractérise des structures spatiales avec des traits structuraux différents. Par exemple, le variogramme pépité correspond à un bruit blanc (absence de structure spatiale), le variogramme gaussien décrit une structure avec une grande régularité spatiale qui apparaît « lisse », le variogramme exponentiel décrit une structure beaucoup plus irrégulière. La somme de variogrammes théoriques permet de modéliser des structures spatiales plus complexes. Par exemple, le variogramme théorique de la Figure 2.3 est la somme d'un variogramme pépité et d'un variogramme gaussien.	29



3.14	Variogramme expérimental et variogrammes théoriques de . Les points correspondent au variogramme expérimental, la ligne correspond à un modèle sphérique, la courbe en tirets correspond à un modèle gaussien et la ligne en pointillés correspond à un modèle exponentiel . . . . .	64
3.15	Interpolation des estimés suivant les 3 modèles théoriques de variogramme (sphérique, gaussien et exponentiel) . . . . .	65
3.16	Représentation de l'avancement (points verts) des bananiers sur un cercle. Pour simplifier le schéma les points sont représentés superposés alors qu'ils devraient tous être placés sur le périmètre du cercle. Le stade phénologique moyen de la parcelle est donné par l'angle et la variabilité est obtenue à partir de la distance la plus courte entre le centre de gravité des et le périmètre du cercle. . . . .	68
3.17	Représentation de la durée d'un cycle de floraison et de l'hétérogénéité des conditions de croissance estimés par la procédure d'ajustement du modèle pour chacune des 1188 CF. Les lignes verticales correspondent aux quantiles à 15 % et à 85 % de et les lignes horizontales correspondent aux quantiles à 15 % et à 85 % de . Les points rouges correspondent aux CF qui ont été choisies pour la suite des analyses. . . . .	72
3.18	Représentation des 4 CF choisies (points) et de leur ajustement par le modèle (ligne rouge).	73
3.19	Représentation de et de à 3 temps depuis la première floraison de la CF4 . . . . .	73
3.20	Représentation de l'évolution de la diversité des stades phénologiques au cours du temps pour chacune des 4 CF. L'enveloppe à 95 % de est aussi représentée. . . . .	74
3.21	Courbe de floraison de la parcelle. Les points rouges correspondent aux floraisons observées et les points noirs aux floraisons simulées. La ligne bleue représente l'évolution de la diversité des stades phénologiques au cours du temps évaluée à partir de l'indice .	76
3.22	Cosineogrammes invertis des stades phénologiques simulés dans la parcelle pour différents temps après la plantation de la parcelle. Les points correspondent aux cosineogrammes invertis et les lignes correspondent aux modèles théoriques ajustés . . . . .	76
3.23	Évolution du ratio d'auto-corrélation spatiale des stades phénologiques au cours du temps depuis la plantation de la parcelle . . . . .	76
4.1	Organisation of the fields of Farm A (a-subfield) and Farm B (b-subfield) . . . . .	84
4.2	Average total precipitation per month and average temperature per month. These averages were calculated from a weather station located on Farm A that has been operating since 2000 . . . . .	85
4.3	Maps of the altitude and slope in Farm A (a and c) and Farm B (b and d) . . . . .	85
4.4	Decomposition of the variability. Each arrow that connects a variable V1 to V2 can be read as V2 derives from V1 . . . . .	87
4.5	Irregular number (n) and location of observations for three different cohorts beginning on Mondays 04/18/2016, 05/16/2016 and 07/13/2016, respectively, in field 8 located in Farm A. Banana trees that have bloomed during these cohorts are represented by a dot .	89
4.6	Variation of by cohort in Farm A and Farm B. The vertical dotted lines represent the limit between each period P1 to P4. The black curve represents ( ). The grey area represents intervals encompassing 95 % of the values for each cohort . . . . .	90
4.7	Decomposition of the variogram of according to the formula . . . . .	92
4.8	Farm A Period 2. Block kriging maps of the variability ( ), of the variability induced by the cohort ( ), and of the variability induced by environmental factors ( ) . . . . .	94
4.9	Field 8, Farm A. Block kriging maps of the variability ( ), of the variability induced by the cohort ( ), and of the variability induced by environmental factors ( ). ** represents p-value <0.005 . . . . .	94



