THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Sciences Agronomiques

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau
Portée par

Unité de recherche HortSys

Estimation spatialisée des rendements d'une culture pérenne en Afrique de l'Ouest : le cas du manguier au Sénégal

Présentée par Julien SARRON
Le 09 décembre 2019

Sous la direction de Éric MALÉZIEUX
et Émile FAYE

Devant le jury composé de

Véronique BELLON-MAUREL, Ingénieure Général, IRSTEA	Présidente
Delphine BURGER-LEENHARDT, Directrice de Recherche, INRA	Rapportrice
Marie-Hélène JEUFFROY, Directrice de Recherche, INRA	Rapportrice
Pierre-Éric LAURI, Ingénieur de Recherche, INRA	Examinateur
Thomas NESME, Professeur, Bordeaux Sciences Agro	Examinateur
Éric MALÉZIEUX, Chercheur, CIRAD	Directeur de thèse
Émile FAYE, Chercheur, CIRAD	Invité

UNIVERSITÉ DE MONTPELLIER

Montpellier SupAgro

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Sciences Agronomiques

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau
Portée par

Unité de recherche HortSys

Estimation spatialisée des rendements d'une culture pérenne en Afrique de l'Ouest : le cas du manguier au Sénégal

Présentée par Julien SARRON
Le 09 décembre 2019

Sous la direction de Éric MALÉZIEUX
et Émile FAYE

Devant le jury composé de

Véronique BELLON-MAUREL, Ingénieure Général, IRSTEA	Présidente
Delphine BURGER-LEENHARDT, Directrice de Recherche, INRA	Rapportrice
Marie-Hélène JEUFFROY, Directrice de Recherche, INRA	Rapportrice
Pierre-Éric LAURI, Ingénieur de Recherche, INRA	Examinateur
Thomas NESME, Professeur, Bordeaux Sciences Agro	Examinateur
Éric MALÉZIEUX, Chercheur, CIRAD	Directeur de thèse
Émile FAYE, Chercheur, CIRAD	Invité

UNIVERSITÉ DE MONTPELLIER

Montpellier SupAgro

A formation égale, la vérité triomphe. Formez-vous, armez-vous de sciences jusqu'aux dents (...) et arrachez votre patrimoine culturel.

- Cheikh Anta Diop
Conférence à la jeunesse, Niger 1984

à ma mère, Béatrice

Remerciements

Voilà. Voilà trois années d'écoulées. Trois années riches et intenses pour parvenir à écrire ces pages. La thèse est un parcours qui marque et il est bien difficile de résumer ce temps écoulé en quelques mots.

Ce travail n'aurait pas pu avoir lieu sans la confiance, la bienveillance et l'enthousiasme de mon directeur de thèse, Éric Malézieux, et de mon encadrant Émile Faye. J'ai vraiment apprécié travailler avec vous deux. Merci Éric pour tes connaissances et les nombreuses interactions scientifiques, très enrichissantes, qui m'ont permis d'approfondir mes compétences. Merci Émile pour les innombrables conseils rédactionnels et techniques, pour les échanges quotidiens (scientifiques ou non), et pour les nombreux dépannages durant mon séjour au Sénégal.

Toute la partie terrain de cette thèse (et c'est un gros morceau) n'aurait pas pu avoir lieu sans Cheikh Bassirou Sané qui m'a apporté à la fois ses connaissances du manguier et de la flore du Sénégal et qui a permis toutes les interactions avec les producteurs. Merci et bonne continuation à toi. Je remercie également, tous les producteurs de mangues qui nous ont donné accès à leur terrain pour travailler. Sans eux, ce travail ne serait rien.

Je remercie Philippe Borianne qui a permis d'introduire les réseaux de neurones dans cette thèse et pour sa grande contribution dans la construction des outils d'analyse d'images. Merci aussi pour les débats enrichissants toujours annotés d'une pointe d'humour. Merci à Thibaut Nordey qui m'a aidé à la mise en place des premiers algorithmes d'analyse d'images et pour ses différents retours sur mes travaux.

Je tiens à remercier aussi Frédéric Normand et Pierre-Éric Lauri pour leurs nombreux conseils et précieuses connaissances sur le manguier et l'écophysiologie. Merci à Louise Leroux, Camille Lelong-Richaud, Frédéric Borne et Paterne Diatta d'avoir accepté de participer à mon comité de thèse et pour les recommandations qu'ils ont apportés à ce travail. Merci à Jean-Yves Rey et Henri Vannière pour leurs connaissances sur l'agronomie des fruitiers au Sénégal et les nombreux échanges sur le terrain.

Je remercie Marie-Hélène Jeuffroy et Delphine Burger-Leenhardt d'avoir acceptées d'être rapportrices de cette thèse, et Véronique Bellon-Maurel, Thomas Nesme et Pierre-Éric Lauri d'examiner ce travail.

Une large part de cette thèse s'est déroulée au département horticole de l'ISRA de Cambérène au Sénégal. Je remercie tous les collègues du Centre pour le Développement de l'Horticulture (CDH) pour leur accueil. Merci à Jeanne Diatta et Lamine Diamé pour les échanges que nous avons pu partager. Je remercie les collègues et le personnel du CIRAD au Sénégal, plus particulièrement Sylvie Lewicki, pour les moments de partage et de convivialité. Je garde un très bon souvenir de nos échanges divers et variés et des différentes occasions de retrouvaille que j'ai pu avoir au sein de cette grande famille. Merci aussi à Louise et Florian qui ont grandement facilité mon aménagement à Montpellier.

Merci aux collègues de l'unité HortSys pour leur accueil et les moments agréables passés au sein de l'équipe. Merci à Pauline qui partage les mêmes efforts de fin de thèse (courage !). Merci également à Alioune pour nos discussions sur le doctorat, la mangue et le Sénégal.

Merci aux différents stagiaires qui ont participé à ce travail : Mounir, Guilhem, Lilia et Nicéphore.

Ces deux ans et demi au Sénégal représentent également une partie de ma vie que je ne suis pas prêt d'oublier. J'y ai vécu des moments très riches sur le plan professionnel comme personnel. Pour ce dernier

pan je tenais à remercier mes amis de Dakar : Matar, Elzo, Justine, Boya, et j'en oublie. Je pense également à Baptiste et Mélissa et à la team de Saint Louis. Tous, m'ont apporté du soutien moral pour la thèse ainsi que de nombreux moments de partage. Je ne vous oublie pas et j'espère vous revoir très vite ! Dédicace à Matar et à sa famille pour leur accueil chaleureux et les très bons repas de fêtes.

Je remercie mes amis de la France qui m'ont accueilli à Paris ou à Montpellier : Marc, Mélodie et Carole. Je remercie également ceux qui sont venu jusqu'à Dakar pour me rendre visite : Chacha, Élise, Matthieu (qui revient même), Beccon, Mathieu, Cyprien, Popol et encore Carole, Mélodie et Marc (même si c'est pour un jour, ça compte quand même). Merci à mes colocs, Pierre, Beccon et Chacha pour leur soutien quotidien et la relecture de ce mémoire. Un grand merci, du fond du cœur, à Armel, compagnon de route et soutien permanent qui m'a donné la motivation quotidienne pour ce travail.

Enfin, je tenais à remercier tous mes amis que je n'ai pas cités et qui m'ont soutenu durant ces trois ans ainsi que ma famille, mes tantes, mon parrain et ma marraine. Merci à mes grands-parents, mon père et mes deux frères. Leur visite au Sénégal m'a fait énormément plaisir. Ce travail achève un long parcours auquel ils ont participé pour une large part et, pour mon père, investi sans relâche depuis de si longues années. Je leur dédie ce mémoire.

Productions scientifiques et supports de communication de la thèse

Articles scientifiques à comité de lecture

Sarron J.; Malézieux É.; Sané C.A.B.; Faye É. **Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV**. Remote Sens 10:1900. https://doi.org/10.3390/rs10121900

Sarron J.; Faye É.; Nordey T.; Diatta J.; Malézieux É. **Production gaps at the tree scale: concept and application to mango in West Africa.** *Agron. Sustain. Dev.* [*soumis*].

Sarron J.; Faye É.; Malézieux É. **Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa.** [*à soumettre*].

Sarron J.; Faye É.; Malézieux É. **Understanding drivers of mango yields in Sub-Saharan Africa: insights from regional agronomic diagnosis.** [*à soumettre*].

Communications orales avec actes

Sarron J.; Sané C.A.B.; Borianne P.; Malézieux É.; Nordey T.; Normand F.; Diatta P.; Faye É. **Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions?** XXX. International Horticultural Congress, 2018, Istanbul, Turkey [*Acta Hortic., en press.*].

Borianne P.; Sarron J.; Borne F.; Faye É. **Deep Mangoes: from fruit detection to cultivar identification in color images of mango trees**. DISP'19, 2019, Oxford, UK, , arXiv:1909.10939.

Communications orales sans actes

Sarron J.; Malézieux É.; Sané C.A.B.; Faye É. **Cartographie de la production des vergers de mangue à partir des paramètres de structure des arbres et de l'occupation des sols évalués par drone.** AgriNumA 2019 : Symposium "Agriculture Numérique en Afrique", 2019, Dakar, Sénégal.

Sarron J.; Sané C.A.B.; Diatta P.; Diatta J.; Malézieux É.; Faye É. **Plant diversity and productivity in Senegalese mango orchards: Evidences from UAV photogrammetry.** 4th World Congress on Agroforestry, 2019, Montpellier France.

Posters scientifiques

Sarron J.; Sané C.A.B.; Borianne P.; Diatta P.; Faye É. **PixYield – L'analyse d'images photographiques comme outil pour l'estimation de rendement : le cas du manguier au Sénégal**. Journée du CIRAD, 2018, Montpellier, France.

Sarron J.; Sané C.A.B.; Diatta J.; Faye É.; Malézieux É. **Estimating and modelling productivity of mango trees from endogenous factors. A methodology set up and validated for African Orchards.** XV European Society for Agronomy Congress, 2018, Genève, Suisse.

Faye É.; Sarron J.; Malézieux É. **Integrating innovative digital tools to estimate yields of perennial crops at different spatial scales in the tropics.** 5th Global Science Conference on Climate-Smart Agriculture, 2019, Bali, Indonesia

Communication de presse

« Vers une application pour prédire la production de mangues au Sahel », CIRAD.fr, février 2019

« Un outil numérique pour renforcer la filière mangue », Scidev.net, mars 2019

« Agriculture numérique : le Sénégal montre l'exemple » AFP et le Monde, mai 2019

Vidéo de communication de l'institut de convergence #DigitAg

Cette thèse a été labélisée et a bénéficié d'une bourse de l'Institut de Convergences Agriculture Numérique #DigitAg (ANR-16-CONV-0004).

Cette thèse a également bénéficié d'une bourse du comité d'organisation du 4th World Congress on Agroforestry pour la participation en tant que 'keynote speaker' du symposium 'Agroforestry mapping'.

Une partie de cette thèse a été réalisée au sein du Centre pour le Développement de l'Horticulture (CDH) de l'Institut Sénégalais de Recherche Agricole (ISRA).







Table des matières

Introduction générale 13

Contexte et Problématiques 15

I. L'Afrique de l'Ouest : une région hétérogène 17

I.1. Définition géographique, milieu physique et climatique 17

I.2. Contexte socio-économique 21

I.3. Productions agricoles 22

I.4. Sécurité alimentaire et changements globaux 23

II. Les arbres fruitiers en Afrique de l'Ouest et leurs problématiques 27

II.1. Avantages de l'arbre dans les systèmes de culture 27

II.2. Systèmes à base de fruitiers en Afrique de l'Ouest 29

II.3. Problématiques rencontrées par les filières fruitières en Afrique de l'Ouest 31

III. Cas d'étude : le manguier au Sénégal 33

III.1. Description de l'espèce et production 33

III.2. Architecture et botanique du manguier 35

III.3. Cycle phénologique et asynchronisme 37

III.4. La filière mangue au Sénégal 39

Objectifs de la thèse 42

État de l'art 45

I. Facteurs déterminant la production de l'arbre 47

I.1. Architecture de l'arbre et interception lumineuse 47

I.2. Phase reproductive et établissement du fruit 49

I.3. Les pratiques culturales 51

II. De la production de l'arbre au rendement du verger 52

II.1. Structure des vergers 52

II.2. Hétérogénéité spatiale du rendement 55

II.3. Intégration des différentes échelles 56

III. Outils et méthodes pour la caractérisation de l'arbre et du verger 57

III.1. La détection des fruits 57

III.2. Structure des arbres 60

III.3. Description des vergers par télédétection 61

Matériels et méthodes 63

I. Cadre conceptuel et déroulement de la thèse 65

II. Zone d'étude et plan d'échantillonnage 67

II.1. Les Niayes 67

II.2. Typologie des vergers de mangues de la zone des Niayes 70

II.3. Echantillonnage 72

III. Acquisition des données 74

III.1. Suivi photographique pour estimer le nombre de fruits 74

III.2.Calibration des modèles d'estimation de la production 74

III.3. Structure des arbres et carte d'occupation des sols 74
III.4. Indice de charge pour l'estimation du rendement du verger 76
III.5. Données pédoclimatiques 77
III.6. Données d'enquêtes 77
Chapitre 1 : Estimation et étude de la production du manguier 79
I. Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions? 82
II. Deep Mangoes: from fruit detection to cultivar identification in colour images of mango trees 92
III. Production gaps at the tree scale: concept and application to mango in West Africa 107
Chapitre 2 : Extrapolation et analyse du rendement à l'échelle du verger 127
I. Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV 130
II. Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa 155
III. Understanding drivers of mango yields in Sub-Saharan Africa: insights from regional agronomic diagnosis 178
Discussion 197
I. Outils pour la caractérisation de l'arbre et du verger 199
I.1. Estimation de la production de l'arbre par analyse d'image 199
I.2. Caractérisation de vergers complexes 205
II. Méthodes d'échantillonnage et estimation du rendement du verger 209
II.1. Utiliser l'imagerie drone pour l'estimation des rendements 209
II.2. Combiner les variabilités intra-verger et l'estimation de la production 212
II.3. Extrapolation au bassin de production 215
III. Facteurs de variabilité du rendement des cultures pérennes à différentes échelles 217
III.1. Facteurs de variabilité du rendement 217
III.2. La variable de charge en fruit ('*fruit load*') 224
IV. Perspectives de la thèse 225
IV.1 De nouvelles méthodologies pour l'étude des cultures fruitières 225
IV.2. Développement d'un outil de gestion et de partage de données pour informer les filières 226
Conclusion générale 229
Références 231
Annexes 243

Introduction générale

L'Afrique de l'Ouest fait face à de nombreux changements à la fois démographiques, environnementaux et socio-économiques. L'agriculture doit répondre à la fois à ces changements tout en garantissant la sécurité alimentaire et en s'adaptant au changement climatique (van Ittersum et al. 2016). Dans les régions tropicales, comme partout ailleurs, les arbres fruitiers rendent de nombreux services écosystémiques que ce soit pour la sécurité alimentaire (nutrition et santé), le revenu des producteurs et pour le milieu (biodiversité, fertilité du sol, régulation du microclimat, etc.). La majeure partie des fruits en Afrique de l'Ouest est produite dans une diversité de petits vergers, 80% des producteurs possèdent moins de 2 hectares (Lowder et al. 2016), dont la structure (densité de plantation, arrangement spatial des arbre, composition spécifique, etc.) est parfois complexe. Dans les vergers on observe de faibles rendements en fruits et, bien souvent, des pratiques culturales inadaptées (de Bon et al. 2014). Cependant, les informations sur les systèmes de culture à base d'arbres fruitiers sont bien souvent imprécises ou manquantes en Afrique de l'Ouest, que ce soit de l'échelle de l'arbre à l'échelle du bassin de production. Ces données d'observation concernent les rendements, mais également les surfaces de production, les pratiques culturales et la structure des vergers (Carletto et al. 2015). Ce manque de données est notamment expliqué par la diversité et la complexité des systèmes de cultures. En effet, à l'heure actuelle de nombreux outils sont disponibles pour caractériser les monocultures (imagerie drone, analyses d'images, méthodes statistiques) mais se trouvent inadaptés pour les vergers complexes. Ce manque d'information sur les cultures fruitières en Afrique de l'Ouest ne permet pas d'évaluer convenablement les systèmes de culture et de concevoir des itinéraires techniques adaptés (Doré et al. 2008).

Le but de cette thèse est de parvenir à mettre en place des outils fiables et adaptés pour l'estimation et la spatialisation des rendements des cultures pérennes en Afrique de l'Ouest. Pour ce faire nous nous intéressons à un cas d'étude : le manguier dans le bassin de production des Niayes au Sénégal. Cette zone d'étude est une des principales zones du pays pour la production de mangues et la première pour la mangue d'exportation. Les systèmes de culture du manguier y sont également variés entre les vergers extensifs (ou de 'cueillette'), les vergers diversifiés, parfois similaires à des agroforêts, et les monocultures intensives. En caractérisant les différents systèmes aux échelles de l'arbre et du verger, des méthodes d'analyses sont mises en place pour déterminer les effets des facteurs environnementaux, des pratiques et des structures de l'arbre et du verger sur les rendements.

Dans ce mémoire, nous détaillerons, dans un premier temps, le contexte et les problématiques de la thèse en décrivant la place des arbres fruitiers dans les systèmes agricoles d'Afrique de l'Ouest ainsi que le cas d'étude de la thèse : le manguier et sa production au Sénégal. Dans une seconde partie, nous détaillerons l'état de l'art sur les facteurs impactant l'élaboration de la production de l'arbre et du rendement du verger ainsi que sur les outils disponibles pour l'estimation et la spatialisation des rendements. Dans les

parties suivantes, la zone d'étude, le plan d'échantillonnage et les outils employés seront décrit. Ensuite, deux chapitres décriront, sous formes d'articles scientifiques, les résultats obtenus au cours de la thèse. Enfin, la dernière partie reviendra sur les principaux points de discussion de la thèse.

CONTEXTE ET PROBLEMATIQUES

Les systèmes à base de pérennes en Afrique de l'Ouest

I. L'Afrique de l'Ouest : une région hétérogène

I.1. Définition géographique, milieu physique et climatique

a. Limites géographiques et hydrographie

On définit plus communément l'Afrique de l'Ouest comme la région couvrant la partie occidentale de l'Afrique au sud du Sahara (Fig. 1). Cette définition est issue à la fois d'une histoire commune (empires précoloniaux Ouest-Africains et, pendant la période coloniale, l'Afrique Occidentale Française) et d'une géographie particulière, située entre le Sahara et l'océan Atlantique (Atlas de l'Afrique de l'Ouest : CILSS 2016). Dans ce mémoire nous définirons l'Afrique de l'Ouest comme la région comprenant les 15 pays membres de la Communauté Economique des Etats de l'Afrique de l'Ouest (CEDEAO), un espace d'échange et de coopération économique et politique, et la Mauritanie, soit une région de 6.1 millions de km².

L'Afrique de l'Ouest est traversée en son nord, et sur toute sa longueur, par le Sahel qui constitue la zone de transition climatique entre le désert du Sahara (au nord) et le domaine soudanien (au sud). À l'est, on considère le fleuve Niger et la ligne entre le mont Cameroun et le lac Tchad comme les frontières naturelles orientales de la région (CILSS 2016). Le relief de l'Afrique de l'Ouest est plus marqué au sud où l'on distingue plusieurs zones de montagne comme les hauts plateaux Guinéens (regroupant plusieurs massifs dont celui du Fouta Djalon, 1 000 m) et le plateau de Jos au Nigéria (1 300 m). Ainsi la région est souvent considérée, à tort, comme homogène alors qu'elle possède une très grande diversité de paysages et d'écosystèmes du fait des nombreuses zones climatiques qu'elle traverse et de la diversité de ses sols.



Fig. 1. Pays de l'Afrique de l'Ouest (espace CEDEAO et Mauritanie) et leur capitale (source fond de carte : FAO).

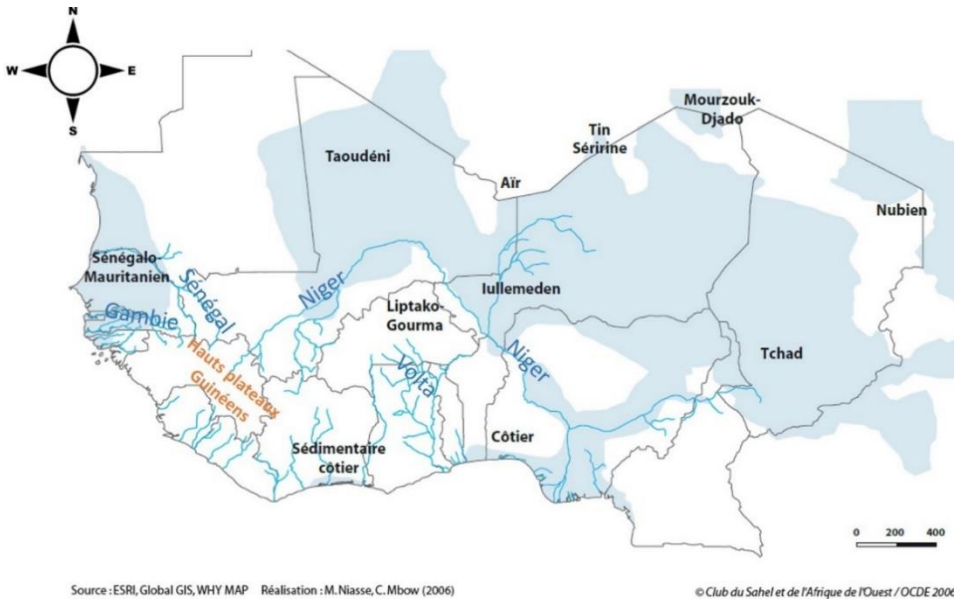


Fig. 2. Principaux fleuves d'Afrique de l'Ouest et limites des trois principales nappes phréatiques fossiles (source de la carte : CSAO/OCDE 2006).

Au sud de l'Afrique de l'Ouest, les hauts plateaux guinéens, situés en zone tropicale semi-humide, reçoivent une très grande quantité de précipitation (> 2 000 mm par an) (Sall et al. 2007). Dans cette zone, surnommée le « château d'eau de l'Afrique de l'Ouest », se situent les sources des principaux fleuves de la région : le Niger (long de 4 180 km), le Sénégal (1 750 km) et le fleuve Gambie (1 120 km) (CSAO/OCDE 2006; Sall et al. 2007). Les principaux cours d'eaux traversent donc des zones à la pluviométrie très élevée, avant de venir alimenter des zones où la pluviométrie est plus faible (le Sahel). L'Afrique de l'Ouest dispose d'une réserve importante d'eau présente dans les nappes souterraines réparties en trois grands ensembles (Fig. 2) dont l'un couvre l'entièreté du Sénégal où se situe notre zone d'étude (CSAO/OCDE 2006). On estime aujourd'hui les ressources en eaux de la région à 1 300 milliards de m^3 dont seulement 1% sont consommées, majoritairement par l'agriculture (75%). La profondeur de ces nappes sédimentaires, parfois supérieure à 1 000 m, constitue un frein à leur exploitation qui nécessite de lourds moyens (ex : forage). Ainsi seules les nappes superficielles sont majoritairement exploitées par les petits agriculteurs via l'installation de puits. Le niveau de ces nappes ainsi que les eaux de surfaces sont grandement impactées par les changements climatiques et la pression démographique (voir Partie I.4).

b. Climat, type de sol et zones agroécologiques

Du fait de son étendue géographique, l'Afrique de l'Ouest se trouve sous l'influence de trois régimes climatiques (classification de Köppen-Geiger) : un climat semi-aride au nord, un climat tropical de savane au centre, et un climat tropical humide au sud. La répartition des précipitations suit donc un gradient du nord au sud (Fig. 3) : le sud de la région reçoit plus de précipitations, entre 1 500 et plus de 2 500 mm par

an, généralement réparties sur une plus longue période de l'année, tandis que le nord reçoit entre 200 et 400 mm (Kouassi et al. 2010 ; Fick and Hijmans 2017). De la même manière, la saisonnalité et les températures suivent un gradient du nord au sud avec une influence de l'océan lorsque l'on se rapproche des côtes (Kouassi et al. 2010). Les premières pluies et l'augmentation des températures apparaissent sur la zone de convergence des flux atmosphériques qui forment le front intertropical (FIT). Au mois de mars le FIT se situe sur les zones côtières du sud. Puis, sous l'effet de la circulation atmosphérique, le FIT remonte progressivement vers le nord et les précipitations augmentent. Aux mois de juin et juillet, le FIT couvre alors l'ensemble du Sahel (mousson). En septembre le FIT redescend vers l'équateur. De ce fait, le nord de l'Afrique de l'Ouest n'a qu'une seule saison des pluies, tandis que le sud a deux saisons des pluies (mars et septembre).

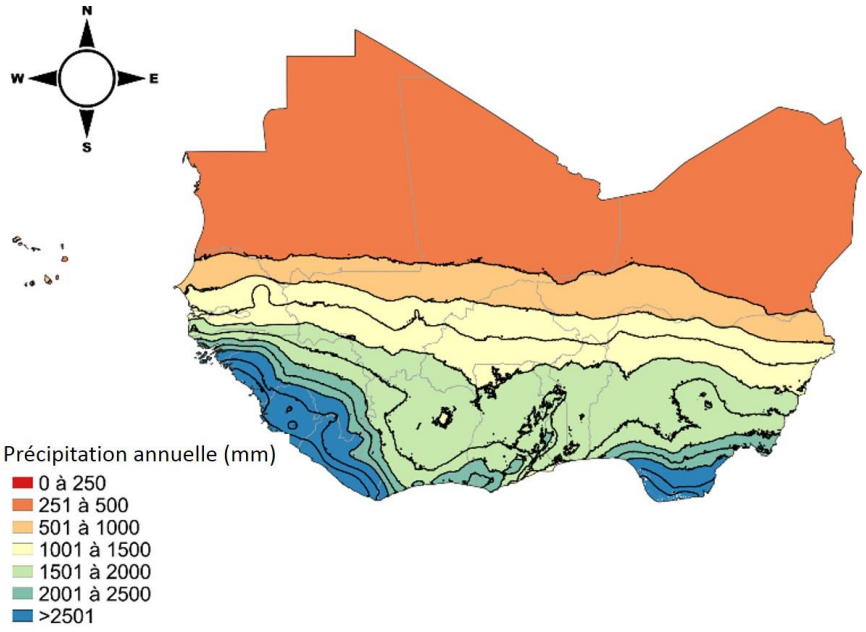


Fig. 3. Carte des isohyètes (lignes de précipitation) de l'Afrique de l'Ouest. Un isohyète est une ligne reliant les points d'égales quantités de précipitations (source des données : WorldClim 2, Fick and Hijmans 2017).

Selon le régime climatique, l'humidité du sol et sa qualité se trouvent impactées. Au Sahel, les sols sont légers et sableux et leur fertilité est faible (Issaka et al. 1996). Ces sols sont adaptés à la culture de certaines céréales (sorgho, maïs, mil) et légumineuses (arachide) mais restent peu productifs du fait de leur très faible humidité et fertilité (Bationo et al. 2007). Ils sont également sensibles à l'érosion du fait de leur composition majoritaire en sable et de leur structure légère (Bayala et al. 2019). Dans les zones de bas-fonds (ex : Niayes au Sénégal), la fertilité a tendance à être meilleure du fait de l'humidité plus abondante et on retrouve des sols caractéristiques de zones humides sédimentaires (Issaka et al. 1996 ; Fall and Fall 2001). Les bas-fonds constituent des régions importantes pour l'agriculture où l'horticulture (maraîchage et arboriculture) est pratiquée. Au Sahel, et comme sur le reste de l'Afrique, on observe une balance négative dans la fertilité

des sols: le faible taux de matières organiques n'est pas compensé par une application suffisante d'intrants (Bationo et al. 2007). Ces effets participent fortement à l'appauvrissement de la qualité des sols.

Les gradients climatiques et la diversité de types de sol délimitent plusieurs domaines agroécologiques (ou agro-climatiques) qui traduisent des ensembles écologiques et floristiques observés aussi selon un gradient nord-sud (Arbonnier 2009). En Afrique de l'Ouest, on retrouve ainsi : le domaine saharien, le domaine sahélien, le domaine soudanien, le domaine guinéen et le domaine centrafricain (Fig. 4). Dans ce mémoire nous parlerons majoritairement du Sahel où se situe notre région d'étude, la région des Niayes (Sénégal). Cette zone recouvre les domaines sahélien et sahélo-soudanien plus au sud et recoupe une large partie de 5 pays d'Afrique de l'Ouest : le Sénégal, la Mauritanie, le Niger, le Burkina Faso et le Mali. La flore du Sahel est une flore de savane sèche ouverte (ou steppe arbustive) composée majoritairement de plantes herbacées, d'arbustes (ex : *Ziziphus mauritiana*, *Moringa oleifera*) et de quelques espèces arborées majoritairement *Balanites aegyptiaca* et les espèces du genre *Acacia* (Arbonnier 2009). Cette zone agro-climatique est caractérisée par un climat semi-aride avec une seule saison des pluies de juin/juillet à août/septembre (Panthou et al. 2018). Le régime pluviométrique de la région est faible (200 à 400 mm par an) mais extrêmement variable spatialement avec parfois de forts événements pluvieux et localisés. Cette variabilité et la dépendance de l'agriculture aux précipitations rendent cette région très vulnérable au changement climatique (voir Partie I.4.b).

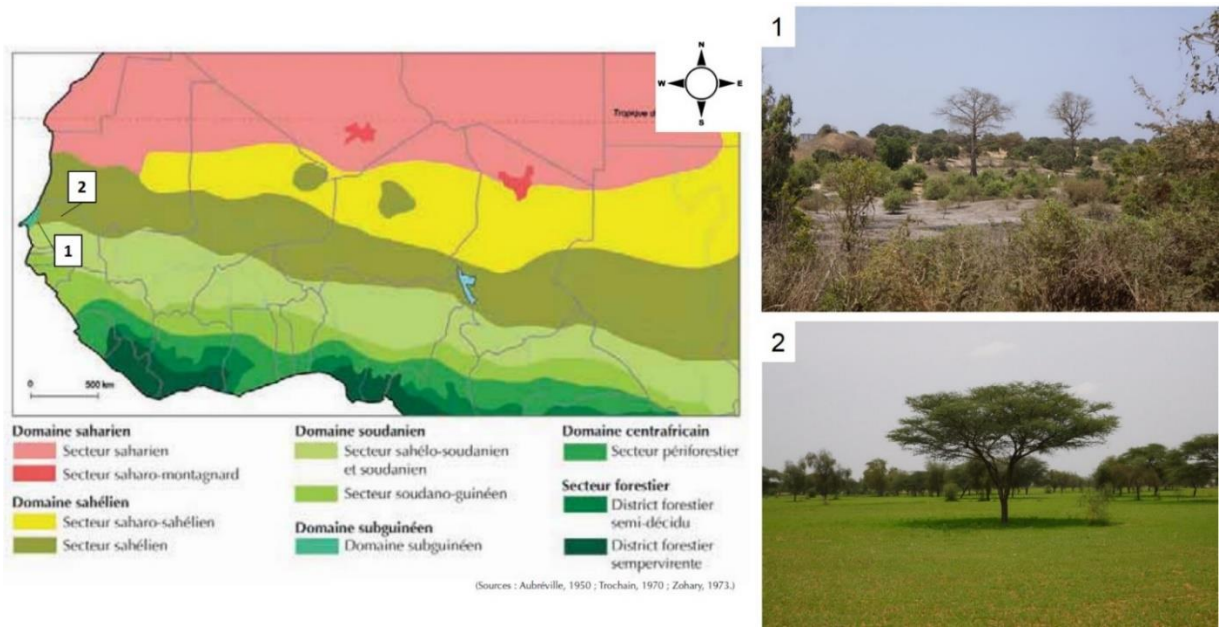


Fig. 4. Domaines agroécologiques de l'Afrique de l'Ouest. Le Sahel regroupe le domaine sahélien et le sous-domaine sahélo-soudanien limitrophe. À l'intérieur même de cette bande on retrouve une diversité floristique et paysagère qui évolue avec les saisons. Par exemple, la région des Niayes (image 1), zone de bas-fonds, et la zone du Ferlo (image 2 au centre un Acacia tortilis raddiana) en saison des pluies (source carte : Arbonnier 2009 ; source photo : J. Sarron).

I.2. Contexte socio-économique

En 2015, la population de l'Afrique de l'Ouest se situait autour de 350 millions d'habitants dont 45% se trouvaient au Nigéria, pays le plus peuplé (United Nations 2019). La population est très jeune (Fig. 5), l'âge médian est estimé à 18 ans, et en constante augmentation avec une projection de 800 millions d'habitants pour 2050 ; soit la région avec la plus forte croissance démographique dans le monde. La croissance démographique est accompagnée d'une croissance économique, le Produit Intérieur Brut (PIB) ayant augmenté de 4 à 8% par an selon les pays entre 2012 et 2015 (African Development Bank 2018).

Malgré ces bonnes performances, la région fait face à une forte présence de l'emploi informel, non régulé par l'État, qui concerne notamment le secteur agricole conduisant à une grande précarité des travailleurs. Ainsi, bien qu'en réduction, le pourcentage de la population vivant sous le seuil de pauvreté était de 43% en 2013 (African Development Bank 2018). La croissance démographique constitue un défi majeur pour la région tant sur le plan économique qu'alimentaire (FAO 2015 ; van Ittersum et al. 2016). Les études montrent que le taux d'urbanisation augmente, en partie à cause d'un exode rural (poussé par les changements environnementaux), de même que le nombre de personnes arrivant sur le marché du travail, la population étant actuellement très jeune (Courtin and Guengant 2011). Maintenir les populations actives dans le secteur agricole tout en augmentant leur revenu constitue une piste pour garantir l'avenir économique des populations rurales et augmenter l'autosuffisance alimentaire des pays (Hollinger and Staatz 2015).

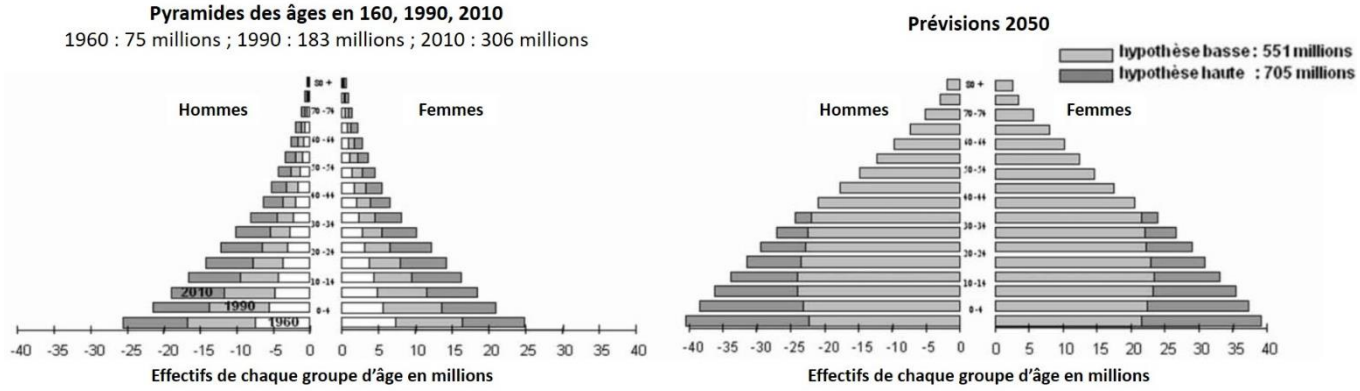


Fig. 5. Pyramide des âges et démographie en Afrique de l'Ouest (1960, 1990 et 2010) et prévisions pour 2050. Les prévisions sur ce graphique affichent une population à 700 millions qui a depuis été réévaluée et se situe aujourd'hui autour de 800 millions (source : Courtin and Guengant 2011).

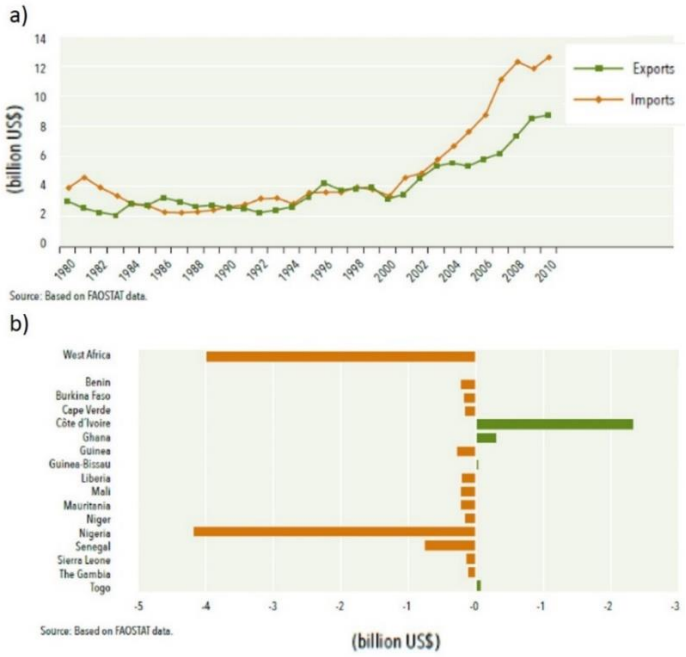


Fig. 6. Balance commerciale en milliards de dollars des produits alimentaires en Afrique de l'Ouest (2013) : (a) évolutions des importations et exportations des produits alimentaires de la région, (b) balance commerciale (exportation – importation) par pays (source figure : Hollinger and Staatz 2015).

I.3. Productions agricoles

En 2017, l'agriculture représentait 28% du PIB de l'Afrique de l'Ouest, soit une part toujours très importante (African Development Bank 2018). À titre de comparaison, la part de l'agriculture dans le PIB était de 3.4% dans le monde et de 1.5% dans l'Union Européenne. L'agriculture en Afrique de l'Ouest est principalement pratiquée sur des petites surfaces et 80% des exploitations font moins de 2 ha (Lowder et al. 2016). On observe également une variabilité de systèmes de culture qui sont parfois très diversifiés (Sibhatu et al. 2015). Les petits producteurs ont ainsi tendance à pratiquer à la fois une agriculture vivrière (d'autosuffisance) et de rente en fonction de l'accès aux intrants et aux marchés, et du contexte socio-économique. Selon les chiffres de la Food and Agriculture Organization (FAO), en termes de surface, le sorgho, le mil, le maïs, le niébé et le riz constituent les cinq espèces les plus cultivées en Afrique de l'Ouest sur 61.6 millions d'hectares (FAOSTAT 2017). Mis à part le riz, les cultures annuelles sont pluviales, c'est-à-dire qu'elles sont cultivées durant la saison des pluies rendant ces productions dépendantes des aléas climatiques (Challinor et al. 2007). Selon une étude, la production agricole totale de la région a augmenté entre 1985 et 2015, poussée par la croissance démographique, l'augmentation du niveau de vie et le changement de régime alimentaire (Hollinger and Staatz 2015). Cependant, cette augmentation ne s'est pas concrétisée pour toutes les productions. Ainsi, si les produits agricoles de base tels que le maïs et les

oléagineux ont connu la plus forte croissance sur cette période - entre 1 et 3% par an et par habitant - le mil, le riz et les fruits ont connu une augmentation moindre autour de 1% par an et par habitant. L'étude souligne également que l'augmentation de la production agricole n'est pas systématiquement suivie par une augmentation de la valeur ajoutée. Malgré ces performances relativement bonnes, l'Afrique de l'Ouest reste déficitaire dans sa balance commerciale des produits agricoles (Fig. 6a). Ainsi, bien que la région ait augmenté la production de certaines denrées de rente (maïs et oléagineuses), elle importe plus qu'elle n'exporte et seulement 4 pays ont une balance positive (Fig. 6b).

I.4. Sécurité alimentaire et changements globaux

a. Crises alimentaires de 2007-2008 et état actuel de la sécurité alimentaire

La forte dépendance de l'Afrique de l'Ouest aux importations et aux marchés internationaux fait que la région reste vulnérable aux aléas de plusieurs natures. La crise alimentaire mondiale (ou « émeute de la faim ») de 2007-2008 et qui a touché en partie la région a exposé les limites d'une vision globale de la sécurité alimentaire (Koning and van Ittersum 2009 ; Godfray et al. 2010 ; van Ittersum et al. 2013). Les différents facteurs ayant provoqué la crise sont : la stagnation des rendements agricoles, l'inégale répartition de la production alimentaire mondiale, une volatilité des prix, et la compétition entre les différentes utilisations des productions agricoles, c'est-à-dire entre alimentation de base pour l'Homme, l'alimentation du bétail et l'utilisation pour l'industrie (ex : biocarburant) (van Ittersum et al. 2013 ; Godfray and Garnett 2014). Depuis les années 1980, la Révolution Verte (i.e. intensification des systèmes de culture) a eu pour effet de concentrer les investissements sur l'augmentation de la production d'aliments de base, souvent des céréales riches en énergie, laissant de côté des cultures pourtant importantes telles que les fruits et les légumes (Gómez et al. 2013). Ainsi, la crise de 2007-2008 a mis en évidence la « faim cachée » redéfinissant la notion même de sécurité alimentaire. Cette dernière ne doit donc plus se mesurer uniquement par la quantité de nourriture produite mais aussi par sa qualité nutritionnelle et son accessibilité (FAO 2008).

La situation actuelle de la sécurité alimentaire en Afrique de l'Ouest est mitigée. D'après la FAO, la prévalence des personnes en sous-alimentation a bien diminué de 12.3% à 10.7% de la population totale entre 2005 et 2014 (FAO 2018). Cependant, on observe une augmentation récente et très forte du nombre de personnes sous-alimentée du fait des conflits armés (ex : nord-Mali) et du changement climatique. En 2018, 56 millions de Ouest-Africains étaient en situation de sous-alimentation (15.1%, FAO 2018). L'Afrique reste le continent avec la plus faible autosuffisance alimentaire (Fig. 7) alors que sa croissance démographique est la plus forte (van Ittersum et al. 2016). S'ajoute à cela les problèmes de malnutritions liés à une mauvaise alimentation (anémie) ou à une suralimentation (obésité), également en augmentation

(FAO 2018). Dans ce contexte, les fruits occupent une place importante pour la sécurité alimentaire via leur apport en nutriments et leurs bénéfices pour la santé (voir Partie II.1)

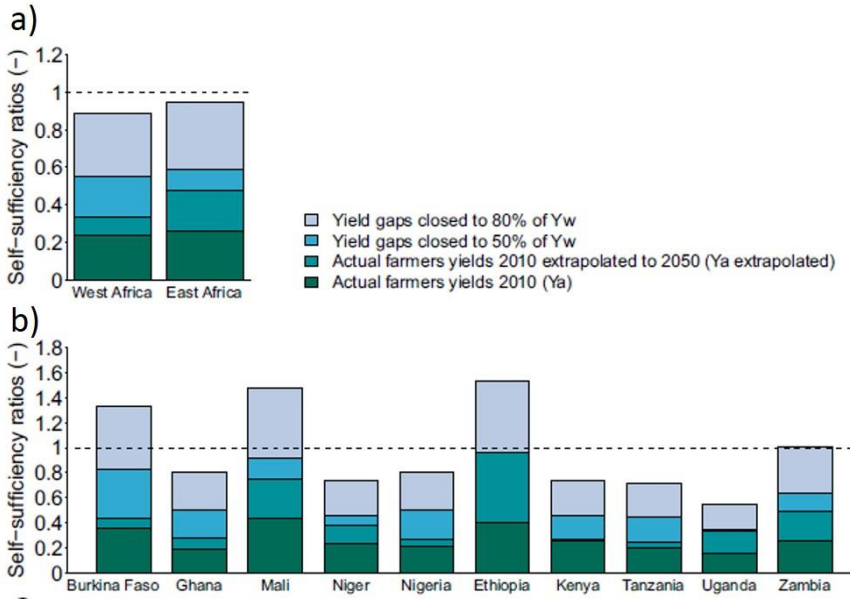


Fig. 7. Projection pour 2050 de l'autosuffisance alimentaire des 5 principales céréales (maïs, mil, riz, sorgho et blé), (a) à l'échelle de l'Afrique de l'Ouest et de l'Est et (b) à l'échelle de plusieurs pays. Ces projections sont basées sur plusieurs scénarios d'augmentation des rendements : rendement réel (Y_a) actuel (2010) ; rendement réel extrapolé avec le taux de croissance actuel, réduction des écarts de rendement avec augmentations des rendements à 50% et 80% du rendement potentiel (Y_w). En 2010, les rendements réels se situaient entre 15 et 27% du potentiel (source : van Ittersum 2016).