4.10	Maps of $\tau$ of farm A at four different periods P1 to P4. The colour scale unit is in days. Negative values correspond to locations where the maturity is reached faster than in other locations and, on the contrary, positive values correspond to locations where the maturity is reached slower . . . . .	97
4.11	Maps of $\tau$ of farm B at two different periods P3 and P4. The colour scale unit is in days. Negative values correspond to locations where the maturity is reached faster than in other locations and, on the contrary, positive values correspond to locations where the maturity is reached slower . . . . .	98
5.1	Example of banana leaf reflectance spectrum . . . . .	103
5.2	Example of band images of banana leaves obtained from Sequoia acquisition at 10 m above the ground . . . . .	103
5.3	Example of annotated RGB image giving the position and identification of the leaves . . .	103
5.4	Vigour map of the field where 35 leaves were sampled for the validation dataset . . . . .	103
5.5	Example of the RGB layer and of a polygon used to identify the 35 leaves of sampled in the field for the validation dataset . . . . .	104
5.6	Database structure . . . . .	104
5.7	Boxplot of the chlorophyll content of leaf 3 of flowering banana tree located in different fields. The fields used for the experiments are highlighted in bold . . . . .	105
5.8	Locations of the Dualex measurements and spectral measurement with the integrative sphere on the leaf. 9 locations are represented on the right part of the leaf. 9 other measurements were made at symmetrical positions on the left part of the leaf . . . . .	106
5.9	Banana plants at different phenological stages (a) Location of chlorophyll measurements on a leaf for the first (b) and second (c) experiments . . . . .	110
5.10	Chlorophyll-content variability in the 3-rank leaf for flowering banana plants in the 15 fields under study . . . . .	113
5.11	a) Mean effect of the leaf rank and phenological stage on the chlorophyll content b) Mean effect of the within-leaf observation's location on the chlorophyll content. The hull represents the confidence interval at 95 % . . . . .	114
5.12	Image acquisition of laid leaves. Left : image acquisition (Sequoia sensor shown top-left). Right : example of Sequoia image (RGB channel) with labelled leaves . . . . .	117
5.13	Multispectral image before (left) and after (right) co-registration. Images shown in false colors with $\tau$ , $\tau$ and $\tau$ . . . . .	118
5.14	$\tau$ estimation by PLS-R on learning and test sets . . . . .	119
5.15	Leaf $\tau$ estimation on selected leaves in canopy. Left : RGB and $\tau$ map detail. Right : $\tau$ - $\tau$ relationship . . . . .	120
5.16	Comparison between $\tau$ and vigor maps. Left and middle : circumference maps. Right : $\tau$ map after kriging . . . . .	121
5.17	Relation entre la teneur en chlorophylle mesurée par le Dualex et la teneur en azote exprimée en pourcentage massique (Figure de gauche, $\tau$ ) ou en masse surfacique (Figure de droite, $\tau$ ) . . . . .	122

# Liste des tableaux

3.1	Description des entrées du modèle . . . . .	45
3.2	Description des variables du modèle . . . . .	45
3.3	Description des paramètres du modèle . . . . .	45
3.4	Coefficients de la régression linéaire générale présentée équation 3.16. Avec la constante de référence de la population (Altitude 0 m, cultivar Williams, irrigué), l'effet de l'altitude, l'effet du cultivar Grande Naine, l'effet de l'irrigation, la variance des résidus, la portée de la structure spatiale des résidus, la variance pépité des résidus. ** p-value 0.005, * p-value 0.05 . . . . .	53
3.5	Paramètres ajustés pour la CF de la parcelle . . . . .	61
3.6	Paramètres de chacune des 4 CF choisies . . . . .	70
4.1	Number of observations used for the variance decomposition by period and farm . . . . .	86
4.2	Decomposition of the variance in two components, the cohort variance and the residual variance . . . . . represents the ratio of cohort variance to the total variance	90
4.3	Decomposition of the spatial structure of . . . . . and correspond respectively to the nugget variance and the partial-sill variance estimated from the variograms of each variable , and . . . . .	91
4.4	Correlation coefficient between maps and maps obtained at each period of each farm. . . . .	93
4.5	Correlation coefficient between maps at different periods. · p-value 0.05, * p-value 0.01, ** p-value 0.001 . . . . .	95
4.6	Correlation coefficient between or maps and the slope. · p-value 0.05, * p-value 0.01, ** p-value 0.001 . . . . .	95
4.7	Mean Block kriging variance of the estimation of and (units : day ) . . . . .	95
5.1	Number of samples by leaf rank and phenological stage. BF : Before Flowering, F : At flowering, AFn : n number of weeks After Flowering . . . . .	105
5.2	Number of samples by phenological stage and leaf rank in experiment 2 . . . . .	110
5.3	Parameter estimates of Model 2 through REML . . . . .	115



# Publications et participations à des conférences

**Lamour, J.**, Naud, O., Lechaudel, M., Le Moguédec, G., Taylor, J. and Tisseyre, B. (submitted 2018) Spatial analysis and mapping of banana crop properties : issues of the asynchronicity of the banana production and proposition of a statistical method to take it into account. *Precision Agriculture*.

**Lamour, J.**, Le Moguédec, G., Naud, O., Lechaudel, M. and Tisseyre, B. (submitted 2019) Model-assisted on-farm experiments : the case of the flowering cycle duration in banana plantations. *Field Crops Research*.

Leroux, C., Jones, H., Pichon, L., Guillaume, S., **Lamour, J.**, Taylor, J., ... and Tisseyre, B. (2018). GeoFIS : An Open Source, Decision-Support Tool for Precision Agriculture Data. *Agriculture*, 8(6), 73. doi :10.3390/agriculture8060073

**Lamour, J.**, Naud, O., Lechaudel, M. and Tisseyre, B. (2017). Mapping properties of an asynchronous crop : the example of time interval between flowering and maturity of banana. Paper presented at the 11th European Conference on Precision Agriculture (ECPA 2017), John McIntyre Centre, Edinburgh, UK, July 16–20 2017, *Advances in Animal Biosciences*, 8(02), 481–486. doi :10.1017/s2040470017000449

**Lamour, J.**, Leroux, C., Naud, O., Lechaudel, M., Le Moguédec G. and Tisseyre, B. (accepted 2019). Disentangling the sources of chlorophyll-content variability in banana crops using an optical chlorophyll meter. 12th European Conference on Precision Agriculture (ECPA 2019), SupAgro, Montpellier, France, July 8–11 2019

Rabatel, G., **Lamour, J.**, Moura, D. and Naud, O. (accepted 2019). A multispectral processing chain for chlorophyll content assessment in banana fields by UAV imagery. 12th European Conference on Precision Agriculture (ECPA 2019), SupAgro, Montpellier, France, July 8–11 2019



# Remerciements

Ce travail de thèse a été réalisé avec l'aide et l'encadrement de nombreuses personnes.

Je remercie la Compagnie Fruitière de m'avoir fait confiance en alternance puis en thèse et particulièrement Alain Normand et Gilbert Bonnefont qui m'ont donné beaucoup de responsabilités et d'opportunités dès mon arrivée dans l'entreprise en 2014.

Je remercie l'ensemble des personnes qui m'ont encadré dans la thèse. Merci Bruno Tisseyre d'avoir accepté de diriger la thèse et de m'avoir suivi depuis le début de mes études à Supagro. Merci Olivier Naud pour ton suivi quotidien, les nombreuses discussions et ta grande ouverture d'esprit ! Merci Mathieu Léchaudel pour tes conseils et ton intérêt tout au long de la thèse. Je garde un très bon souvenir de mon « stage » de formation en Guadeloupe et j'aurais aimé pouvoir passer plus de temps sur place. Un grand merci Gilles Le Moguédec de m'avoir apporté ton aide très précieuse. Tu m'as appris énormément de choses depuis mon premier stage au CIRAD. Un grand nombre des travaux présentés dans la thèse n'auraient pas été possibles sans ton implication. Merci d'avoir été autant disponible pour travailler avec moi, sur ton grand tableau blanc et sur R ! Merci pour tes grandes qualités de patience, d'écoute et ta pédagogie. Merci aussi Ryad Bendoula pour les cafés coachings presque quotidiens et toujours très utiles. J'espère que nous pourrions travailler un jour ensemble.