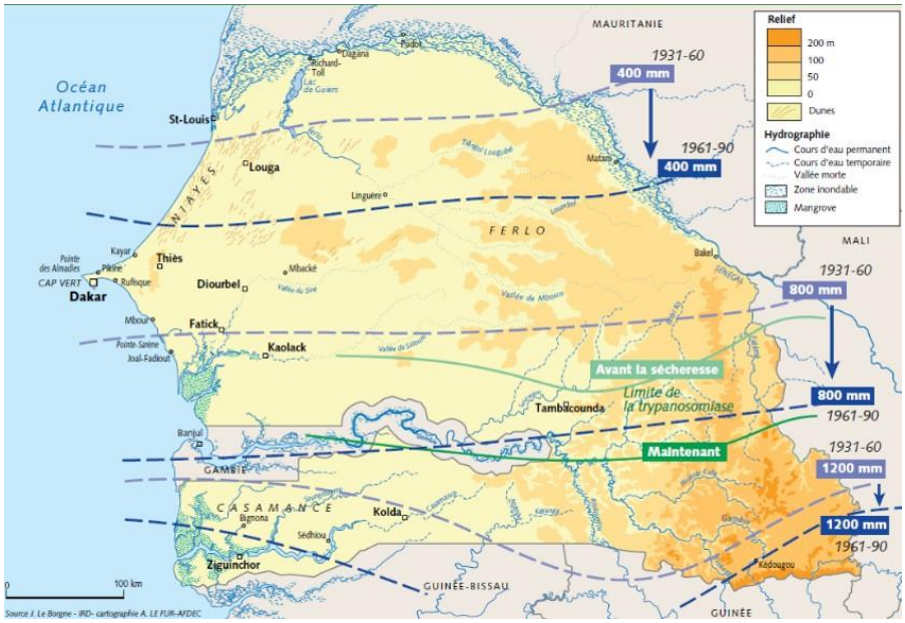


Fig. 8. Déplacement des isohyètes vers le sud observé au Sénégal à la suite des sécheresses de 1970-1980 (source figure : IRD).

b. Changement climatique et vulnérabilité de l'agriculture

Comme partout ailleurs, l'Afrique de l'Ouest est confrontée au changement climatique mais son agriculture est particulièrement vulnérable parce qu'elle est en grande partie basée sur des cultures pluviales qui ne sont pas irriguées (Challinor et al. 2007 ; Wheeler and von Braun 2013). De ce fait, l'Afrique de l'Ouest, et plus particulièrement le Sahel, a déjà connu une longue période de sécheresse dans les années 1970 et 1980 (Held et al. 2005 ; Dong and Sutton 2015 ; Panthou et al. 2018). Cette vague de sécheresse a réduit le niveau de précipitation de 40% par rapport au niveau de 1950 (Held et al. 2005) et a conduit à un déplacement des précipitations vers le sud (exemple au Sénégal : Fig. 8). Plusieurs études ont montré l'impact de ces variations de la pluviométrie (Ray et al. 2015) et des températures (Sultan et al. 2019) sur les principales cultures annuelles de la région (maïs, riz, mil, sorgho). Les prévisions futures relatives au changement climatique prévoient une réduction de 11% en moyenne d'ici à 2050 du rendement du riz, du blé et du maïs due à l'augmentation des températures (Roudier et al. 2011). Même si ces dernières années les climatologues ont constaté une augmentation des précipitations ('retour pluviométrique'), certaines études prévoient que l'allongement de la saison sèche va continuer et impacter les plantes (cultivées ou non) (Panthou et al. 2018). Le Sahel plus spécifiquement reste la région la plus vulnérable et concentre les plus fortes réductions de rendement passées et, selon les prévisions, futures (Roudier et al. 2011 ; Sultan et al. 2019). Dans certaines zones du Sahel, la durée de la saison des pluies est ainsi passée de 5 mois, en 1950, à 3 mois dans les années 2000 (Aguiar 2009). Sur les côtes, la réduction du niveau des nappes phréatiques et la montée du niveau de la mer, conduisent à une salinisation des sols déjà observée par exemple dans les Niayes au Sénégal (Aguiar 2009).

Les études de l'effet du changement climatique sur les cultures pérennes dans les zones semi-arides à tropicales sont plus rares. Plusieurs études se sont focalisées sur l'impact du changement climatique sur le développement et la production du cacao en Afrique de l'Ouest (Läderach et al. 2013 ; Schroth et al. 2016). Mais, à notre connaissance, les études sur le changement climatique et les arbres fruitiers en Afrique de l'Ouest restent inexistantes. A l'échelle globale, la recherche sur les fruitiers est principalement limitée par le fait que les informations (rendement, surface, etc.) et les modèles de culture relatifs aux arbres fruitiers sont peu nombreux (Mathur et al. 2012 ; Normand et al. 2015). Mathur et al. (2012) ont démontré que l'augmentation des températures devrait avoir un effet sur les productions de fruits tropicaux et subtropicaux, comme la mangue, qui devraient profiter d'une augmentation des températures et des conditions d'humidité. Ces résultats sont contradictoires avec ceux de Normand et al. (2015) qui soulignent un effet négatif du changement climatique sur la production du manguier, via l'augmentation des événements climatiques extrêmes et leurs impacts sur la phénologie et la floraison. Ces deux études soulignent la nécessité d'accroitre les efforts de recherche concernant les effets du changement climatique

sur les arbres fruitiers avec une prise en compte des variabilités locales. L'acquisition d'informations sur les rendements des plantes pérennes à différentes échelles font pleinement partie de ces enjeux.

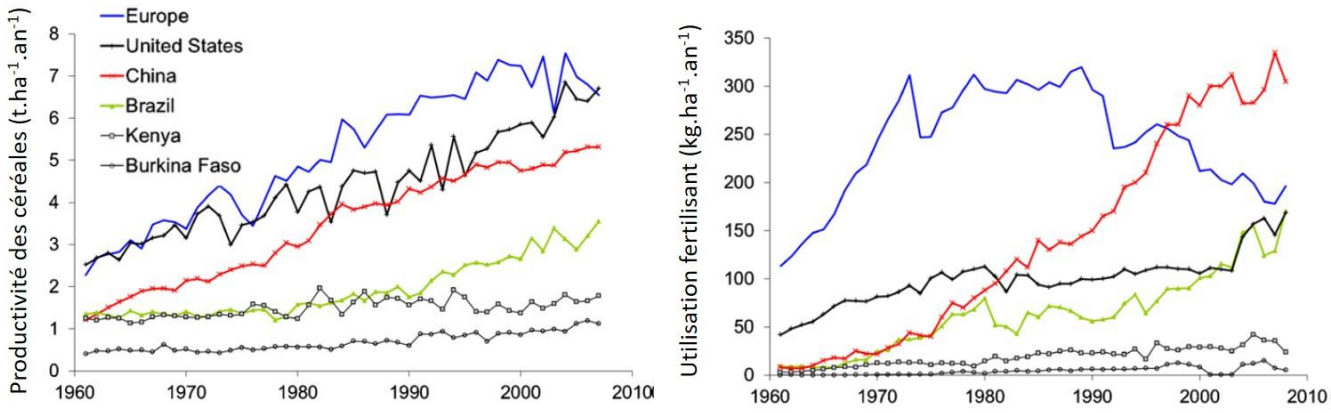


Fig. 9. Evolution de la productivité moyenne des céréales et de l'utilisation de fertilisant dans plusieurs pays de 1961 à 2008 (source : Tittonell and Giller 2013).

c. Voies vers l'autosuffisance alimentaire

Avec les projections démographiques et le changement de régime alimentaire, la demande en céréale de la population ouest-africaine devrait tripler d'ici 2050 par rapport à 2007. L'agriculture doit également traiter avec l'augmentation des pressions sur les écosystèmes et la ressource en eau dont elle est fortement dépendante notamment au Sahel (Aguiar 2009 ; Roudier et al. 2011 ; Camara et al. 2019). À l'heure actuelle, l'agriculture ouest-africaine reste familiale, pratiquée sur de petites surfaces et avec un faible niveau d'intrants (Affholder et al. 2013 ; Tittonell and Giller 2013 ; Lowder et al. 2016). Du fait de la faible fertilité des sols et des niveaux d'intrants (Fig. 9), la relative augmentation de la production agricole en Afrique de l'Ouest s'est principalement établie grâce à une augmentation des terres cultivées (Bationo et al. 2007). La région apparaît bien souvent comme un espace vide que l'on peut facilement convertir en terre cultivable (Tittonell and Giller 2013). Ainsi, contrairement à d'autres régions, les rendements sont faibles en Afrique de l'Ouest. Une étude de van Ittersum et al. (2016) estime que les rendements des principales céréales n'ont augmenté que d'environ 30 $kg.ha^{-1}.an^{-1}$ entre 1991 et 2014. En Afrique de l'Ouest, les écarts entre rendement observé et potentiel sont élevés et la réduction des yield gaps semble une des solutions les plus sérieuses pour parvenir à répondre à l'augmentation de la demande alimentaire (Fig. 7, van Ittersum et al. 2016).

Certains auteurs évoquent la nécessaire intensification de l'agriculture ouest-africaine via l'utilisation de variétés améliorées et l'apport d'intrants (Sanchez 2010). Or, la relation entre augmentation de la production agricole et la sécurité alimentaire est largement débattue (Koning and van Ittersum 2009 ; Dury et al. 2017)

et la crise alimentaire de 2007-2008 prouve qu'augmenter la quantité d'aliments produits, en particulier les céréales, ne suffit pas à réduire la pauvreté et la malnutrition (Gómez et al. 2013). Se surajoutent à ces questions les impacts environnementaux et sociaux négatifs de l'agriculture intensive (Altieri 2002 ; Pingali 2012). L'agriculture intensive se caractérise par des écosystèmes simplifiés où quelques espèces et variétés améliorées sont cultivées au moyen d'une forte demande en intrants et en ressources fossiles (Matson 1997 ; Malézieux 2012). Ce modèle agricole a cependant de nombreux impacts négatifs sur l'environnement dont : de fortes émissions de gaz à effet de serre, l'érosion et la perte de fertilité des sols, la réduction de la biodiversité, la pollution des nappes phréatiques ou encore les risques pour la santé des agriculteurs. Certains de ces impacts négatifs sont déjà visibles en Afrique de l'Ouest, par exemple en ce qui concerne les risques liés aux usages des pesticides (de Bon et al. 2014).

La nécessaire augmentation des rendements doit donc se faire au regard de ces contraintes environnementales et sociétales et prendre en compte les particularités de l'agriculture de la région (petites structures, diversité des systèmes, présence de plantes pérennes, cultures pluviales). Aujourd'hui, il existe plusieurs démarches d'agricultures qui ont été conceptualisées afin d'intensifier durablement l'agriculture (Cassman 1999 ; Doré et al. 2011). Ces démarches sont parfois imbriquées et peuvent correspondre à des principes - comme c'est le cas de l'agroécologie (Altieri 2002) ou de l'agriculture biologique (Andriamampianina et al. 2018) - ou à des techniques agronomiques - comme les associations culturales (Malézieux et al. 2009 ; Himmelstein et al. 2017). Parmi ces modèles, les systèmes agroforestiers (Malézieux 2012 ; Mbow et al. 2014a) ou associant des arbres fruitiers apportent des solutions intéressantes pour parvenir à une réduction de la pauvreté et l'insécurité alimentaire tout en garantissant une adaptation aux contraintes sociétales et environnementales (voir Partie II).

II. Les arbres fruitiers en Afrique de l'Ouest et leurs problématiques

II.1. Avantages de l'arbre dans les systèmes de culture

a. Services écosystémiques

Les travaux de cette thèse se focalisent sur les cultures pérennes en Afrique de l'Ouest, plus précisément les arbres fruitiers. Les services écosystémiques rendus par les arbres, fruitiers ou non, sont nombreux et il est difficile d'en faire une synthèse exhaustive (Malézieux et al. 2009 ; Simon et al. 2010 ; Demestihas et al. 2017). Les arbres fournissent notamment des services de production (bois et fruits), jouent un rôle dans la biodiversité, régulent le microclimat et préservent le sol et la ressource en eau (Demestihas et al. 2017). Dans les systèmes diversifiés à base de plantes pérennes (ou systèmes agroforestiers) s'ajoutent les interactions entre les différentes espèces cultivées arborées ou non (Fig. 10, Malézieux et al. 2009). Dans

certaines situations l'augmentation de la diversité apporte des interactions bénéfiques et augmente les rendements des plantes cultivées via différents mécanismes comme la régulation des bioagresseurs (Ngo Bieng et al. 2013 ; Gidoin et al. 2014). Les effets des arbres dans le maintien de la fertilité et de l'humidité du sol et dans la réduction de l'érosion ont été démontrés dans de nombreuses études (Isaac et al. 2005 ; Félix et al. 2018). Les systèmes agroforestiers offrent également des avantages dans la résilience au changement climatique via la régulation de la température et la formation d'un microclimat à l'échelle de la parcelle. Par exemple, l'ajout d'arbres d'ombrage permet de réduire la vulnérabilité du cacao à l'augmentation des températures (Schroth et al. 2016). Les systèmes agroforestiers sont également une solution pour la mitigation du changement climatique en favorisant par exemple le stockage du carbone dans le sol ou dans la biomasse aérienne des arbres (Mbow et al. 2014a).

En Afrique de l'Ouest, certains systèmes à base de plantes pérennes sont très bien étudiés. C'est le cas par exemple des systèmes à base de cacao (Gidoin et al. 2014 ; Schroth et al. 2016) ou encore les parcs agroforestiers sahéliens associant les cultures pluviales (mil et sorgho) avec *Faidherbia albida* (Félix et al. 2018 ; Bayala et al. 2019). Cependant, à ce jour, les études se focalisant sur les services écosystémiques rendus par les arbres fruitiers en Afrique de l'Ouest sont plus rares. Alors que l'évaluation des systèmes de culture est nécessaire pour permettre leur amélioration et leur diffusion auprès des producteurs (Doré et al. 2008, 2011), cette évaluation se trouve limitée en Afrique de l'Ouest par le manque de données expérimentales et de terrains (voir Partie II.2.b).

b. Sécurité alimentaire

Bien que les performances (agronomique et services écosystémiques) des systèmes de cultures à base d'arbres fruitiers restent à être plus spécifiquement étudiés, ces espèces ont un intérêt supplémentaire concernant la sécurité alimentaire (Gómez et al. 2013). On estime à environ 800 le nombre d'espèces fruitières recensées et consommées en Afrique Sub-Saharienne (Cemansky 2015). Les fruits sont la source de nombreux éléments nutritifs (sucres, vitamines, antioxydants) et possèdent de nombreux bénéfices pour la santé humaine, notamment dans la prévention de certaines maladies comme les cancers ou le diabète (Carlier et al. 1992 ; WHO 2004). Ainsi, les fruits sont vus comme une des solutions pour diversifier les régimes alimentaires et réduire la malnutrition dans le monde et en particulier en Afrique de l'Ouest (Stadlmayr et al. 2013 ; Gómez et al. 2013). Les fruits sont importants dans ces pays car ils apportent une source nutritive complémentaire en période de soudure c'est-à-dire à la fin de la saison sèche et durant la saison des pluies (de juin à septembre), lorsque que les stocks de céréales viennent à manquer (Fifanou et al. 2011 ; Jamnadass et al. 2011).

Les arbres fruitiers sont aussi une source de revenu importante ou complémentaire pour de nombreux producteurs (ex : Monteiro et al. 2017). C'est plus particulièrement le cas des fruits à haute valeur ajoutée, comme les agrumes ou la mangue, dont la production est valorisée sur les marchés locaux voire internationaux (Rey and Dia 2010). La vente de ces fruits permet au producteur de dégager un revenu conséquent lui permettant de subvenir aux besoins alimentaires (en achetant des aliments non produits sur l'exploitation) ou non du foyer (Monteiro et al. 2017).

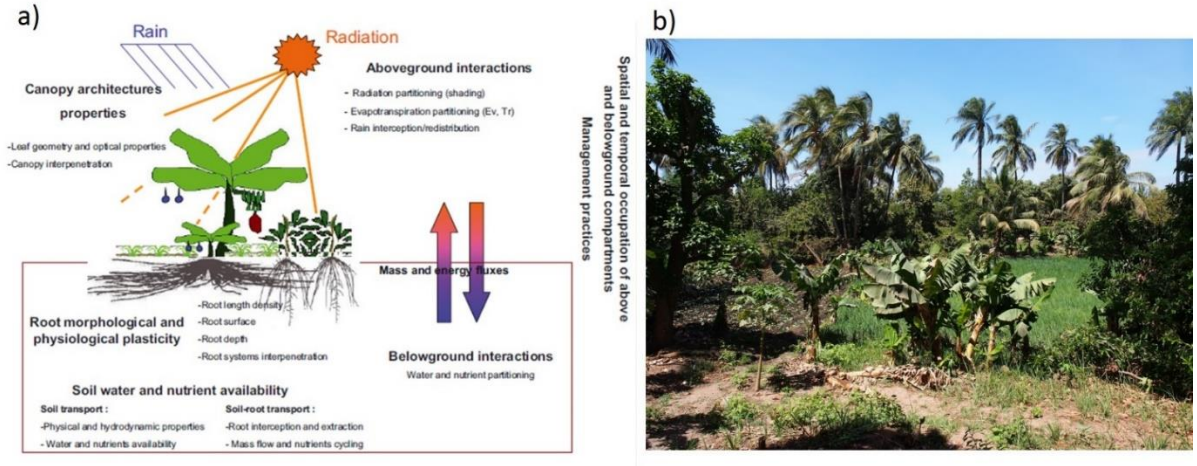


Fig. 10. (a) interactions et compétitions pour les ressources (eau, nutriments et radiations solaires) entre les plantes dans les systèmes agroforestiers. (b) exemple d'un système agroforestier : riziculture et plantation de bananiers et de palmiers en Casamance, Sénégal (source figure : Malézieux et al. 2009 ; source image : É. Faye).

II.2. Systèmes à base de fruitiers en Afrique de l'Ouest

L'arbre fruitier occupe donc une place d'importance pour la sécurité alimentaire, l'économie des agriculteurs et la résilience des systèmes agricoles en Afrique de l'Ouest (Mbow et al. 2014a, b). Dans cette région, les systèmes de culture basés, en partie (systèmes diversifiés) ou entièrement (monoculture), sur des arbres sont variés (ex : Grechi et al. 2013). Malgré le grand nombre d'espèces fruitières présentes en Afrique de l'Ouest, seules quelques-unes sont cultivées à grande échelle et valorisées sur les marchés locaux, régionaux et internationaux. Ainsi, les principaux arbres cultivés dans la région sont les manguiers, les anacardiers et les orangers, dont la production en Afrique de l'Ouest s'élevait respectivement à 2.4, 1.4 et 1.3 million de tonnes en 2017 (FAOSTAT 2017). En termes de surfaces, l'anacardier représentait 2.8 millions d'hectares, suivi des manguiers (500 000 ha) et des orangers (65 000 ha). Les données de la FAO indiquent que la production de fruits en Afrique de l'Ouest est en constante augmentation, alors que les rendements stagnent pour le manguier et l'anacardier (Fig. 11). L'augmentation de leur production résulte donc d'une augmentation des surfaces (Rey and Dia 2010 ; Monteiro et al. 2017). Par exemple, les surfaces en anacardier ont presque triplé depuis 2000 où elles étaient de 900 000 ha (FAOSTAT 2017). Ces espèces

apparaissent comme des sources de revenus importantes pour les producteurs (Fare et al. 2017 ; Monteiro et al. 2017). Mais aussi pour les pays, à l'instar de la Guinée-Bissau dont l'exportation de noix de cajou représente plus de 90% des exportations totales (Monteiro et al. 2017). Le modèle d'étude de cette thèse est le manguier qui est donc une culture d'importance pour l'économie et la sécurité alimentaire dans la région (Rey and Dia 2010). Les particularités de cette espèce (physiologique) et sa culture en Afrique de l'Ouest seront détaillées dans la partie suivante (voir Partie III).

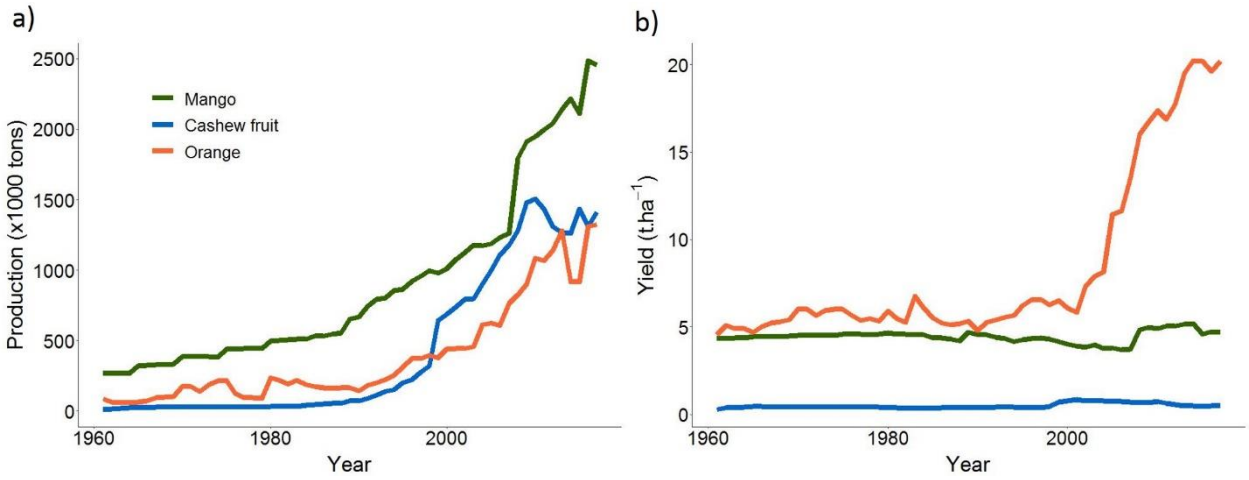


Fig. 11. Evolution (a) de la production et (b) des rendements des trois principaux fruits cultivés (mangue, pomme et noix de cajou et orange) en Afrique de l'Ouest (source données : FAOSTAT 2017).

La plupart des vergers à base de fruitiers sont de petites tailles (1 à 10 ha) et sont diversifiés : de nombreux petits producteurs cultivent une large diversité d'espèces fruitières (Vandermeer et al. 1998). La construction des systèmes (choix des espèces, densité de plantation, etc.) est issue d'un équilibre entre les contraintes des producteurs et leur choix (espèces, pratiques) et leur motivation (Malézieux 2012 ; Mbow et al. 2014a). La diversification est souvent considérée comme une voie d'intensification des systèmes et permet, dans certaines conditions, une augmentation de la production par rapport aux monocultures (Vandermeer et al. 1998). Cependant, l'étude et la compréhension de ces systèmes diversifiés restent difficiles du fait de leur complexité et de la variabilité des situations (espèces choisies, conditions pédoclimatiques, contexte socio-économique). Les aspects socio-économiques (opportunités comme contraintes) et culturels apparaissent comme des déterminants majeurs dans le choix des espèces et la conduite des cultures (ex : Fifanou et al. 2011). S'ajoute à cela l'évolution spatio-temporelle de la structure des vergers. Chez les petits producteurs, la gestion des arbres se fait souvent de manière localisée (Jagoret et al. 2018), la structure du verger évolue donc localement dans le temps en fonction de plusieurs paramètres intervenant aux échelles de l'arbre (mortalité, maladie) et du verger (historique du verger, stratégie du producteur, contexte socio-économique).

Enfin, il existe quelques exemples de systèmes de monocultures intensives mais leurs surfaces sont moindres comparées à d'autres régions comme l'Europe ou l'Amérique du Nord. En effet, ces systèmes nécessitent des moyens importants pour leur mise en place et correspondent souvent à des structures ou investisseurs étrangers (Rey and Dia 2010).

II.3. Problématiques rencontrées par les filières fruitières en Afrique de l'Ouest

Les problématiques auxquelles les systèmes à base de pérennes en Afrique de l'Ouest doivent faire face sont nombreuses et parfois particulières à chaque filière. Cependant on retrouve deux constantes communes à chaque espèce fruitière et à chaque système : les faibles rendements et le manque de données.

a. Des vergers aux rendements faibles

Comme nous l'avons déjà décrit, les rendements agricoles sont faibles en Afrique de l'Ouest. Ceci prévaut également pour les arbres fruitiers même les principales espèces qui dominent les marchés (manguier, anacardier et oranger). En effet, de nombreuses études signalent la faible productivité des arbres et des vergers (Rey and Dia 2010 ; Monteiro et al. 2017 ; Akohoue et al. 2018). Dans la plupart des cas, les arbres sont cultivés avec peu d'investissements et de connaissances dans l'utilisation des intrants (irrigation, fertilisation) et des pratiques de taille (Monteiro et al. 2017). S'en suit une mauvaise maîtrise des itinéraires techniques par les producteurs (Jamnadass et al. 2011 ; de Bon et al. 2014). Par exemple, dans les systèmes à base d'agrumes au Bénin, on estime à plus de 90% le nombre de vergers où les pratiques recommandées sont peu diffusées ou mal suivies (Akohoue et al. 2018).

Les espèces comme le manguier et l'anacardier sont certes adaptées aux climats tropicaux à semi-arides, mais leur production est sensible à certains facteurs comme la disponibilité en eau (Crane et al. 2009). Seuls les vergers mono-spécifiques avec des densités de plantation et des niveaux d'intrants plus importants parviennent à des rendements élevés. Et encore, certaines situations dans des monocultures n'obtiennent pas des rendements satisfaisants. Par exemple, les rendements de mangues peuvent être inférieurs à 5 $t.ha^{-1}$ (nord Côte d'Ivoire, Rey et al. 2007) quand ils peuvent dépasser les 50 $t.ha^{-1}$ dans d'autres régions du monde (Menzel and Le Lagadec 2017).

Les raisons de cette faible productivité sont multiples. D'abord les arbres fruitiers subissent un désinvestissement des politiques de recherche et de développement (Jamnadass et al. 2011 ; Mbow et al. 2014b ; Cemansky 2015). En effet, malgré leur avantage pour la sécurité alimentaire, les arbres fruitiers, contrairement aux céréales, ont bénéficié de moins d'attention de la part des politiques alimentaires en Afrique de l'Ouest (Gómez et al. 2013). Ceci induit un appui technique difficile des filières et un faible apport de la recherche expérimentale en ce qui concerne l'évaluation des pratiques culturales et la sélection

variétale (Cemansky 2015). Ensuite, la complexité biologique des espèces tropicales est également mise en cause. Les arbres fruitiers tropicaux sont des modèles complexes dont le comportement de reproduction est alternant : une année pouvant être productive et la suivante non (Monselise and Goldschmidt 1982). La phénologie est également asynchrone, signifiant, qu'à l'échelle de l'arbre, plusieurs stades phénologiques différents peuvent être observés sur une même branche (Dambreville et al. 2013a). À l'échelle du verger, cela se traduit par une hétérogénéité spatiale de production. Ces phénomènes sont difficilement compris et très peu étudiés en Afrique de l'Ouest (Diatta 2018).

b. Le manque de données

Le second frein principal au développement des cultures fruitières est le manque de données fiables dans les différents types de systèmes (ex : Andriamampianina et al. 2018). Bien que des bases de données, comme celles procurées par la FAO (ex : FAOSTAT) soient disponibles, elles ne donnent des informations qu'à l'échelle des pays et, dans de plus rares cas, à l'échelle régionale. À l'heure actuelle, les bases de données ou les relevés statistiques, de l'échelle de l'arbre à l'échelle du bassin de production, sont soit de mauvaise qualité soit inexistantes. Il existe donc très peu d'études et d'informations décrivant les systèmes de culture à base de pérennes en Afrique de l'Ouest. Ces informations concernent autant les données expérimentales que les données d'observation issues des vergers (structures, surfaces et niveaux de production) (Carletto et al. 2015). Dans certains cas, des données sont disponibles mais elles se révèlent bien souvent de mauvaise qualité. La difficulté à décrire les systèmes complexes aggrave ce manque d'informations (Vandermeer et al. 1998). Enfin, à l'heure actuelle, il n'existe pas de modèle de culture pour les arbres fruitiers tropicaux, ce qui empêche d'utiliser ces outils pour la caractérisation des productions (Normand et al. 2015).

Ce manque de données impacte tous les acteurs des filières fruitières. Tout d'abord, les petits producteurs sont dans l'incapacité d'estimer précisément leur rendement, même quelques jours avant la récolte. Pourtant, la donnée de rendement est primordiale pour les petits producteurs qui peinent à tirer parti de leur production : la méconnaissance exacte de la quantité de fruits dont ils disposent les empêche de maximiser leur bénéfice. Dans les monocultures, une estimation peut être réalisée mais elle reste peu fiable car basée sur un échantillonnage couplé à l'expertise du producteur (Anderson et al. 2019). Dans ces systèmes, l'estimation du rendement, deux à trois mois avant la récolte, fait partie des tâches critiques pour la préparation de la logistique et la négociation sur le marché de l'export (Payne et al. 2014 ; Linker 2017). Les données sont également importantes pour les acteurs en aval des producteurs tels que les acheteurs et les exportateurs qui souhaitent connaître les volumes disponibles mais également avoir une traçabilité des produits (origine, doses d'intrants utilisés, etc.). Enfin, le manque d'informations freine également la recherche et le développement qui ne disposent pas des données nécessaires (rendements, pratiques, surface,

etc.) pour réaliser un appui technique (ex : outil d'aide à la décision) ou évaluer les systèmes de culture (Doré et al. 2011).

Il existe pourtant des méthodologies et des outils de mesure conçus pour la caractérisation de la structure et des rendements des vergers. Ces outils sont basés sur la télédétection à partir d'images satellites (Rahman et al. 2018) ou de données prises au sol (ex : LiDAR, Stein et al. 2016). Ils sont développés sur des vergers monospécifiques avec plus ou moins de succès. Il reste que ces outils ne sont pas ou peu utilisés en région tropicale du fait de la complexité des vergers (réduisant l'efficacité de la télédétection) et de leur coût élevé. S'ajoutent à cela les contraintes climatiques expérimentées en région tropicale : la forte couverture nuageuse réduit considérablement les fenêtres d'observation de ces régions par l'imagerie satellitaire (Burke and Lobell 2017). De plus, l'utilisation de ces outils est plus limitée chez les petits producteurs du fait de la complexité de la structure des vergers et de nombreuses contraintes techniques (couverture nuageuse, petite taille des parcelles, etc.) (Usha and Singh 2013). Le développement d'outils 'low-cost' ou frugaux (drone, capteur visuel simple, smartphone, etc.) offre des opportunités notamment pour les petits agriculteurs. Par exemple, le drone, qui n'est pour l'instant que très peu utilisé, pourrait avoir de réels avantages pour caractériser les systèmes agroforestiers (Pádua et al. 2017).

III. Cas d'étude : le manguier au Sénégal

III.1. Description de l'espèce et production

a. Taxonomie

Le manguier (*Mangifera indica L.*) est une espèce rattachée à la famille des Anacardiaceae qui comporte d'autres espèces tropicales cultivées comme l'anacardier et le pistachier (Bompard 2009). Le sous-genre *Mangifera* comporte 69 espèces dont 26 sont comestibles et caractérisées par un fruit au goût acide et astringent. L'espèce *M. indica* est la plus cultivée et consommée et fait partie des seules espèces du genre *Mangifera* à être adaptée à des climats semi-arides, comme le Sahel.

La culture du manguier remonterait à 4 000 ans et son bassin d'origine se situerait entre le nord-est de l'Inde et la Birmanie (Bompard 2009). On distingue deux types de sous-espèces de manguier selon leur mode de reproduction et leur centre de diversité. Un premier groupe, originaire des régions subtropicales (Inde), est à graine monoembryonnaire. Le second, originaire des zones tropicales (sud-est de l'Asie), est à graine polyembryonnaire (Mukherjee and Litz 2009). L'Inde est la zone d'origine de nombreuses variétés ('Alphonso', 'Dashehari', 'Langra', 'Rani Pasand', etc.) qui se maintiennent par multiplication végétative. La mangue a d'abord été introduite en Afrique de l'Est par les Arabes (XIIIe siècle), mais ce sont les

Portugais et les Espagnols qui, depuis leurs établissements coloniaux en Inde et en Asie du Sud-Est, diffusent l'espèce en Afrique de l'Ouest (au XIXe siècle), puis en Amérique du Sud (Rey et al. 2004 ; Mukherjee and Litz 2009). Vers 1860, la mangue est introduite en Floride qui va constituer un second centre de diversité d'où seront créées de nombreuses variétés appelées « floridienne » ('Haden', 'Tommy Atkins', 'Kent', 'Keitt', etc.). Les variétés floridiennes sont adaptées à différentes zones agroécologiques contrairement aux variétés indiennes, souvent improductives en dehors de leur centre de domestication.

b. Production dans le monde

Aujourd'hui on ne dénombre pas moins de 1 000 variétés de mangue dans le monde mais seulement une quinzaine sont cultivées et commercialisées à grande échelle (Mukherjee et Litz 2009). Le manguier est cultivé dans une centaine de pays, sous des climats subtropicaux (Brésil, Afrique du Sud), tropicaux (Asie du Sud) ainsi que sous des climats semi-arides et tropicaux secs (Pérou, Queensland en Australie, Sahel en Afrique) (Mitra 2016). La mangue est le 5ème fruit le plus produit dans le monde (en tonnage) - derrière les bananes/plantains, les agrumes, le raisin et les pommes - avec une production de 51 Mt en 2017 (FAOSTAT 2017). Elle constitue le fruit le plus important en Asie où est produit 76.5% de la production mondiale, notamment par l'Inde (16 Mt) et la Chine (4.3 Mt). Les continents Américain et Africain représentent ensuite respectivement 12.6 et 10.8% de la production mondiale (Mitra 2016). Cependant, l'Afrique de l'Ouest est la région du monde, avec l'Afrique de l'Est, où l'augmentation de la production de mangues a été la plus forte cette dernière décennie : +93 % entre 2000-05 et 2010-15 (Fig. 11a et 12). Les principaux pays exportateurs sont le Mexique (23%), le Brésil (14.3%) et le Pérou (10.3%) qui exportent majoritairement vers les Etats-Unis et l'Union Européenne, deux régions à très forte demande (Ternoy et al. 2006). Cependant la part de mangues exportée dans le monde ne représentait que 3% de la production totale en 2017 (FAOSTAT 2017).

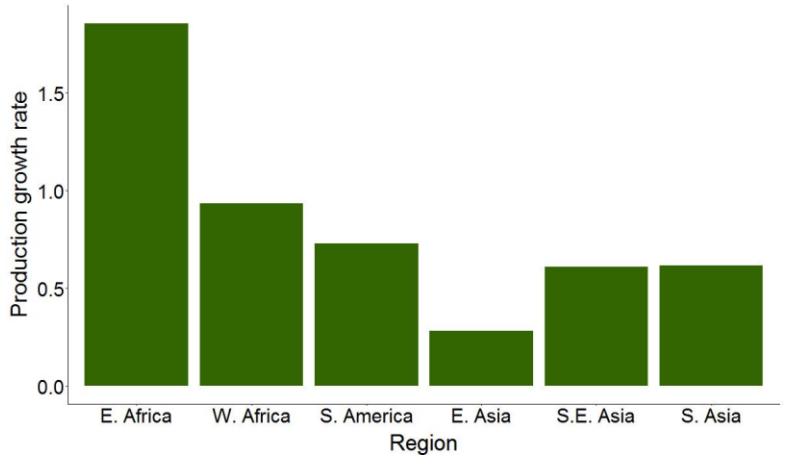


Fig. 12. Taux de croissance de la production de mangues dans les principales régions de production dans le monde entre 2000-05 et 2010-15 (source données : FAOSTAT 2017).

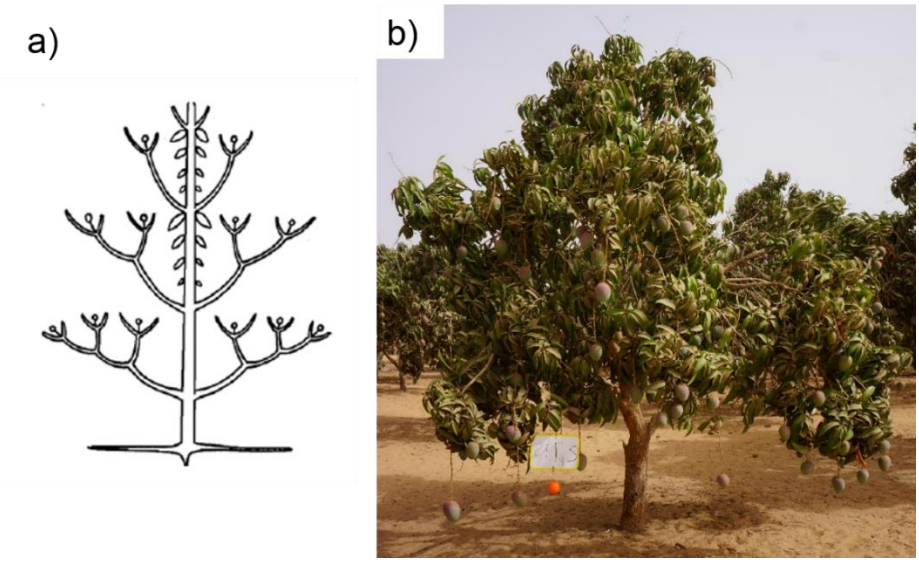


Fig. 13. Architecture du manguier : (a) le modèle de Scarrone (Hallé et al. 1978) et (b) manguier de variété 'Kent' en production.

III.2. Architecture et botanique du manguier

Le manguier (Fig. 13) peut atteindre une hauteur de 40 mètres et vivre plusieurs centaines d'années. Sa morphologie suit un modèle architectural de Scarrone (Fig. 13a., Hallé et al. 1978), défini par un tronc monopodial qui porte des étages de branches sympodiales, c'est-à-dire caractérisées par une croissance terminale (Normand et al. 2009 ; Normand and Lauri 2018). Selon les modes de conduite et les variétés, le manguier entre en production entre 3 à 5 ans après la plantation (Mukherjee and Litz 2009). Le système racinaire est caractérisé par une ou plusieurs racines pivotantes pouvant atteindre 6 mètres de profondeur et étant accompagnées de très nombreuses racines latérales et proches de la surface (Crane et al. 2009). Ce système pivotant confère une résistance de l'arbre aux vents cycloniques et à la sécheresse.

Les feuilles sont persistantes et peuvent rester jusqu'à 3-4 ans sur l'arbre. Leur forme est variable selon les variétés (Mukherjee and Litz 2009). Les fleurs du manguier sont regroupées en inflorescence qui ont une forme de panicule et qui peuvent compter jusqu'à plusieurs milliers de fleurs en fonction de la variété (Fig. 14). Les inflorescences apparaissent majoritairement à partir du bourgeon apical de l'unité de croissance (floraison sympodiale), ce qui conduit à un développement des fruits à la périphérie du houppier (Normand and Lauri 2018). Les fleurs peuvent être hermaphrodites, qui sont les seules à produire des fruits, ou uniquement mâles (Iyer and Schnell 2009). La proportion des deux types dépend des conditions environnementales et de la variété. Les fleurs hermaphrodites restent minoritaires, ce qui laisse penser que leur nombre peut être un facteur limitant la nouaison et donc la production de l'arbre.



Fig. 14. Inflorescence de manguier, variété 'Kent' (source photo : J. Sarron).

La mangue est une drupe (à noyau mono- ou polyembryonnaire) de forme et de couleur très variables et fonction de la variété (Fig. 15). Le fruit peut donc être allongé ou ovale et se terminer par un bec ou non. Sa longueur varie de 2.5 à 30 cm pour les variétés les plus grosses. Les couleurs de l'épiderme (ou péricarpe) varient également du vert au rouge/jaune selon la variété, mais également du stade de maturité et de l'exposition au soleil (Nordey et al. 2014a). Le mésocarpe (partie située entre le noyau et le péricarpe), d'épaisseur variable, a un aspect plus ou moins fibreux et est la partie consommée contenant les matières nutritives. Avant maturité, le fruit reste acide, astringent et riche en acide ascorbique (vitamine C) mais leur taux chute plus ou moins rapidement, pendant la maturation, en fonction de la variété et des conditions extérieures (Lalel et al. 2003). Avec la maturation, la mangue devient riche en vitamine A, en sucre (dont 11% de fructose) et en antioxydants, des propriétés nutritionnelles intéressantes pour la santé et la sécurité alimentaire (Carlier et al. 1992). Enfin, la mangue est un fruit climactérique (i.e., caractérisé par un pic d'éthylène pendant le murissement), ce qui lui confère un avantage pour l'exportation (Lalel et al. 2003).

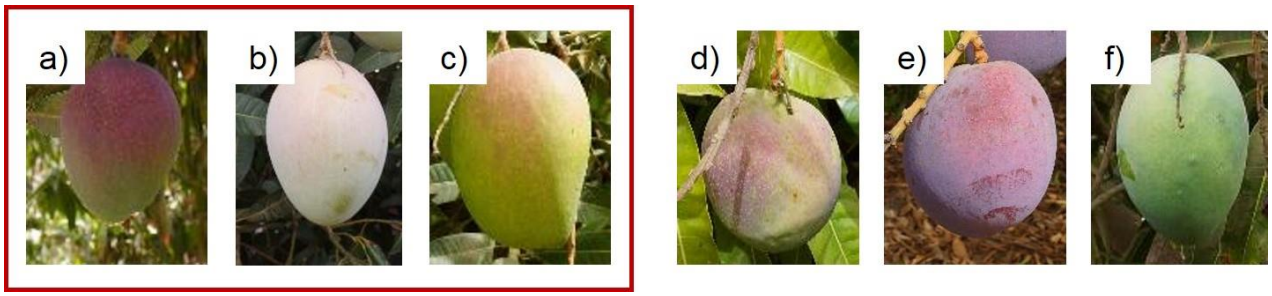


Fig. 15. Variabilité de forme et de couleur en fonction de la variété de mangue dans la région des Niayes au Sénégal. Variétés : (a) 'Kent', (b) 'Keitt', (c) 'Boucodiékhal', (d) 'Divine', (e) 'Sanguine' et (f) 'Papaye'. Le cadre rouge indique les trois variétés d'étude de la thèse (source photo : J. Sarron).

III.3. Cycle phénologique et asynchronisme

La croissance végétative du manguier a lieu généralement après la récolte et suit en réalité un rythme de petites phases successives de croissance (flush, Fig. 16 et 17). À chaque flush correspond l'apparition d'une unité de croissance (UC) dont le développement est dépendant de la température (Sukhvibul et al. 1999 ; Dambreville et al. 2013a, b). Suite à son initiation, l'UC devient soit une pousse végétative (phase végétative), soit une inflorescence selon les conditions environnementales. La différenciation de l'UC en inflorescence est induite par des conditions différentes selon les climats (Sukhvibul et al. 1999). On retiendra que, dans les zones subtropicales et semi-arides, la floraison est initiée par des températures fraîches (18°C le jour/10°C la nuit) qui correspondent à la saison sèche sous ces climats (Ramírez and Davenport 2010). La condition hydrique semble cependant avoir un impact sur l'intensité de la floraison et sur la longueur des inflorescences. Généralement les producteurs démarrent donc l'irrigation à l'apparition des premières fleurs (González et al. 2004). La fructification dure plusieurs mois selon les variétés et les conditions hydriques et de température. Le fruit se gorge d'abord d'eau et de matière sèche et augmente de volume (phase d'élongation) (Léchaudel et al. 2007). Sur pied, les fruits arrivent à maturité de manière variable. Il est difficile de déterminer le stade de maturité car, le fruit étant climactérique, il n'y a pas de paramètres visuels simples pour son estimation. Par exemple, la couleur du fruit varie à la fois avec la maturation mais aussi en fonction des conditions de lumière (Nordey et al. 2014).

La succession des flushes peut se décaler dans le temps à l'échelle d'un même arbre. L'initiation et le devenir (pousse végétative ou inflorescence) d'une UC dépend à la fois de paramètres structuraux (ex : position dans la branche et diamètre de l'UC mère, surface foliaire) et temporels (âge de l'UC mère) (Normand et al. 2009 ; Dambreville et al. 2013a). S'ajoute à cela un effet des différents UC entre elles conduisant à une superposition des cycles de croissance (Fig. 17). En effet, il existe une compétition pour la ressource en carbone entre les fruits et les pousses végétatives en croissance qui se surajoute à un effet temporel de la floraison sur la phase végétative suivante. Ceci caractérise le coût de la reproduction : plus l'arbre dépensera d'efforts à la floraison, plus il sera chargé en fruits mais ceci impactera négativement l'initiation des UCs du cycle suivant (Capelli et al. 2016). Ainsi un arbre très chargé une année aura un développement végétatif et reproductif réduit l'année suivante. Les influences spatio-temporelles entre les UC conduisent à un décalage dans l'apparition des flushes sur l'arbre. Ceci abouti à l'observation de plusieurs stades phénologiques particuliers dans le même arbre, ce que l'on appelle asynchronisme phénologique (Fig. 16).

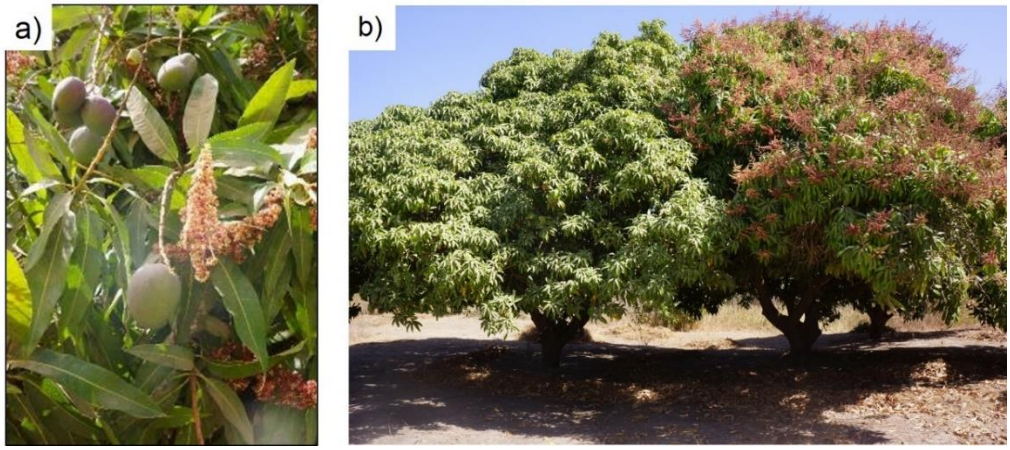


Fig. 16. Phénomènes (a) d'asynchronisme et (b) d'alternance sur le manguier. L'apparition successive des inflorescences (flushes) conduisent à des asynchronismes de production (source photo : J. Sarron).