Les données présentées ont été recueillies avec l'aide de nombreuses personnes que je remercie. Je tiens à remercier particulièrement Paul Esnault et Philippine Sotteau qui ont effectué leur stage de master sur des sujets connexes à cette thèse. Je remercie aussi Paul d'avoir continué en VIE et d'avoir participé à de nombreuses campagnes de mesures. Je remercie l'ensemble des personnes de la Plantation du Haut Penja (sans oublier toutes les personnes du club) qui m'ont bien accueilli et qui ont facilités les expérimentations. J'ai de très bon souvenirs de mes passages au Cameroun, j'ai adoré mes missions sur le terrain, parce que les paysages sont si beaux, parce que l'atmosphère de la plantation est si dépaysant mais aussi grâce à vous et ce que vous avez partagés avec moi. Je remercie aussi l'équipe agronomie en entier et notamment Aurélien Pugeaux, Jacques Simo et Emmanuel Andjengo. Je pense aussi à Richard Congo, Paul Dontsop et Honoré qui m'ont guidé dans le labyrinthe des parcelles qu'ils connaissent parfaitement.

Je remercie Gilles Rabatel et Daniel Moura de m'avoir accompagné au Cameroun pour la mission drone. Merci Gilles pour avoir investi autant de temps sur l'analyse des données. Merci Daniel d'avoir pris à cœur cette mission, d'avoir accepté de rester plus longtemps pour finir l'acquisition des données, et merci de m'avoir montré (un peu) de ton savoir-faire et de tes connaissances. Le travail que tu as réalisé pour faciliter les acquisitions avec des interfaces et garder le drone en état de fonctionnement dans des conditions tropicales ne se mesure pas en lisant les articles et le mémoire. Je sais cependant pertinemment que la qualité des données et des résultats est due en grande partie à ton travail et à ta rigueur. Merci !

Je te remercie aussi Corentin Leroux pour ton aide et ton optimisme incroyable. Je suis très content que l'on ait réussi à travailler ensemble sur l'article chlorophylle !

Enfin, merci aux personnes d'ITAP et particulièrement aux personnes du bâtiment MINEA pour leur accueil pendant la durée de la thèse. Je ne me lance pas dans l'exercice périlleux de nommer toutes ces personnes, pour ne pas prendre le risque d'en oublier. Je ne ferais qu'une exception pour mes collègues (et amis !) de bureau : Anis, Maxime, et Adel (Monsieur Adel!).



# Résumé

La culture bananière dédiée à l'export est industrielle et implantée sur de grandes surfaces monovariétales. Cette culture est intensive en intrants, en main d'œuvre et est soumise à des contraintes environnementales et sociétales qui imposent une amélioration des pratiques agricoles. L'Agriculture de Précision (AP) est une approche technique qui est apparue sur les grandes cultures pour optimiser leur rendement, la qualité des récoltes et pour diminuer leur impact environnemental. Cette approche s'appuie sur les données spatialisées des exploitations pour réaliser un état des lieux de la production. Son objectif est de mettre en évidence les conditions variables de croissance et d'identifier celles qui sont contrôlables afin d'optimiser les procédés de production en fonction du potentiel local dans l'exploitation. La culture bananière présente des particularités par rapport aux grandes cultures qui n'ont pas été étudiées et qui remettent en cause l'utilisation directe des méthodes d'analyse des données des exploitants. Notamment, les bananiers sont asynchrones, leur cycle de développement n'est pas saisonnier et n'est pas synchronisé par le mode de culture. En conséquence, les parcelles sont constituées de bananiers à différents stades phénologiques. Ainsi, contrairement aux grandes cultures qui sont, elles, synchrones, les observations que l'on peut faire sur une bananeraie à un moment donné ne dépendent pas seulement des conditions de croissance mais aussi du stade phénologique variable des plants.

L'objectif de cette thèse est de proposer de nouvelles méthodes pour valoriser les données produites en bananeraies afin de faciliter le diagnostic agronomique dans une démarche d'AP. Les données sur lesquelles nous avons travaillé sont issues d'observations classiquement réalisées dans les bananeraies pour gérer les récoltes. Nous nous sommes aussi intéressés à la télédétection et nous avons proposé des méthodes d'analyse qui permettent d'étudier la variabilité spatiale en diminuant le biais lié à l'asynchronisme.