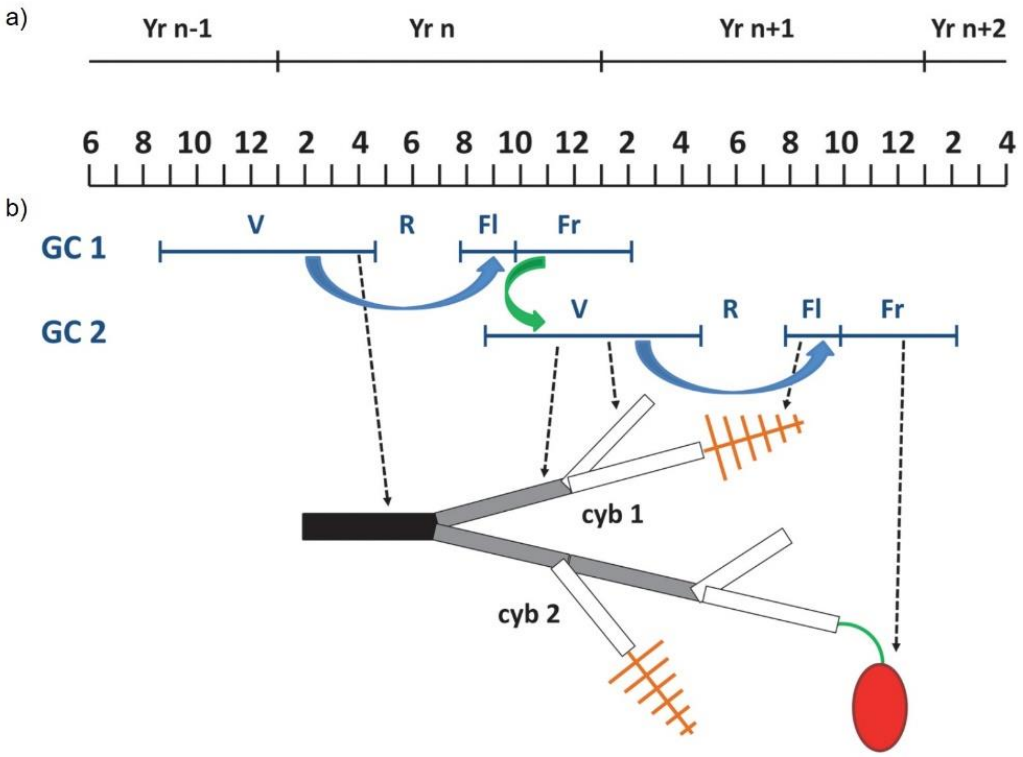


Fig. **17.** (a) cycle de croissance de la variété 'Cogshall' (île de la Réunion, hémisphère sud) et (b) succession des unités de croissance durant ces deux cycles. Les stades phénologiques sont : la croissance végétative (V), la période de repos (R), la floraison (Fl) et la fructification (Fr). Les flèches bleues montrent les interactions intra-cycle entre les phases V et Fl, les flèches vertes indiquent les interactions entre la phase Fl et la phase V du cycle suivant (source : Normand et Lauri 2018).

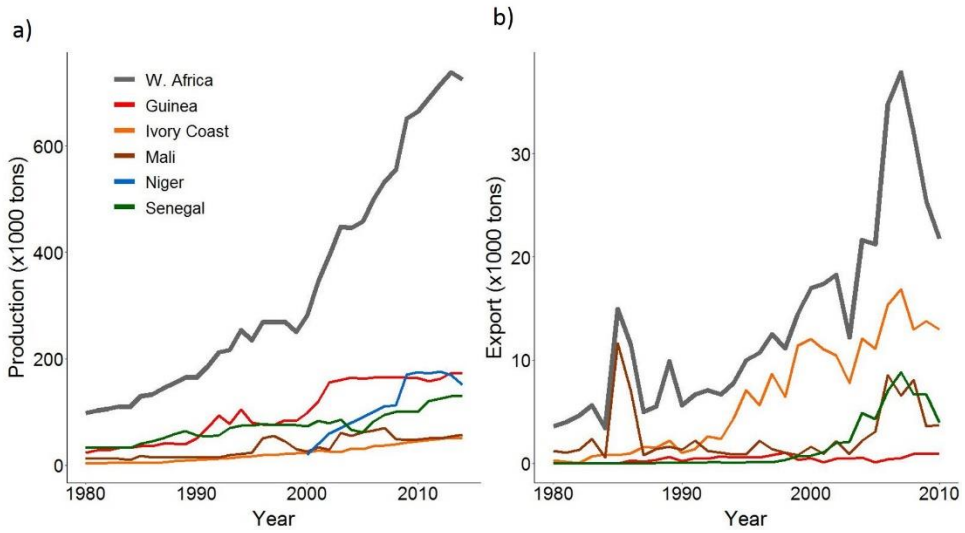


Fig. 18. Evolution (a) de la production et (b) des exportations de mangues dans les principaux pays producteurs d'Afrique de l'Ouest (hors Nigeria) (source : FAOSTAT 2017).

III.4. La filière mangue au Sénégal

a. Chiffres et bassins de production

Le Sénégal (Fig. 1) est l'un des pays qui, parmi les 16 pays d'Afrique de l'Ouest, a observé la plus forte croissance économique cette dernière décennie. Avec un territoire de 196 000 km² et une population de 16 millions d'habitants, le Sénégal est le pays situé le plus à l'ouest du continent et se trouve sous l'influence de plusieurs zone agri-climatiques dont le Sahel (Fig. 4). Le Sénégal fait partie des pays d'Afrique de l'Ouest où la production de mangues suit une très forte augmentation passant de 32 000 t en 1980 à 131 000 t en 2014 (+400 %, Fig. 18a). La majeure partie de la production est consommée en frais et il y a peu de transformation (Rey and Dia 2010). En 2017, on estimait que 14 000 t de mangues était exportées soit environ 10% de la production totale (ANSD 2017). À l'échelle de la région, le Sénégal est le 4ème producteur de mangues en Afrique de l'Ouest derrière le Nigeria, le Mali et la Guinée mais il est le 2ème exportateur (soit 18% de l'exportation régionale) derrière la Côte d'Ivoire (Fig. 18b). L'exportation de mangues est relativement récente : la quantité exportée s'élevait à seulement 730 tonnes en 2000 (FAOSTAT 2017). Les exportateurs, majoritairement constitués de sociétés privées, ne représentent qu'un petit nombre d'exploitations : 83 % de l'exportation de mangues au Sénégal était assuré par 5 exportateurs en 2006, et une société assurait à elle seule 50% de l'approvisionnement (Rey and Dia 2010). La production du reste des vergers (~80% en termes de surface) est destinée au marché local et régional.

Les principaux bassins de production de la mangue se situent en Casamance, dans le delta du Sine-Saloum et dans les Niayes (Fig. 19). Bien qu'il soit difficile d'évaluer les productions, Rey et Dia (2010) ont estimé

que la Casamance représentait 57% de la production, suivie par la région des Niayes (37%). Cependant l'essentiel de l'exportation (70 à 80%) est assurée par la région des Niayes qui possède l'avantage d'être située à proximité des axes routiers et du port de Dakar (Ternoy et al. 2006). La situation climatique est favorable dans les Niayes et permet de produire des fruits plus tardivement dans l'année. Ceci confère au Sénégal un avantage compétitif sur les marchés régionaux et internationaux. La zone des Niayes, dans laquelle se situe notre étude, est présentée dans le Matériels et Méthodes de ce mémoire.

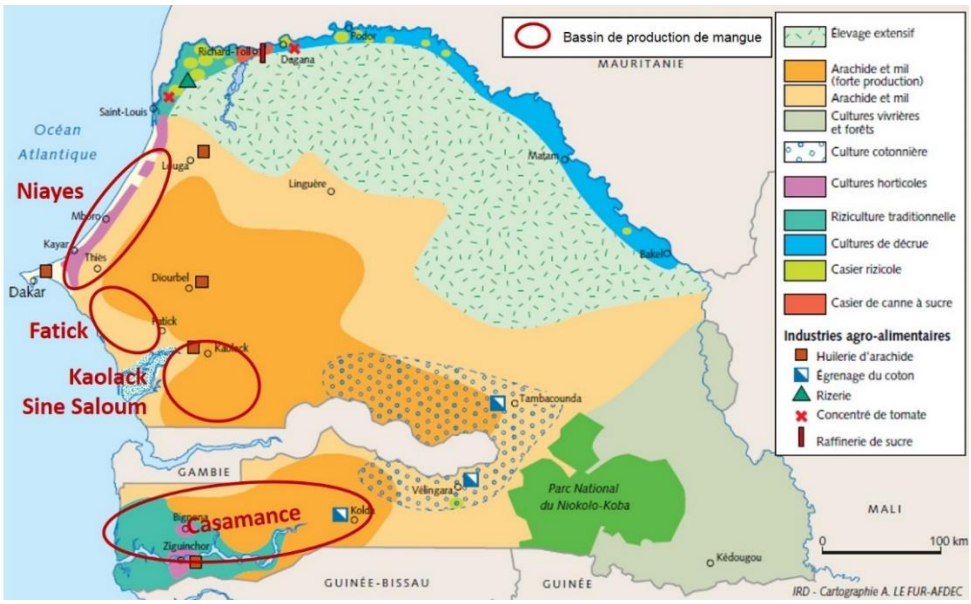


Fig. 19. Principaux bassins de production agricoles du Sénégal (source : IRD).

b. Variétés cultivées et contraintes de la filière

Il existe de nombreuses variétés de mangues au Sénégal (Fig. 15, Rey et al. 2007). Les premières mangues destinées à l'exportation étaient de variété 'Amélie'. Depuis les années 2000, on assiste cependant à une transformation des vergers. En effet, l'exportation a nettement augmenté du fait de la forte demande en variétés floridiennes ('Kent' et 'Keitt'), plus avantageuses sur le marché d'export (peu de fibres, bonne tenue durant le transport, coloration, gros calibre), et le surgreffage des vergers. Aujourd'hui les variétés les plus cultivées sont donc les variétés 'Kent' et 'Keitt' dans les monocultures mais également dans les petits vergers. Il existe cependant d'autres variétés de mangues dites 'locales' qui sont soit consommées sur place ('Boucodiékhal', 'Papaye') soit utilisées comme porte greffe ('Séwé', 'Dieg Bou Gatt') car nanisantes et trop fibreuses. Selon les bassins et les variétés, la production s'étale d'avril (Casamance) à septembre pour les variétés tardives dans les Niayes (Fig. 20).

La filière mangue fait face aux problématiques du manque de données et des faibles rendements évoquées précédemment (Partie II.3). Outre ces contraintes communes à l'ensemble des productions fruitières, la

mangue au Sénégal doit faire face à des facteurs limitants d'ordre environnemental et biotique. Tout d'abord le pays, tout comme le reste de l'Afrique de l'Ouest, fait face au problème de la mouche des fruits qui occasionnent des dégâts sur la production (jusqu'à 70%, Vayssières et al. 2008). La mouche et plus particulièrement *Bactrocera Invadens*, introduite en 2004, pose des problèmes pour l'exportation. L'insecte est en effet soumis à quarantaine en Europe (principale région importatrice de mangues ouest-africaines) et sa détection provoque la destruction de la totalité de la cargaison (Rey and Dia 2010). La filière a beaucoup de mal à gérer ce ravageur du fait qu'elle n'est pas vraiment structurée. Il arrive que des conteneurs soient envoyés à l'export sans contrôle de la qualité, pénalisant ainsi tout le reste des producteurs-exportateurs en cas de détection de la mouche des fruits une fois en Europe.

D'autres ravageurs sont observés comme les termites qui causent des dépérissements sur les arbres, en particulier en situation de stress hydrique (Sane 2016). L'anthracnose et d'autres maladies fongiques provoquent également des dégâts sur la qualité du fruit et sont souvent mal maîtrisées.

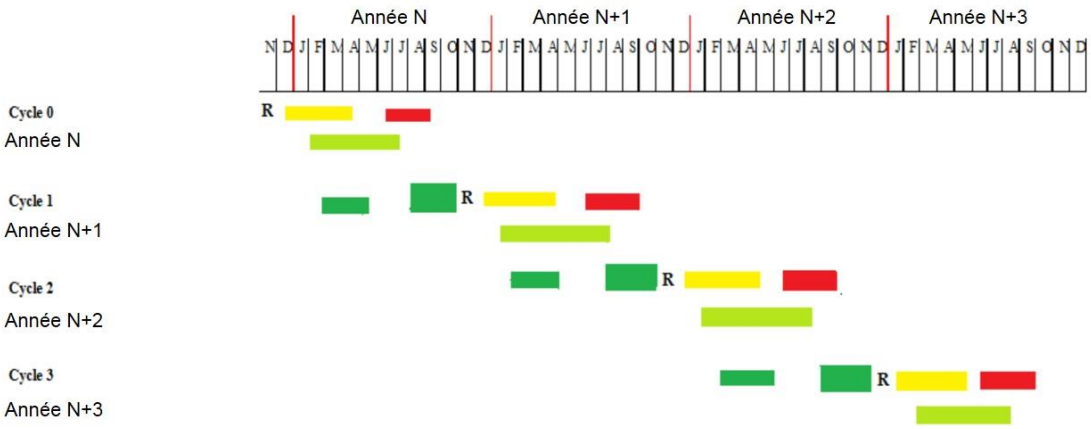


Fig. 20. Cycle phénologique de la variété 'Kent' dans la zone des Niayes au Sénégal. Les rectangles verts foncés indiquent les périodes de croissance végétative, les jaunes celles de la floraison, les verts clairs celles de la fructification et les rouges celles de la récolte (source : Diatta 2018).

Objectifs de la thèse

L'agriculture ouest-africaine doit donc faire face à de nombreux défis comme la croissance démographique, la dépendance aux importations et le changement climatique, avec des perspectives de diminution des rendements (van Ittersum et al. 2016). Le Sahel est une région où ces contraintes sont sans doute les plus fortes et exercent des pressions sur l'environnement, notamment la ressource en eau (Aguiar et al. 2010). Ces facteurs impactent aussi la sécurité alimentaire des petits exploitants agricoles, à travers les rendements des cultures notamment les arbres fruitiers (Mathur et al. 2012). Pour garantir un apport de nourriture en quantité et en qualité, et ce de manière durable pour l'environnement, l'agriculture peut s'engager dans plusieurs voies comme l'agroécologie et l'agroforesterie déjà pratiquées à petite échelle en Afrique de l'Ouest (Jagoret et al. 2011 ; Félix et al. 2018).

Dans ce contexte, les systèmes de culture à base de fruitiers offrent de nombreux avantages : une contribution à la sécurité alimentaire (autosuffisance et santé), la diversification des revenus, la résilience des systèmes et les services écosystémiques rendus par les arbres (Malézieux et al. 2009). Cependant, les filières fruitières en Afrique de l'Ouest pâtissent de faibles rendements et d'un manque de données décrivant les systèmes de culture. Les informations sur les pratiques culturales, la structure des vergers (composition spécifiques et configuration) sont souvent inexistantes (Carletto et al. 2015). Quant aux rendements, leur mesure est souvent imprécise à l'échelle du pays et ponctuelle voire inexistante aux échelles inférieures. Pourtant la donnée de rendement reste primordiale à la fois pour les producteurs et les filières mais également pour la recherche. Sans données fiables, la conception et l'évaluation de systèmes de culture plus performants économiquement et écologiquement tourne au ralenti.

Le cas du manguier au Sénégal est assez emblématique de cette situation. La culture de mangues existe depuis de nombreuses années au Sénégal, comme en Afrique de l'Ouest. Dans la région, il existe une diversité des systèmes de culture même dans un seul bassin de production (Grechi et al. 2013). Cette diversité peut induire des écarts de rendement entre les types de systèmes qui peut se traduire par une forte variabilité spatio-temporelle de productivité à l'échelle du bassin de production. S'ajoute à cela une variabilité spatiale de la production à l'intérieur des vergers, principalement dans les vergers complexes mais que l'on peut retrouver également dans les vergers intensifs. Ces variabilités intra et inter-vergers n'ont jamais été mesurées dans le cas de la mangue au Sénégal. Pourtant leur connaissance serait nécessaire à la fois pour la filière (producteur et autres acteurs) mais aussi pour la recherche. De plus, les effets des facteurs causant la variabilité des rendements du manguier aux différentes échelles (arbre, verger et bassin de production) ne sont pas connus avec précision dans le contexte ouest-africain : la faible qualité des données empêche toute étude précise des facteurs d'hétérogénéités et de variabilités intra- et inter-verger.

Ces dernières années, le développement d'outils permettant l'acquisition massive de données notamment par télédétection (drone, satellite etc.) et les algorithmes de traitements de ces données (réseau de neurones) apportent de nombreux avantages pour l'agriculture (Usha and Singh 2013 ; Kamilaris and Prenafeta-Boldú 2018). Or, les méthodes d'estimation et de spatialisation des cultures pérennes ne sont, à l'heure actuelle, pas adaptées au contexte de l'agriculture ouest-africaine, plus particulièrement chez les petits producteurs. L'utilisation d'outils frugaux offre des opportunités mais des adaptations reste nécessaire pour les arbres fruitiers en particulier dans des systèmes de culture complexes.

L'estimation et la spatialisation des rendements des cultures pérennes constituent donc un enjeu stratégique pour la sécurité alimentaire et pour l'adaptation de ces cultures aux changements du milieu. Or, les méthodes d'estimation et de spatialisation des cultures pérennes ne sont, à l'heure actuelle, pas adaptées au contexte de l'agriculture ouest-africaine, plus particulièrement chez les petits producteurs. Ce manque est constaté de l'échelle de l'arbre à celle du bassin de production, voire du pays. Les objectifs de cette thèse sont donc :

- **de développer des méthodes fiables d'estimation de la productivité des cultures pérennes en Afrique de l'Ouest : de la production de l'arbre aux rendements dans le bassin de production ;**

- **d'imbriquer ces différentes échelles pour extrapoler la production de l'arbre vers le rendement du verger puis de celui du verger vers la production du bassin ;**

- **d'analyser et comprendre les facteurs de variabilité de la production aux différentes échelles.**

Le chapitre suivant permet d'avoir un aperçu de l'état de l'art sur les facteurs impactant la production de l'arbre, unité de base du verger. Ce chapitre, apporte ensuite des éléments de compréhension sur le rendement à l'échelle du verger et sur sa variabilité spatiale. Enfin, nous verrons quelles sont actuellement les méthodes disponibles pour l'estimation de la production de l'arbre et du rendement du verger. Cet état de l'art permet de poser les définitions du rendement à chaque échelle et de décrire les outils qui seront utilisés tout au long du mémoire (voir également le Matériels et Méthodes).

ÉTAT DE L'ART

Caractérisation et rendement des arbres fruitiers

I. Facteurs déterminant la production de l'arbre

I.1. Architecture de l'arbre et interception lumineuse

L'arbre fruitier constitue l'unité qui compose le verger. Contrairement aux cultures annuelles, sa productivité est dépendante de son architecture et donc des précédents cycles de production (âge). Dans sa définition la plus précise, la production de l'arbre se définit comme le poids total de fruits produits durant un cycle de production (exprimé en kg). Or, la fructification, c'est-à-dire le développement du fruit, nécessite l'apport d'éléments carbonés (glucides) majoritairement fabriqués par la photosynthèse des feuilles (Grossman and DeJong 1994). L'architecture de l'arbre - en déterminant la morphologie, le nombre et l'organisation des différents organes (feuilles, pousses, racines, etc.), dont les branches qui constituent le houppier - joue un rôle capital dans l'interception de la lumière et donc la production d'éléments carbonés. L'architecture correspond à une définition botanique et physiologique de la forme et de l'agencement dans l'espace des composantes de l'arbre (i.e., le modèle architectural défini par Hallé : Hallé et al. 1978 ; Costes et al. 2010). Le modèle architectural varie en fonction de l'espèce et du cultivar (Fig. 21) : les traits génétiques contrôlent en partie la morphologie (position des méristèmes, direction de la croissance, longueur et port des branches, etc.) et la temporalité des stades phénologiques (Lauri and Trottier 2004 ; Dambreville et al. 2013a). De plus, les facteurs génétiques impactent directement la production de biomasse via le contrôle de l'efficience de la photosynthèse et des dimensions (largeur, longueur et épaisseur) des feuilles, paramètres qui varient selon les espèces et les variétés.

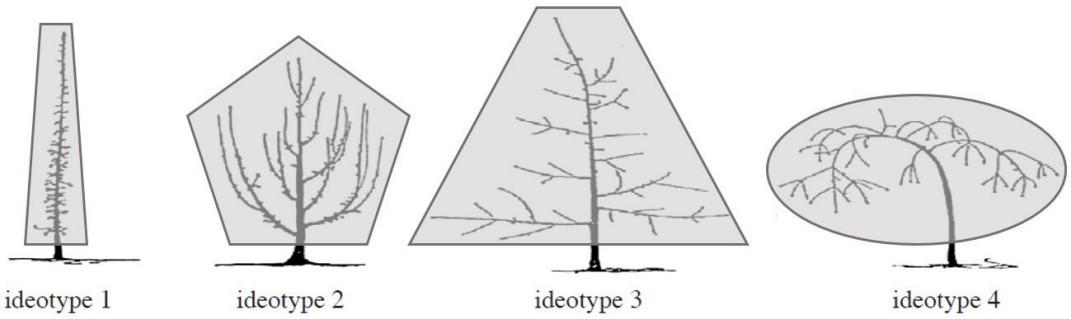


Fig. 21. Variabilité interspécifique (idéotypes) de l'architecture du pommier. Les formes grises représentent la structure approximée de chaque idéotype (source : Costes et al. 2010).

De nombreux facteurs environnementaux conditionnent l'architecture de l'arbre et son expression génétique. En effet, les facteurs génétiques contrôlant le partitionnement de la biomasse dans l'arbre varient sous l'influence de l'âge, du microclimat et de la disponibilité en ressources (variation phénotypique, Coleman et al. 1994). Parmi les facteurs environnementaux, la température et la disponibilité en eau restent les plus importants dans le développement des plantes. Ces facteurs conditionnent l'activité photosynthétique, la croissance végétative et la production de biomasse (Schaffer et al. 2009). Outre ces

facteurs, la variation des radiations solaires ou l'environnement lumineux, la disponibilité en nutriments et le CO_2 impactent la croissance foliaire et/ou l'efficience de la photosynthèse (Grossman and DeJong 1994 ; Stephan et al. 2008).

L'architecture du houppier, issue d'un équilibre entre des processus endogènes et des contraintes exogènes, est donc le principal moteur de la production de l'arbre en permettant l'interception de la lumière (Stephan et al. 2008). Chez la plupart des espèces, la production de l'arbre est nulle les premières années après la plantation. Puis, la production augmente, avec le développement du houppier et l'augmentation de la surface foliaire et de l'interception lumineuse, et peut atteindre un palier ou non selon les conditions de production (densité de plantation, pratiques de taille) et les espèces (Fig. 22a, Menzel and Le Lagadec 2014, 2017). Ainsi la production de fruits est bien souvent positivement corrélée au volume du houppier, à la surface foliaire et à l'interception lumineuse (Fig. 22b).

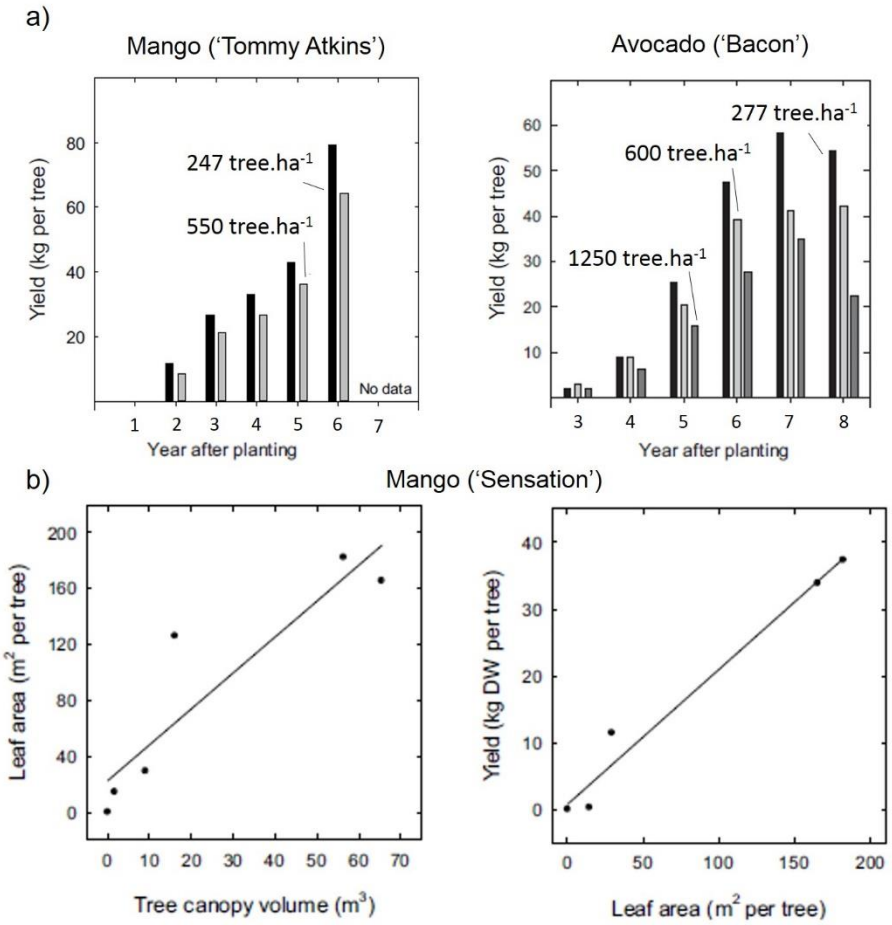


Fig. 22. Relation entre la production de l'arbre et sa structure : (a) effet de l'âge et de la densité de plantation sur la production du manguier et de l'avocatier ; (b) relation entre la surface foliaire, le volume du houppier et la production pour le manguier (source : Menzel and Le Lagadec 2014, 2017).

Dans ce mémoire nous utiliserons plutôt le terme de structure, définition parfois similaire à l'architecture mais qui correspond à l'ensemble des variables morphologiques décrivant l'arbre (sa hauteur, la surface et le volume du houppier, le diamètre du tronc, etc.). La structure permet de décrire l'arbre dans son ensemble, sans descendre à des échelles plus fines (branche, pousse, est.) et regroupe donc des variables facilement mesurables, à la main ou à l'aide d'outils adaptés, pour caractériser la morphologie de l'arbre (Fig. 21).

I.2. Phase reproductive et établissement du fruit

Le nombre total de fruits et leurs caractéristiques individuelles (dimension, poids, qualité et couleur) s'établissent au cours du cycle de reproduction qui regroupe plusieurs stades successifs aboutissant à la formation et la maturation des fruits. Généralement, la phase reproductive comprend quatre stades principaux qui sont, dans l'ordre chronologique d'apparition : 1) l'initiation de la pousse ; 2) la floraison 3) le développement du fruit ; 4) et la maturation du fruit (ex pour le manguier : Hernández Delgado et al. 2011). A chaque stade, la production de l'arbre peut être caractérisée par différentes composantes : le nombre de fleurs/inflorescences (à la floraison), le nombre de nœuds et le nombre de fruits par pousse/inflorescence (au stade de développement du fruit), le nombre total de fruits (à la maturation du fruit) et leurs caractéristiques (ex : le poids du fruit).

a. Effets de la structure et des paramètres environnementaux

De nombreuses études ont mis en évidence l'importance de l'architecture de l'arbre sur les composantes de la production telles que le nombre de fruits produits, leur taille et leur qualité (Normand et al. 2009 ; Costes et al. 2010 ; Léchaudel et al. 2013 ; Dambreville et al. 2013a). À l'échelle de la branche, la structure et le nombre de feuilles portées par la branche ont un impact direct sur la floraison (Ramírez et al. 2010), le nombre de fruits produits et leur qualité (Léchaudel et al. 2002 ; Lauri and Trottier 2004). À l'échelle de l'arbre, et comme vu plus haut, la structure du houppier détermine directement le potentiel de production (nombre de sites reproducteurs) et donc le nombre de fruits (Fig. 22, Menzel and Le Lagadec 2014, 2017).

En plus de ces facteurs endogènes, plusieurs facteurs exogènes interagissent et déterminent et/ou impactent les composantes du rendement au sein d'un même arbre (Ramírez and Davenport 2010). Tout d'abord, l'induction de la floraison est contrôlée par la température chez la plupart des espèces surtout dans les milieux tempérés (Borchert 1991). En plus de jouer sur l'initiation florale, la température ainsi que l'eau sont connus pour intervenir dans le nombre de fleurs et leur taille (Ramírez and Davenport 2010). Chez plusieurs espèces, comme le pêcher ou manguier, le développement et la physiologie du fruit sont également induits par la température et la ressource en eau (Grossman and DeJong 1994 ; Nordey et al. 2014). Au cours de son développement, le fruit va grossir notamment en absorbant de l'eau et des matières carbonées,

cette absorption se fait via la transpiration qui est impactée par de nombreux paramètres : l'épaisseur du fruit, l'humidité extérieure et la température.

L'arrangement spatial des branches et le houppier, plus ou moins poreux à la lumière, modifient le microclimat de l'arbre (Lauri et al. 2004 ; Stephan et al. 2008 ; Léchaudel et al. 2013). En conséquence, la température, l'interception lumineuse et l'humidité varient spatialement dans le houppier (Fig. 23). Par exemple, les feuilles externes captent mieux la lumière que les feuilles internes à cause des effets d'ombrage où l'activité photosynthétique diminue (Stephan et al. 2008). La variabilité spatiale du microclimat dans l'arbre crée alors des variabilités dans la production et la qualité des fruits qui s'observent au sein d'un même arbre et à l'échelle de la canopée (Connor et al. 2014).

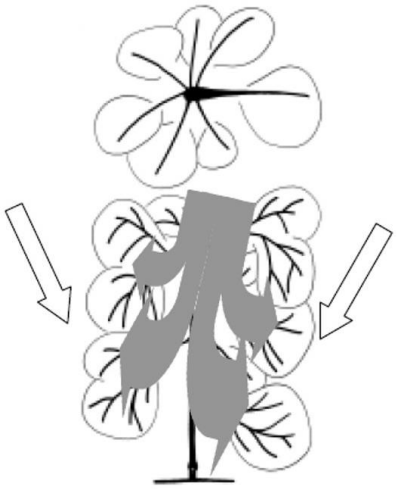


Fig. 23. Distribution de la lumière dans le houppier du pommier (taille de formation centrifuge). La distribution du microclimat (lumière, température et humidité) est impactée par l'architecture de l'arbre. La taille participe au contrôle de ce microclimat (source : Lauri et al. 2004).

b. Phénomènes d'alternance et d'asynchronisme

Lors du développement de l'arbre la succession temporelle des cycles phénologiques végétatifs et reproductifs s'ajoute à des facteurs structurels et environnementaux (température et luminosité) et déterminent le devenir des méristèmes (Dambreville et al. 2013a). Chez le manguier, selon les conditions environnementales, les précédents cycles et la structure de la pousse (longueur, nombre de feuille, position du méristème) le méristème va se différencier en pousse végétative ou en pousse reproductive (Capelli et al. 2016). Chez le pommier, la structure de la pousse (longueur et nombre de feuilles portées) détermine la différenciation du bourgeon terminal en pousse végétative ou en inflorescence (Lauri and Trottier 2004). Cette particularité conduit à l'alternance de la production : un arbre peut produire une année (année 'ON') et avoir une baisse de la production l'année suivante (année 'OFF'). Le profil d'alternance d'un arbre (i.e.,

la succession des années ON et OFF) varie en fonction de plusieurs paramètres dont l'espèce et la variété, le climat sous lequel il évolue et les pratiques culturales (Monselise and Goldschmidt 1982 ; Dambreville et al. 2014).

L'asynchronisme constitue également une contrainte pour la production des arbres fruitiers. La température impacte la phénologie de l'arbre c'est-à-dire l'initiation et la durée de croissance des organes (Borchert 1991). Dans les zones tropicales et subtropicales (voir Contexte et problématiques – Partie III), le rôle de la température est complexe et ce facteur interagit avec des paramètres intrinsèques à l'arbre (Dambreville et al. 2013a). Cette particularité définit le caractère asynchrone des espèces pérennes dans les régions tropicales : la croissance et le développement de la branche sont indépendants du reste de l'arbre, conduisant à l'apparition des nouvelles pousses par flushs successifs. L'asynchronisme phénologique, observé en zone subtropicale et tropicale, augmente les phénomènes d'hétérogénéité de production à l'échelle de l'arbre (voir Fig. 16 et 17).

I.3. Les pratiques culturales

Pour certaines espèces, le choix du matériel génétique en début de cycle est primordial car, sauf en cas de regreffage ou de sur-greffage, il va impacter la production de l'arbre durant toute sa période productive (Lu et al. 2013 ; Connor et al. 2014). En arboriculture, il est commun d'utiliser un porte-greffe (Reddy et al. 2003 ; Menzel and Le Lagadec 2014). Son effet sur la structure du houppier n'est pas négligeable et il est courant d'utiliser des variétés nanisantes dans les vergers à haute densité afin de réguler la taille du houppier (Reddy et al. 2003). En arboriculture fruitière, les pratiques culturales regroupent les interventions humaines visant à augmenter le nombre et la qualité de fruits produits et à réduire les phénomènes d'alternance et d'asynchronisme. On peut diviser les pratiques en deux catégories : i) les pratiques de taille ; et ii) l'apport d'intrants.

La taille vise principalement à améliorer l'interception lumineuse dans le houppier et il en existe plusieurs types (Costes et al. 2010). Dès la plantation, le producteur applique une taille de formation en manipulant la croissance de l'arbre (retrait de bourgeons ou de pousses), pour modeler son architecture et encourager le développement végétatif les premières années (Lauri et al. 2004). Après quelques années, le producteur manipule le houppier annuellement pour maintenir son volume et sa production en retirant par exemple les branches les plus vieilles ou celles les moins exposées au soleil (Connor et al. 2014 ; Normand and Lauri 2018). La taille a également pour but de réduire les phénomènes d'alternance et d'asynchronisme chez les espèces tropicales (Ramírez and Davenport 2010). Chez le manguier, une taille annuelle est souvent pratiquée, après le cycle de production, pour réduire les variabilités de production intra et inter-arbre (Crane et al. 2009).

L'apport d'intrants permet de pallier les stress environnementaux et biotiques. L'irrigation reste primordiale pour certaines productions notamment dans les zones tropicales et semi-arides à cause des précipitations irrégulières. Chez le manguier, un déficit en eau peut conduire à une réduction du nombre de fleurs, à une chute physiologique des fruits et à une réduction de la taille des fruits (Spreer et al. 2009).

En plus des facteurs abiotiques liés au climat, de nombreux facteurs biotiques, maladies et ravageurs, impactent ou réduisent les composantes du rendement. Contrairement aux facteurs environnementaux, il est plus difficile de généraliser sur les maladies et les ravageurs car certains sont spécifiques et leurs abondances et effets dépendent de nombreux paramètres comme le climat ou les pratiques culturales (ex : Sané. 2016).

II. De la production de l'arbre au rendement du verger

La production individuelle d'un arbre varie à la fois en fonction de facteurs endogènes et exogènes. Cette hétérogénéité entre les arbres impacte le rendement du verger. On définit le rendement brut du verger comme la somme de la production de tous les arbres qui le composent :

$$Y = \frac{\sum_{j=1}^{N} P_j}{A} \times 10^3 \quad (1)$$

où : Y est le rendement du verger (en t.ha^{-1}) ; N est le nombre d'arbres dans le verger ; P_j est la production de l'arbre j en kg.; et A est la surface productive du verger en ha.

La définition du rendement peut varier selon les objectifs. On peut ainsi s'intéresser au rendement brut ou récolté, c'est-à-dire au poids de fruits récoltés par le producteur (qui peut laisser les fruits de petits calibres ou immatures dans l'arbre) dans un verger, ou encore au rendement commercialisé, c'est-à-dire correspondant aux fruits qui seront vendus en prenant en compte leur calibre et leur qualité. Dans la suite de cette thèse nous aborderons uniquement le rendement brut (Eq. 1).

II.1. Structure des vergers

Nous définissons, dans ce travail, la structure du verger comme l'ensemble des variables définissant la composition spécifique et variétale et l'arrangement spatial des arbres dans le verger. Dans les systèmes monoculturaux, les arbres ont majoritairement le même âge et sont également de la même variété (au moins à l'échelle de la parcelle). Le rendement du verger est donc principalement impacté par la densité de plantation, c'est-à-dire le nombre de pieds planté à l'hectare. Par exemple, chez le manguier, jusqu'à un certain niveau, la densité de plantation impacte positivement le volume de canopée (= somme des houppiers)

du verger et donc l'interception lumineuse et le nombre de fruits produits (Fig. 24, Menzel and Le Lagadec 2017). Cependant, les nombreuses études – par exemple sur manguier, olivier et avocatier - sur l'impact de la densité sur la production des arbres reflètent que cette variable doit être optimisée pour obtenir le meilleur rendement à l'échelle du verger (Connor et al. 2014 ; Menzel and Le Lagadec 2014, 2017). En effet, une densité trop élevée peut augmenter la compétition entre les arbres pour la lumière (Fig. 25). Par exemple dans les vergers de manguiers, le doublement de la densité, de 800 à 1 600 arbres.ha^{-1}, augmente très peu le rendement du fait de la réduction de la production individuelle des arbres (Fig. 24). Dans ce cas précis, le doublement de la densité peut paraitre contre-productif au regard des coûts engendrés pour l'implantation et le maintien des arbres (Menzel and Le Lagadec 2017). De manière générale, dans les vergers à haute densité, pour équilibrer entre la densité de plantation et la compétition entre les arbres, les producteurs maintiennent le volume de la canopée en appliquant une taille régulière ou en utilisant des porte-greffes adaptés (Connor et al. 2014 ; Menzel and Le Lagadec 2014). La configuration (ou arrangement) spatiale des arbres est également importante. Dans les monocultures, les arbres sont disposés en rangs et leur orientation peut avoir un effet sur l'interception lumineuse totale de la canopée et donc sur le rendement. Par exemple, en comparant des rangs d'oliviers orientés N-S avec des rangs E-O, Gómez-del-Campo al. (2009) ont démontré un effet de l'orientation du rang sur le rendement et sur la distribution spatiale de la production dans la canopée (ex : face Sud plus productive que la face Nord dans un rang orienté E-O).

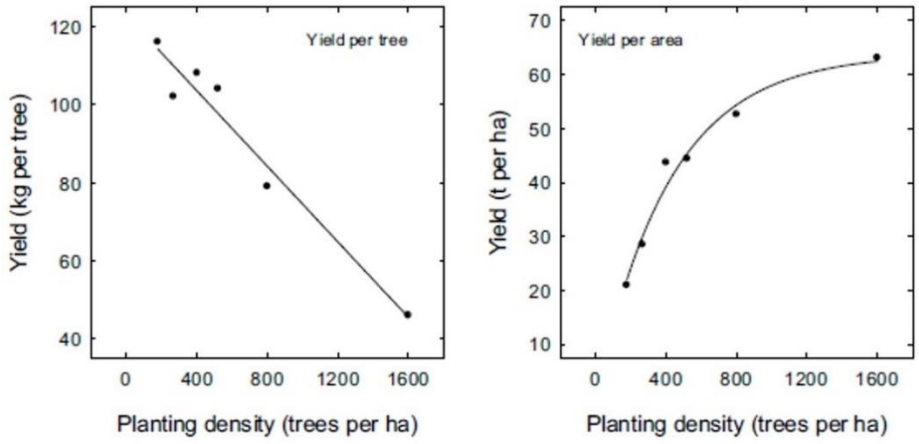


Fig. 24. Effet de la densité de plantation sur la production de l'arbre et sur le rendement dans les monocultures de manguiers. À partir d'une certaine densité, la compétition entre les arbres pénalise le rendement (source : Menzel and Le Lagadec 2017).

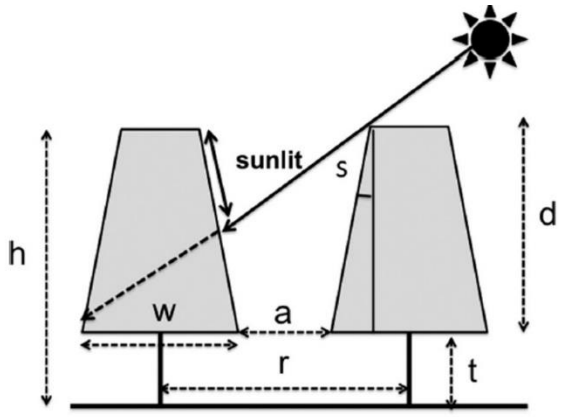


Fig. 25. Schéma de la compétition pour la lumière entre deux rangs dans le verger. La canopée a une profondeur (d), une pente (s) et une largeur à la base (w). La hauteur du rang est (h) égal à d + t avec t la longueur aérienne du tronc. Les deux rangs sont séparés par une distance (r) donnant une largeur libre (a = r-w) (source : Connor et al. 2014).

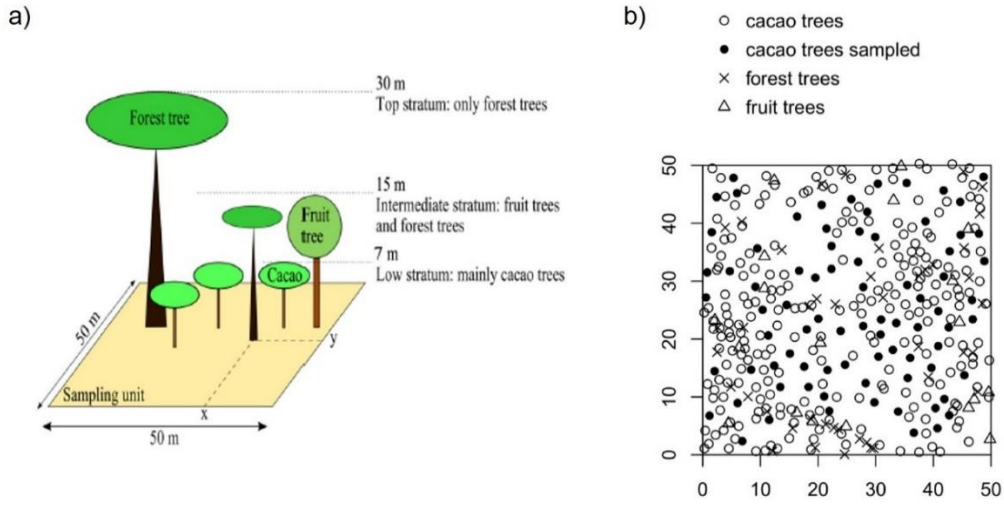


Fig. 26. Complexité et hétérogénéité des systèmes agroforestiers : (a) représentation schématique de la structure verticale et (b) distribution spatiale des espèces dans un système agroforestier à base de cacaoyers au Cameroun (source : Gidoin et al. 2014).

Dans le cas des vergers complexes (agroforesterie) il y a plusieurs espèces présentes dans la parcelle. La structure de ces vergers peut être optimisée comme c'est le cas de certains systèmes en Europe où les arbres sont disposés de manière linéaire et la densité est optimisée (ex : système en allée, Dupraz 1999). Mais dans la majorité des cas et en Afrique de l'Ouest, les vergers agroforestiers évoluent dans des structures complexes où la configuration des arbres apparaît aléatoire : l'arrangement spatial des arbres peut parfois montrer des zones où la canopée est plus ou moins agrégée (plus dense) en fonction de la variation de la densité de plantation, de l'arrangement spatial ou bien de la structure des arbres (Fig. 26 et Gidoin et al. 2014). En plus de la densité et de la configuration spatiale, la composition spécifique (nombre et proportions des espèces et des variétés cultivées, dominance de certaines espèces, etc.) entre également en jeu dans l'élaboration du rendement. Les espèces peuvent interagir positivement ou négativement (compétition)

selon les situations (Malézieux et al. 2009 ; Gidoin et al. 2014). Dans ces vergers complexes, la composition et la densité totale d'arbres (i.e., incluant toutes les espèces) est donc tout aussi importante que la densité de l'espèce d'intérêt dans la définition du rendement (Jagoret et al. 2017).