Le premier travail de la thèse a été de proposer des méthodes de caractérisation de l'asynchronisme des parcelles. Pour cela nous avons définis plusieurs indicateurs : la durée du cycle de développement des bananiers ; l'hétérogénéité des conditions de croissance ; le stade phénologique moyen d'une parcelle et enfin la variabilité des stades présents dans une parcelle. Ces méthodes ont été appliquées sur les données de floraisons d'une plantation industrielle au Cameroun. Nous avons ainsi mis en évidence un effet de l'environnement et des pratiques des producteurs sur l'asynchronisme.

Un deuxième pan du travail de thèse a été de proposer un modèle qui permet d'identifier la variabilité liée à l'environnement à partir de données de production. Le but était de générer des cartes interprétables agronomiquement sans biais lié à l'asynchronisme. Le modèle que nous avons proposé a été appliqué sur un indicateur particulier : le temps entre la floraison des plants et le moment de leur maturité. On suppose que cette méthode est suffisamment générale pour être appliquée à d'autres propriétés comme le poids des régimes à la récolte.

Enfin, le dernier sujet étudié a été l'intérêt de la télédétection dans les bananeraies malgré la diversité des stades phénologiques. Nous avons quantifié l'importance de différentes sources de variabilité sur la teneur en chlorophylle des feuilles de bananier mesurée à partir d'un capteur piéton et nous avons construit un indice qui permet de prédire cette teneur par drone. Des cartes de variabilités de la chlorophylle intra-parcellaire ont été produites par ce moyen mais l'interprétation agronomique doit être faite avec précaution car les stades phénologiques ne sont pas connus et leurs effets ne peuvent être corrigés. Les conditions permettant d'interpréter ces cartes sont discutées.





# Abstract

Exported bananas are produced on industrial plantations covering large areas and relying most of the times on clones of the Cavendish cultivars. This crop is input and labour intensive and is subject to environmental and societal constraints that impose an improvement in agricultural practices. Precision Agriculture (PA) is a methodological approach that has emerged on arable crops to optimize their yield, quality and reduce their environmental impact. This approach uses geo-referenced producer's data and studies their spatial variability to increase the farming performance by adjusting the management by area. Its objective is to highlight the variable growth conditions and identify those that are manageable in order to optimize production processes according to the specific potential of the sites on the farm.

Unlike arable crops, bananas exhibit unique characteristics that must be taken into account when studying production and yields. In particular, banana plants are asynchronous, their development cycle is not seasonal and is not synchronized by the cropping system. As a result, the fields are made up of banana plants at different phenological stages. Thus, unlike arable crops which are synchronous, the observations that can be made on a banana plantation at a given time depend not only on growing conditions but also on the variable phenological stages of the plants.

The objective of this doctoral thesis was to propose new methods for using data produced in banana plantations in order to facilitate agronomic diagnosis in a PA approach. The data used are mainly those recorded to manage the harvests. These are observations classically recorded in banana plantations. We also studied remote sensing data and proposed analytical methods to study spatial variability by reducing the bias due to the asynchronism of banana plants.

The first part of the thesis work consisted in proposing methods to characterize the asynchronism of the fields. For this purpose, we defined several indicators : the average duration of the banana plant development cycle ; the heterogeneity of growing conditions ; and finally the average phenological stage of a field and the within field variability of the stages. These methods were applied to flowering data from an industrial plantation in Cameroon. An effect of the environment and producers' practices on asynchronism was highlighted.

A second part of the thesis consisted in proposing a model that identifies variability related to the environment and producers' choices using production data. The purpose of this method was to generate maps that could be interpreted agronomically, without the asynchronism bias. The model we proposed was applied to a particular property : the time between flowering and commercial maturity. It is assumed that this method is general enough to be applied to other agronomic properties such as the weight of bunches at harvest.

Finally, the last subject we studied was the relevancy of remote sensing in banana plantations despite the diversity of phenological stages. We assessed the importance of different sources of variability on the chlorophyll content of banana leaves measured from a pedestrian sensor and constructed an index to predict this content by drone. Maps of within field variability of chlorophyll content have been produced by this method, but agronomic interpretation must be done with caution because the phenological stages are not known and their effects cannot be corrected. The conditions for interpreting these maps are discussed.