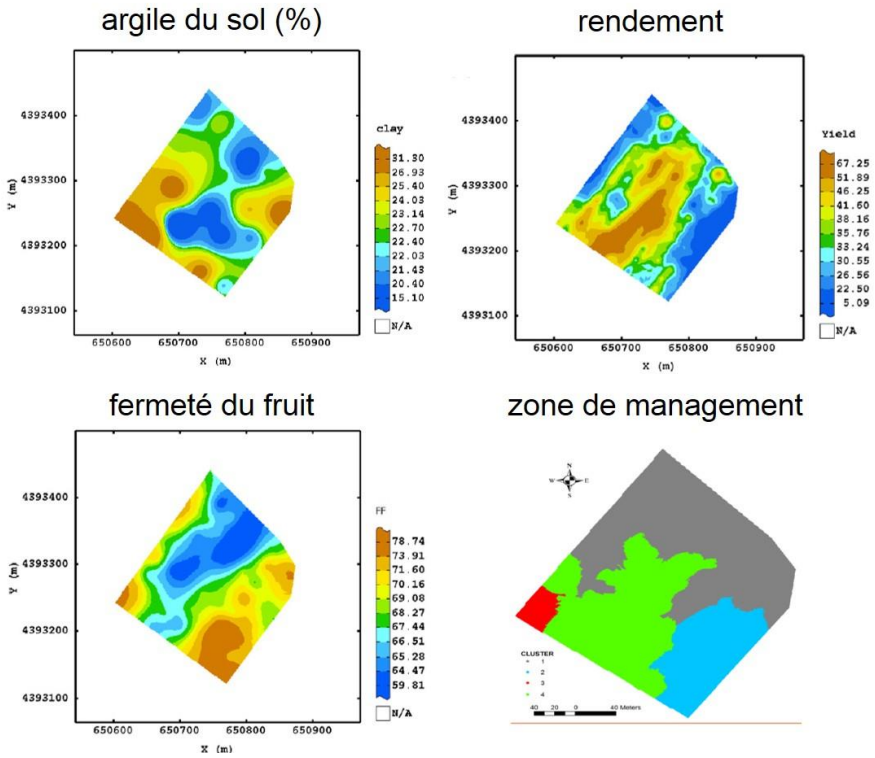


Fig. 27. Variation spatiale des propriétés du sol (contenu en argile), du rendement et de la qualité du fruit (fermeté) dans un verger de pommiers. Ces variables peuvent être utilisée pour délimiter des zones homogènes dans le cadre de l'agriculture de précision (source : Aggelopoulou et al. 2013).

II.2. Hétérogénéité spatiale du rendement

À l'échelle du verger, des hétérogénéités dans le développement des arbres peuvent être causées par les variations spatio-temporelles des caractéristiques du sol, de la topographie, du climat et des interactions intra et inter spécifiques entre les arbres (Perry et al. 2010 ; Aggelopoulou et al. 2013). Ces hétérogénéités et les phénomènes d'alternance induisent une variation spatio-temporelle de la production et de la qualité des fruits au sein du verger (Aggelopoulou et al. 2010 ; Noble et al. 2018). La présence d'hétérogénéité spatiale a pu être mise en évidence (Fig. 27) dans les vergers monoculturaux, comme c'est le cas sur pommier (Aggelopoulou et al. 2010), poirier (Perry et al. 2010) ou agrume (Mann et al. 2011).

Ces variabilités de l'environnement se surajoutent à celle des pratiques qui ne sont pas homogènes à l'échelle du verger, surtout chez les petits producteurs. La gestion des vergers des petits producteurs évolue

dans le temps en fonction du développement des arbres et les vergers sont souvent gérés par zone voire à l'échelle de l'arbre (replantation, regreffage, taille sélective, etc.) (Jagoret et al. 2011). Ainsi, dans les vergers agroforestiers, on observe une plus grande variabilité dans l'âge et la structure des arbres (Fig. 26, Jagoret et al. 2018). Cependant, à notre connaissance, la distribution spatiale du rendement n'a jamais été étudiée dans ce type de verger.

L'agriculture de précision est une approche permettant de contrôler ou de prendre en compte les hétérogénéités des parcelles (Whelan and McBratney 2000). Cette approche se développe progressivement dans les monocultures où l'objectif est de délimiter des zones homogènes de traitements en fonction des hétérogénéités spatiales pour ensuite appliquer les bonnes doses d'intrants en fonction des besoins (Mann et al. 2011 ; Aggelopoulou et al. 2013). Dans les petits vergers, la gestion des hétérogénéités se base sur l'observation et les connaissances du verger par le producteur (zone de fort rendement ou zone précoce, etc.).

II.3. Intégration des différentes échelles

La donnée de rendement est une donnée primordiale à la fois pour le producteur et la recherche. Le problème pour les cultures pérennes réside dans l'intégration des variabilités entre les arbres pour estimer le rendement à l'échelle du verger. L'équation de rendement Eq. 1 est la plus précise pour déterminer le rendement brut. Cependant, la production de tous les arbres est difficilement mesurable (voir Partie III). L'estimation du rendement peut également être faite en considérant une production moyenne :

$$\hat{Y} = \bar{P} \times d \times 10^3 \quad (2)$$

où $\hat{Y}$ est le rendement estimé (en t.ha^{-1}) ; $\bar{P}$ la production moyenne (en kg) et d la densité de plantation (en arbres.ha^{-1}).

En Afrique de l'Ouest, l'estimation de la production avant la récolte n'est réalisée que dans les vergers intensifs pour l'exportation qui disposent de moyens humains et matériels souvent importants pour cette estimation. L'estimation de $\bar{P}$ (Eq. 2) se fait sur la base d'un comptage visuel des fruits, sur un pourcentage d'arbres sélectionnés aléatoirement (autour de 5% des arbres sur des parcelles de 1 à 2 ha), suivi d'une extrapolation en multipliant la production par la densité de plantation d. Cette estimation est à la fois fastidieuse et peu fiable. Par exemple, l'extrapolation de la production des arbres à l'échelle du verger n'intègre pas la variabilité intra-verger.

Pourtant l'intégration des hétérogénéités intra-verger est importante notamment pour la mise en place de méthodes d'échantillonnages (Peeters et al. 2015 ; Anderson et al. 2019). En effet, l'estimation de $\bar{P}$ par un échantillonnage aléatoire d'arbres devient caduque si la distribution spatiale de la production des arbres

n'est pas aléatoire mais suit une tendance (dispersion ou agrégation). Ce problème est exacerbé dans les vergers complexes où les hétérogénéités semblent plus marquées et où la diversité variétale, qui induit une variabilité en production et en maturité (précocité), doit également être prise en compte dans l'estimation du rendement. À l'heure actuelle, très peu d'études se sont penché sur les moyens statistiques ou les modèles nécessaires pour extrapoler la production de l'échelle de l'arbre à l'échelle du verger (Anderson et al. 2019) et, *a fortiori*, aux échelles supérieures (bassin, région, pays).

III. Outils et méthodes pour la caractérisation de l'arbre et du verger

La modélisation est largement utilisée en grande culture pour estimer et étudier les rendements dans différentes conditions pédoclimatiques et de pratiques et de l'échelle parcellaire (Jeuffroy et al. 2014) à l'échelle globale (Burke and Lobell 2017). La situation est différente pour les cultures pérennes : à l'heure actuelle pour les arbres fruitiers les modèles existants se focalisent sur une fonction particulière (ex : interception lumineuse, Stephan et al. 2008) ou un organe (ex : le fruit : Nordey et al. 2014). Il n'existe donc pas de modèle disponible pour étudier la variabilité de rendements en fonction des pratiques et des conditions environnementales (Normand et al. 2015). Du fait de l'absence de modèle de culture pour l'estimation de la production et du rendement, les méthodes d'estimation empiriques du rendement ont pris une place importante en arboriculture. Ces dernières années l'amélioration des algorithmes d'analyse d'image a permis le développement d'outils décrivant par exemple, le nombre de fruits, la structure des arbres ou encore la structure des vergers (Usha and Singh 2013). Ces outils permettent grandement d'améliorer l'estimation empirique des rendements.

III.1. La détection des fruits

À l'échelle de l'arbre, plusieurs outils ont été développés pour faciliter la détection des fruits et l'estimation de la production par analyse d'image (Gongal et al. 2015). À l'heure actuelle, il existe différents systèmes de vision par ordinateur caractérisés par un ou plusieurs capteurs associés à un algorithme de traitement d'image. Les systèmes peuvent être très simples : par exemple, Payne et al. (2013) ont traité des images RVB (rouge-vert-bleu), prises avec un appareil numérique, avec des seuillages sur des indices de couleur et de forme pour détecter des mangues dans le houppier. D'autres systèmes de vision sont plus complexes et couplent des images de différentes sources avec des algorithmes de *machine learning* (i.e., algorithme d'apprentissage automatique) de type *k-nearest neighbours* KNN ou *support vector machine* SVM (Linker et al. 2012 ; Qureshi et al. 2016). Le *machine learning* se base sur l'apprentissage à partir de données d'entraînement afin de réaliser une action de classification, de modélisation, ou de partitionnement.

L'essor très récent (depuis 2012) des réseaux de neurones profonds, ou *deep learning* (Fig. 28), et le développement de machines de calcul plus performantes a permis d'améliorer encore ces systèmes et de rendre les outils plus rapides et plus fiables (Toulon 2018 ; Kamilaris and Prenafeta-Boldú 2018). Le *deep learning* est un assemblage de neurones artificiels qui s'inspirent du fonctionnement du neurone humain : le neurone combine plusieurs informations d'entrées, les pondèrent (affectation d'un poids à chaque information qui fait partie du paramétrage) et produit un message en sortie en utilisant une fonction d'activation (Fig. 28a). Le *deep learning* est donc composé de plusieurs couches de neurones (Fig. 28b) qu'il faut paramétrer (affectation des poids) avec les données d'entrainements. Plus les couches sont nombreuses et le réseau profond, plus la quantité de données nécessaire à l'apprentissage est importante. Les réseaux de neurones ont plusieurs applications parmi lesquelles l'analyse d'image par classification (i.e., détection d'objet et affectation d'une classe à cet objet). Les réseaux convolutifs (convolutional neural network, CNN) sont sans doute les réseaux les plus utilisés pour l'analyse d'image car ils permettent de traiter rapidement l'image par région et d'y détecter les objets d'intérêts (Fig. 29, Koirala et al. 2019b). Les réseaux montrent une plus grande efficacité par rapport aux algorithmes plus classiques (seuillage de couleur, KNN, SVM).

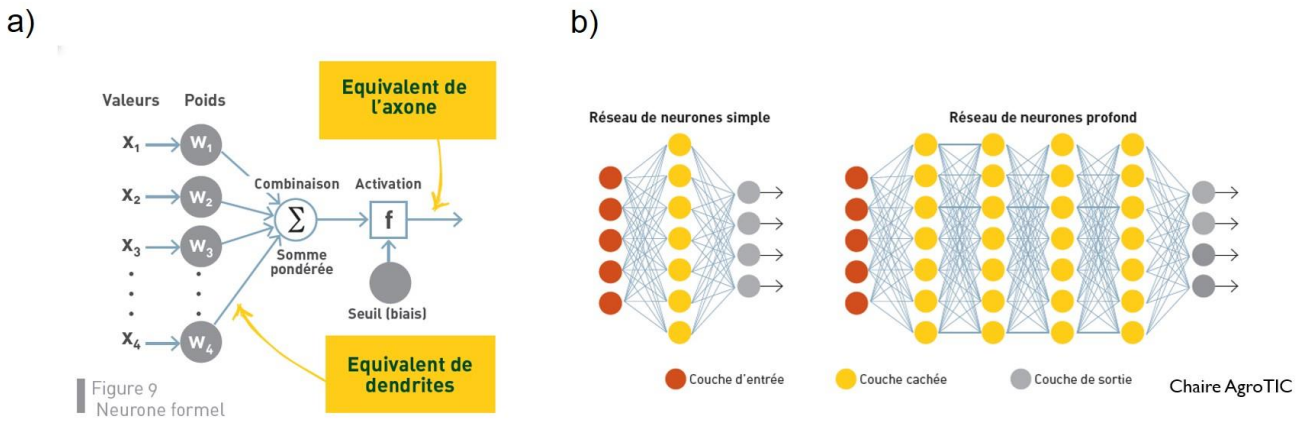


Fig. 28. Structure schématique (a) du neurone formel simple et (b) d'un réseau de neurone (source : Toulon 2018, Chaire AgroTIC).

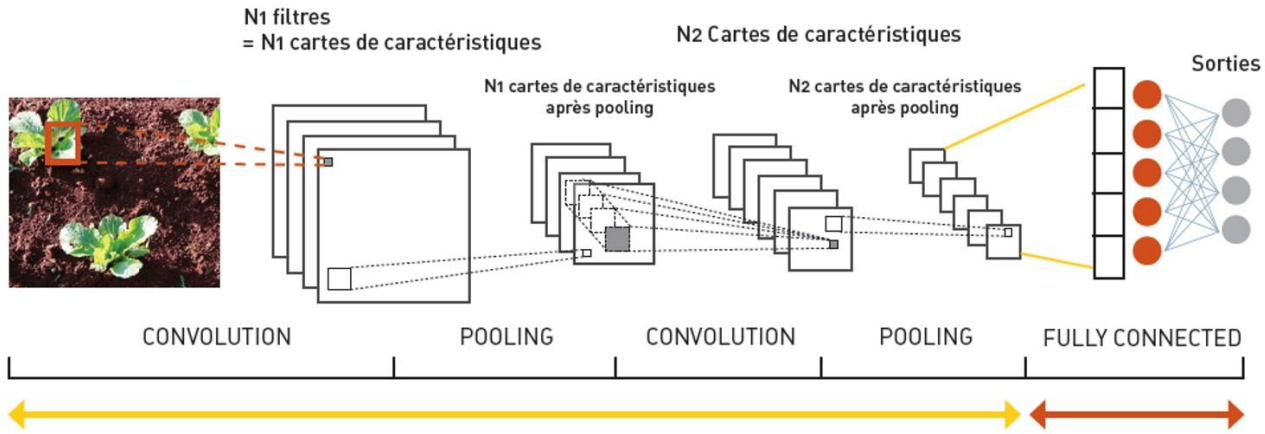


Fig. 29. Architecture simplifiée des réseaux de neurones convolutifs (CNN) (source : Toulon, Chaire AgroTIC).

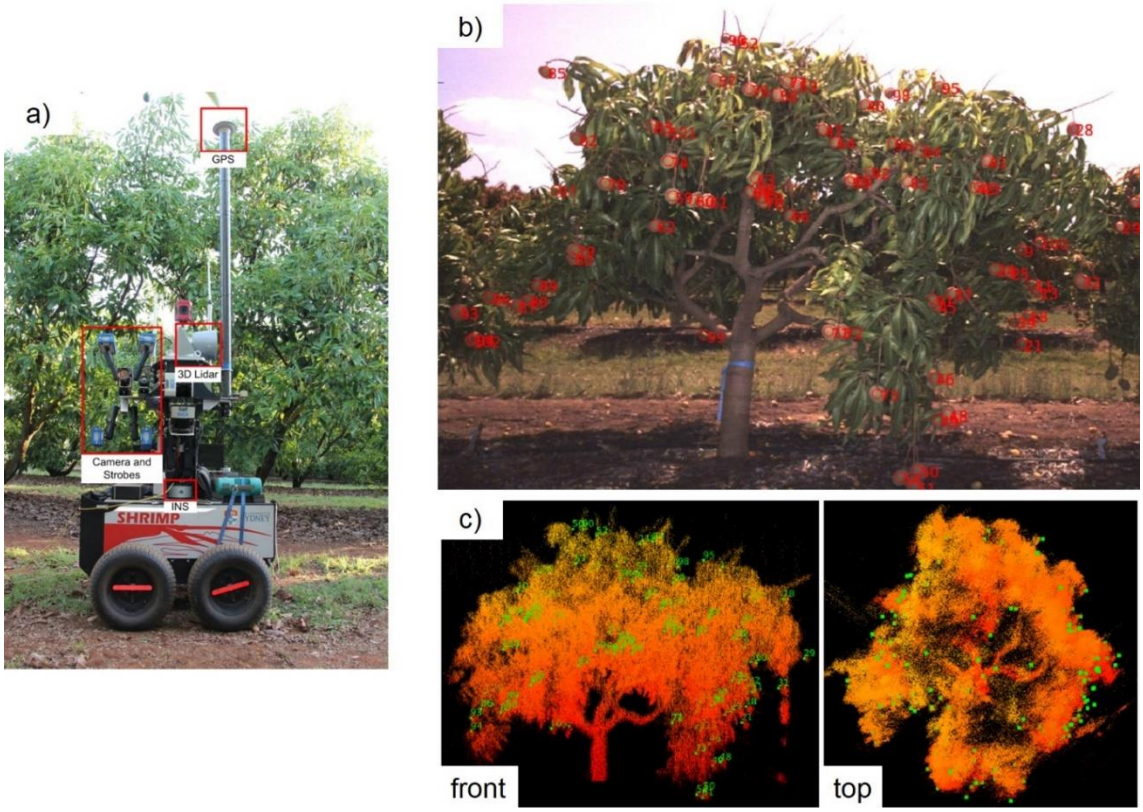


Fig. 30. Robot 'Shrimp' développé par une équipe Australienne : (a) le robot porte un capteur RVB combiné à (b) un réseau de neurones CNN permettant de détecter les mangues. (c) un LiDAR donne une reconstitution du houppier des arbres pour repositionner les fruits dans le houppier. Ce système se déplace en autonomie (GPS) permettant de balayer tous les arbres d'un verger (ici monoculture de manguiers) (source : Stein et al. 2016).

L'essor du *deep learning* s'est traduit par un accroissement de la complexité des systèmes de vision qui peuvent dorénavant traiter un plus grand flux de données qui peuvent être issues de plusieurs sources (ex : Stein et al. 2016). Dorénavant certaines équipes de recherche développent des capteurs en déplacement, sur des plateformes tractées (Payne et al. 2014 ; Nuske et al. 2014) ou des robots (Stein et al. 2016 ; Bargoti and

Underwood 2017), qui permettent de caractériser la structure des arbres (avec un LiDAR) et d'y détecter les fruits (Fig. 30a et b). La détection du fruit par analyse d'image semble donc intéressante pour estimer la production des rendements. La construction d'un jeu d'entraînement sur lequel se basent les algorithmes de *machine learning* reste une contrainte majeure pour ces outils (Koirala et al. 2019b). Le jeu d'entraînement requiert une phase d'annotation manuelle des pixels ou des objets qui peut affecter la qualité de l'entrainement et donc les résultats de l'algorithme (voir Discussion – Partie I).

III.2. Structure des arbres

Comme nous l'avons vu précédemment les volumes de houppier (à l'échelle de l'arbre) et de la canopée (à l'échelle du verger) sont des composantes déterminantes pour les rendements. Plusieurs outils sont capables de déterminer avec plus ou moins de précision la structure de l'arbre. Les capteurs LiDAR offrent la plus grande précision qu'ils soient utilisés au sol (Fig. 30c) ou portés par des drones ou des avions (Wallace et al. 2016). Ces outils sont cependant coûteux et les données obtenues sont très complexes à analyser. De nombreuses études ont démontré les capacités du drone et des capteurs RVB pour parvenir à mesurer la structure des arbres (Lisein et al. 2015 ; Nevalainen et al. 2017 ; Birdal et al. 2017). Cette mesure est permise par le développement de la photogrammétrie, plus particulièrement des algorithmes de *structure-from-motion* (SfM). Le SfM (Fig. 31) permet d'extraire les coordonnées 3D par triangulation à partir d'une série d'images géoréférencées et ayant un certain pourcentage de recouvrement ('overlap').

Ainsi, l'analyse d'images drone permet de construire un nuage de points 3D et de calculer, par extrapolation, des modèles numériques de terrain (MNT). Les MNT sont ensuite utilisés pour extraire des informations sur la structure des arbres comme la hauteur, le volume ou la surface projetée du houppier (Díaz-Varela et al. 2015 ; Birdal et al. 2017). Plusieurs études ont démontré que la précision d'un drone avec un capteur RVB et des algorithmes de SfM se rapprochent de la précision d'un LiDAR (Wallace et al. 2016). La photogrammétrie à partir d'images drone a été utilisée à plusieurs reprises pour déterminer la structure des arbres de l'ensemble d'un verger. Dans ce contexte, les principales applications de ces outils sont la mesure de la biomasse aérienne à des fins de sélection variétale (Díaz-Varela et al. 2015) ou d'agriculture de précision (Torres-Sánchez et al. 2015). La précision de ces données permet aussi d'étudier les variations de volume de la canopée avant et après les pratiques de taille (Jiménez-Brenes et al. 2017).

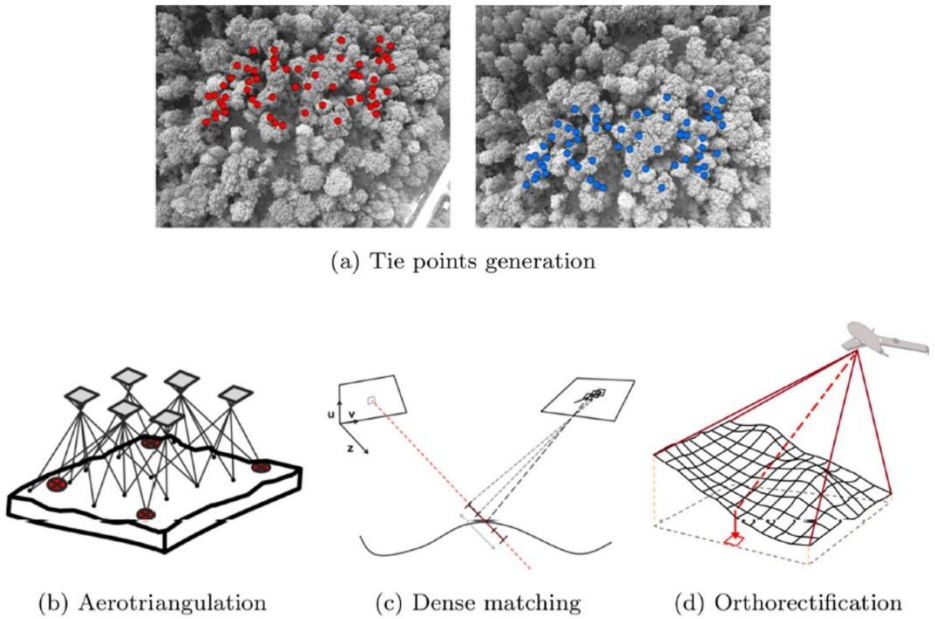


Fig. 31. Etapes du *structure-from-motion* (SfM) : (a) les points de correspondance sont d'abord estimés sur deux images qui se recouvrent, puis (b) une triangulation et (c) une reconstruction 3D sont réalisées. La dernière étape d'orthorectification corrige la géométrie du nuage de points pour produire (d) un modèle numérique et une orthomosaïque (source : Lisein et al. 2015).

III.3. Description des vergers par télédétection

Le drone est également un outil intéressant pour obtenir une carte d'occupation des sols du verger à partir de l'orthomosaïque. Les images à très haute résolution (THR), obtenues par drone, avion ou avec certains satellites (ex : images WorldView), peuvent servir à des procédures de télédétection et de classification orientée objet (Geographic Object-Based Image Analysis, GEOBIA, Blaschke et al. 2014). Par rapport à des procédures centrées sur le pixel (i.e., analyse et classification séparée des pixels), le GEOBIA permet d'obtenir de meilleurs résultats en regroupant les pixels en objets à traiter (phase de segmentation) (Blaschke et al. 2014). Cette procédure permet également de délimiter les objets contenus dans l'image pour en extraire les caractéristiques (Maschler et al. 2018). Plusieurs études ont démontré l'efficacité de classification GEOBIA à partir d'images THR pour classifier des paysages complexes comme des forêts ou des vergers (Jiménez-Brenes et al. 2017 ; Maschler et al. 2018). Par exemple, Maschler et al. (2018) ont utilisé un capteur hyperspectral aéroporté et une procédure GEOBIA afin de différencier 13 espèces arborées. La procédure GEOBIA fonctionne également avec des capteurs RVB moins couteux comme c'est le cas par exemple pour segmenter les arbres d'un verger (Jiménez-Brenes et al. 2017).

À partir de la carte d'occupation des sols il est possible d'extraire de nombreuses informations sur le verger comme sa composition spécifique (en nombre d'individus ou en proportion), la configuration spatiale des arbres (distribution agrégée vs. dispersée, connectivité entre les individus, etc.) ou tout simplement des densités de plantation ou des surfaces (Pádua et al. 2017). La télédétection permet également d'obtenir des informations sur l'état des arbres (vigueur, présence de maladie, état hydrique) voire leur production en utilisant des indices de végétation (Rahman et al. 2018). Ainsi, certaines études se sont penchées sur l'estimation de la production des arbres à partir de la télédétection aérienne avec cependant des résultats contrastés (Ye et al. 2007 ; Mann et al. 2011 ; Rahman et al. 2018). Par exemple, les corrélations entre la production de l'arbre et le NDVI (normalized difference vegetation index), indice largement répandu pour caractériser l'état de végétation, varient de très faible (< 0.18, Ye et al. 2007) à moyenne (= 0.52, Mann et al. 2011) chez les agrumes. Ceci conduit les auteurs à calibrer les modèles de prédiction de la production à chaque nouvelle situation (changement d'échelle ou de zone, changement d'année) (Rahman et al. 2018).

Les méthodes d'analyse d'image ne sont donc pas seulement intéressantes pour la caractérisation de l'arbre et du verger mais aussi pour l'estimation et la spatialisation multi-échelle du rendement. Cependant les outils utilisés par l'analyse d'image peuvent être coûteux ou peu adaptés aux vergers complexes que l'on trouve en majorité en Afrique de l'Ouest. Cette contrainte justifie un des objectifs de la thèse (p. 43) qui était le développement de méthodes fiables pour l'estimation et la spatialisation des rendements du manguier. Les méthodes développées pour ces travaux de thèse sont décrites dans le chapitre suivant (Matériels et Méthodes).

MATÉRIELS ET MÉTHODES

Afin de répondre aux objectifs de la thèse (page 43), un cadre conceptuel et méthodologique a été élaboré afin d'estimer les rendements du manguier en intégrant les différentes échelles d'études (arbre, verger et région). Ce cadre conceptuel a été mis en place sur un bassin de production d'importance pour la production de mangues au Sénégal : la région des Niayes. Dans cette région, une diversité de systèmes de culture à base de manguiers est observée. Nous avons donc choisi de nous baser sur une précédente typologie (Grechi et al. 2013) pour échantillonner des vergers répartis géographiquement dans la zone des Niayes et représentatifs des principaux systèmes de culture.

I. Cadre conceptuel et déroulement de la thèse

La méthodologie utilisée dans la thèse vise à intégrer les différentes échelles d'étude : l'arbre, le verger et la région (Fig. 32). À chaque échelle, des facteurs déterminant ou impactant le rendement agissent et conduisent à une variabilité du rendement qui doit être prise en compte à l'échelle supérieure (voir État de l'art – Parties I et II). Les travaux de thèse se sont donc d'abord intéressés à l'estimation de la production à l'échelle de l'arbre (voir Chapitre 1), unité de base du rendement. À cette échelle, la production a été estimée par analyse d'image (Chapitre 1 – Partie I et Partie II). Ces données de production ont permis d'adapter le concept des *yield gaps* à l'échelle de l'arbre (Chapitre 1 – Partie III) et d'étudier les effets du cultivar, de la structure de l'arbre, des pratiques culturales (taille et intrants) et du climat (température et précipitation). Les données sur la production et la structure de l'arbre ont ensuite été utilisées pour estimer le rendement du verger. Pour cette échelle deux approches ont été utilisées. L'une, basée sur l'imagerie drone (carte d'occupation des sols), l'extraction de la structure des arbres (hauteur de l'arbre et surface et volume du houppier) et la modélisation, permet de spatialiser le rendement (Chapitre 2 – Partie I). L'autre méthode utilise l'extrapolation de la production d'arbres échantillonnés en tenant compte de la structure du verger (Chapitre 2 – Partie II). L'obtention de cartes de rendement permet d'étudier les hétérogénéités spatiales de production potentiellement induites par la structure du verger notamment la composition spécifique et variétale et la biomasse de manguiers. Enfin, l'estimation des rendements de vergers de différents systèmes de culture permet d'utiliser la méthodologie du diagnostic agronomique régional (Chapitre 2 – Partie III). Cette méthode analytique permet de déterminer les variables importantes dans l'élaboration du rendement parmi la structure du verger (composition, biomasse de manguiers), les pratiques culturales et le pédoclimat.

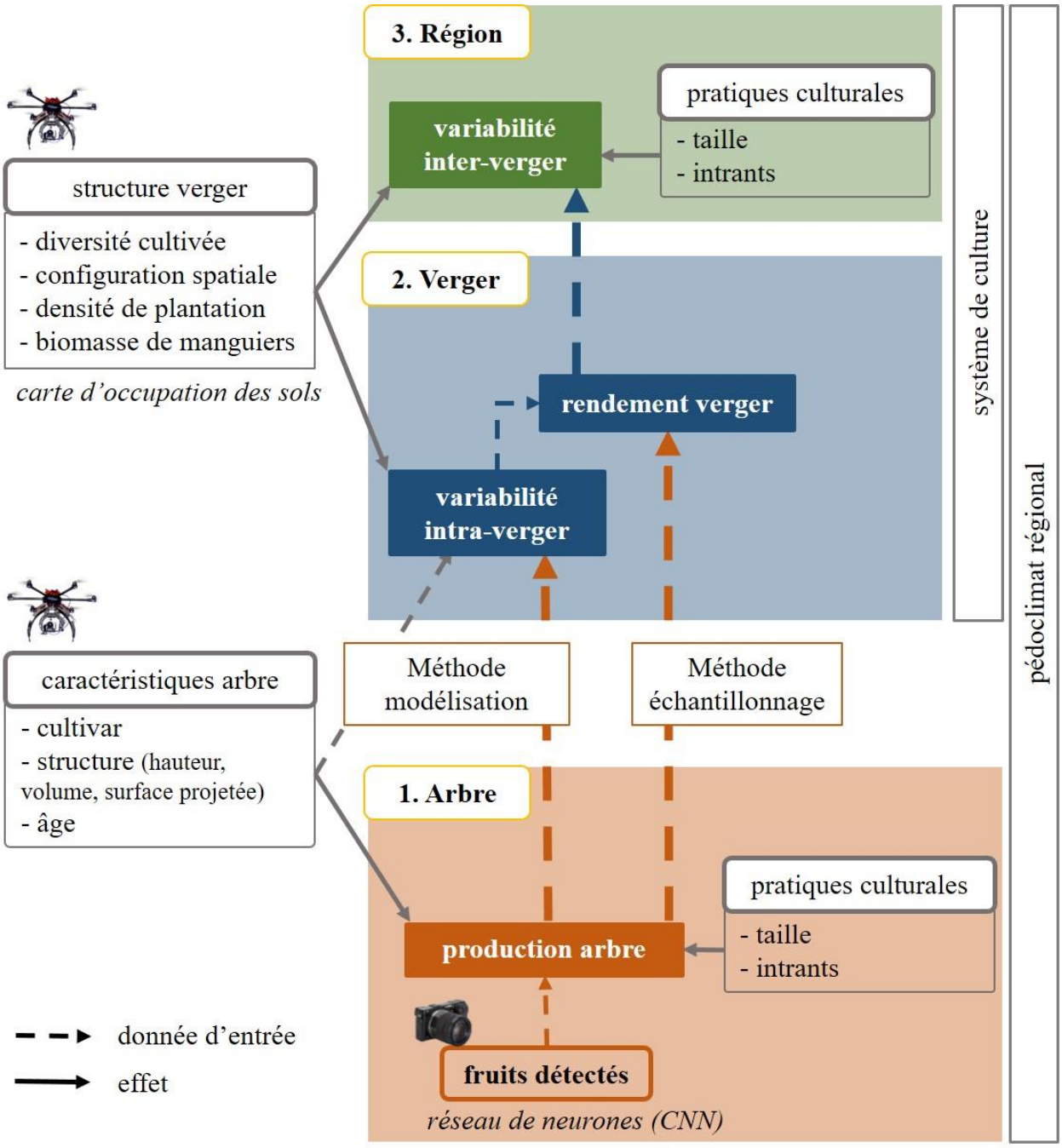


Fig. 32. Cadre conceptuel de la thèse. La production de l'arbre et les facteurs impactant sont étudiés dans une premier temps (Chapitre 1). L'estimation de la production à l'échelle de l'arbre permet de calibrer deux méthodes pour estimer le rendement du verger : l'une basée sur la modélisation (à partir de données obtenues par drone), l'autre basée sur des stratégies d'échantillonnage (Chapitre 2 – Partie I et II). Ces estimations de rendement permettent d'étudier les facteurs de variabilité inter-verger à l'échelle de la région (Chapitre 2 – Partie III).

II. Zone d'étude et plan d'échantillonnage

II.1. Les Niayes

Les Niayes (entre 14°30'-16° N et 16°6'-17°2' O) sont une région écologique et morpho-pédologique située au nord-ouest du Sénégal le long de l'océan Atlantique, entre, au sud, la péninsule du Cap-Vert (Dakar) et, au nord, la vallée du fleuve Sénégal (Saint Louis) (Fig. 33). Le plateau de Thiès constitue la limite physique au sud-est de la région, tandis qu'au nord-est le relief est moins marqué mais le bassin arachidier, bassin agricole caractérisé par la culture de l'arachide et de mil en hivernage (= saison des pluies), constitue la limite agricole des Niayes (Aguiar 2009). Les Niayes sont donc étalées sur une longueur de 180 km et une largeur allant de 30 à 35 km. Bien que situées dans le Sahel, les Niayes disposent d'un climat et d'une végétation associés au domaine guinéen (Fall 1986). Ce microclimat est généré par les courants d'air froid des Açores (venant du nord de l'Atlantique) qui protègent la région des vents chauds et secs venant de l'intérieur (Harmattan) (Fall and Fall 2001). Les températures y sont donc plus fraiches (minimale < 18°C, moyenne maximale de 31°C) et la proximité de l'océan favorise une plus forte humidité.

Les Niayes sont un ensemble de massifs dunaires et de couloirs inter-dunaires, complexifiant la morpho-pédologie de la région (Fall 1986). On distingue plusieurs systèmes dunaires qui se succèdent entre l'océan à l'ouest et l'intérieur des terres à l'est (Fig. 34, Fall 1986 ; Fall and Fall 2001 ; Fare et al. 2017):

- Les dunes vives (ou blanches ou littorales) longent tout le littoral (largeur maximale de 300 m), fortement balayées par les vents, la végétation y est rare et clairsemée, les arbres sont uniquement des filaos (*Casuarina equisetifolia*).
- Les dunes semi-fixées (ou jaunes) par une végétation plus abondante, sont des dunes de largeur variable (quelques mètres à 4 km). C'est un ensemble à l'altitude plus élevée (20 à 30 m) se terminant par un relief abrupt. Ces dunes sont interrompues par des dépressions ('ndioukis') et des lacs (ex : Tamna, Retba).
- Les dunes ogoliennes (ou rouges) et anciennes (ou 'erg') sont des dunes continentales formées durant l'Ogolien (période d'extension du Sahara 15 000 à 20 000 ans BP) et à l'altitude plus faible. Elles sont traversées également de dépressions ('niaye'). Les dunes rouges sont constituées plus particulièrement d'une couverture végétale abondante caractéristique à la fois de la steppe arborée ou, plus rarement, du domaine soudano-guinéen dans les dépressions.

Dans cet ensemble complexe, les niayes au sens strict représentent ces espaces inter-dunaires et dépressions où la nappe phréatique affleure (résurgences de la nappe fossile du quaternaire, voir Fig. 2) notamment durant la saison des pluies. Une niaye est caractérisée à la fois par un gradient pédologique et de végétation (Fig. 35). Les sols aux sommets des dunes rouges (Fig. 34a) et anciennes sont des sols sableux ferrugineux

non lessivés (‘dior’, Fall and Fall 2001). Pauvres en matières organiques ils sont sensibles à l’érosion éolienne et hydrique ainsi qu’aux pratiques culturales. Dans les dépressions, les sols sont hydromorphes (‘dekk’) plus riches en argile et en matière organique, voire parfois tourbeux (Fall 1986). Les bas-fonds sont donc plus fertiles du fait de la présence plus abondante d’espèces végétales hydrophiles (ex : *Nymphaea lotus*) aux points bas et moins sensible au manque d’eau sur les pentes (*Elaeïs guineensis*, *Cocos nucifera*). Sur les dunes on retrouve les espèces typiques du Sahel (ex : *Acacia raddiana*).

Les conditions pédo-climatiques favorables (Fig. 34), notamment dans les bas-fonds (Fig. 35), et la proximité des axes routiers (Fig. 33) ont permis le développement de l’agriculture dans les Niayes et font de cette région la première région pour la production maraîchère (principalement oignon, piment, aubergine et chou) du Sénégal. Selon les sources, les Niayes fourniraient entre 50 à 70% des productions horticoles du pays (Fall and Fall 2001 ; Fare et al. 2017). Cependant, il faut souligner la faible pertinence de ces données du fait du caractère informel des productions et des ventes sur les marchés locaux qui sont très difficiles à évaluer (Duteurtre et al. 2010).

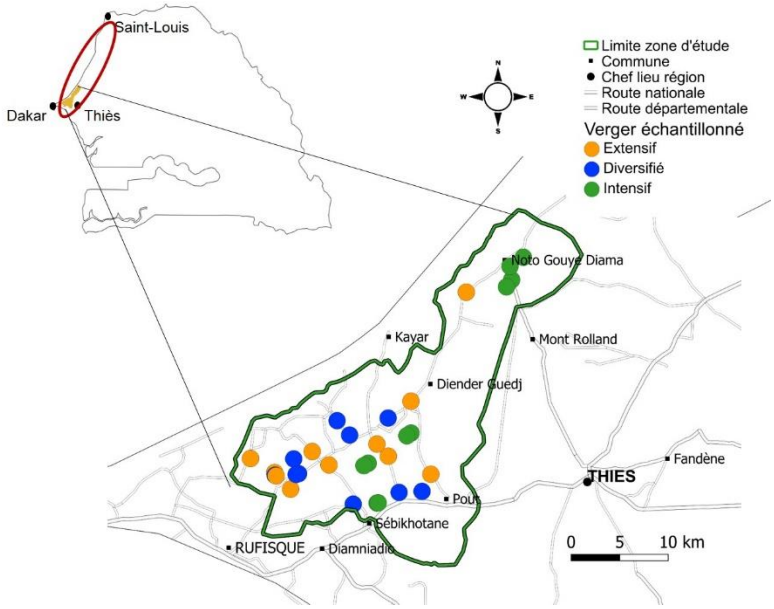


Fig. 33. Localisation de la zone d’étude et des vergers sélectionnés par type de système (extensif, diversifié ou intensif). La zone d’étude se situe au sud des Niayes (ellipse rouge).

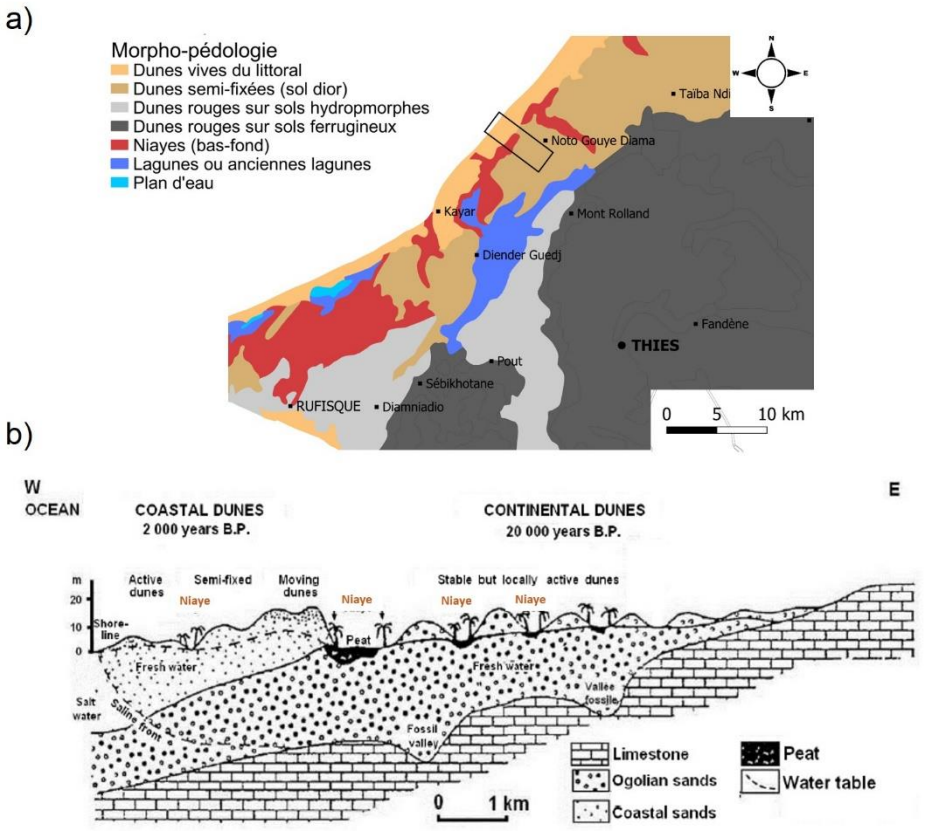


Fig. 34. Milieu physique de la zone d'étude : (a) morpho-pédologie et (b) coupe transversale (cadre sur la carte (a)) ouest-est représentant l'ensemble dunes/inter-dunes de la région des Niayes (source schéma : Fare et al. 2017).

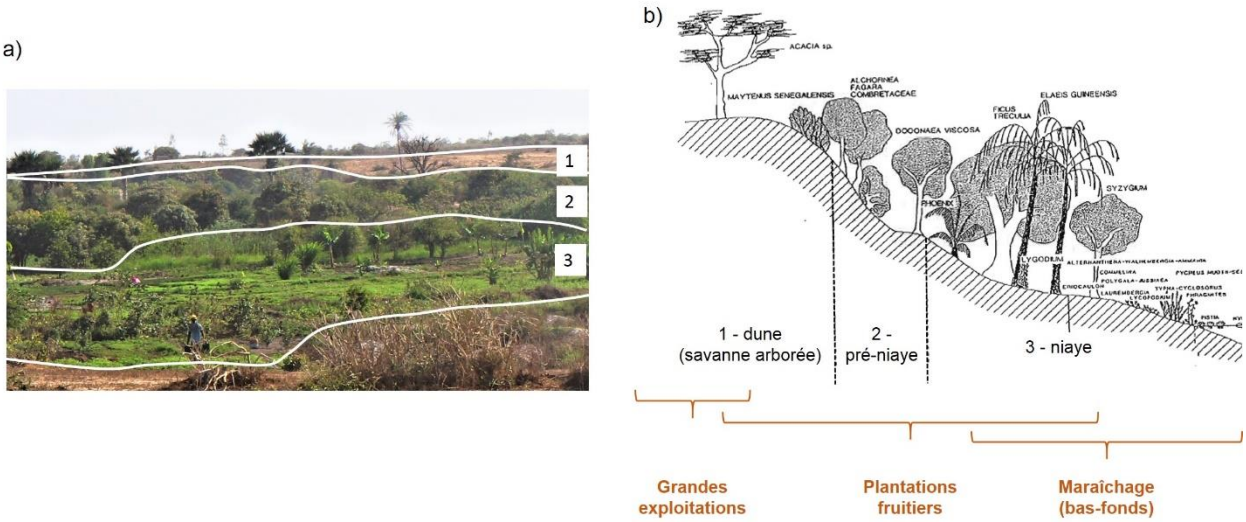


Fig. 35. Photographie (a) et représentation schématique d'une niaye composée d'une dune (1) et d'un bas-fond (2 et 3). Les conditions pédoclimatiques et la topographie contraignent le type de culture (arbre ou maraîchage) et la taille des exploitations (grande ou petite exploitation) (source schéma : Aguiar 2009).

L'hétérogénéité de topographie, de sol et d'accès à l'eau induisent une diversité des systèmes de culture dépendante également de la structure familiale des exploitations (Fare et al. 2017). Fall et Fall (2001) distinguent les petites exploitations (< 1 ha) familiales situées dans les bas-fonds (Fig. 35), des exploitations moyennes (1 à 20 ha) et plus modernes mais plus contraintes par la ressource en eau car situées sur les sols dior (hors du bas fond et nécessitant de l'eau d'irrigation). Enfin, quelques exploitations intensives (> 50 ha) sont présentes pour la production de légumes et de mangues.

Les Niayes sont donc une zone agroécologique importante au Sénégal, notamment pour la sécurité alimentaire et l'économie rurale. Cependant les changements démographiques passés (mouvements de population lors des sécheresses des années 1970 et 1980) et présents (urbanisation) exercent une pression sur le foncier et les ressources hydriques (Camara et al. 2019). S'ajoute à cela l'effet du changement climatique qui se fait déjà ressentir (voir Contexte et Problématiques). La nappe phréatique n'affleure plus tout au long de l'année dans les bas-fonds du fait de l'allongement du nombre de jours secs consécutifs (Aguiar 2009).

Tableau 1. Description des trois principaux types de vergers à base de manguiers dans les Niayes

	Extensif	Diversifié	Intensif
Nombre échantillonné	9	10	10
Surface productive moyenne (ha)	1.32	2.92	2.7
Rendement approximatif ($t.ha^{-1}$)	< 5	5 à 10	20 à 35
Densité de manguiers ($arbres.ha^{-1}$)	147	122	226
% de variété 'Kent'(%)	70	43	80
Arbres cultivés	Manguier	Manguier et autres espèces	Manguier
Maraichage	Faible et vivrier	Elevé et de rente	Nul
Pratiques culturales	Nulle à faible.	Moyenne. Engrais organiques. Irrigation du maraîchage et/ou des agrumes. Certains vergers en agriculture biologique	Elevée. Engrais et produits phytosanitaires systématiques. Désherbage (mécanique et/ou chimique)

II.2. Typologie des vergers de mangues de la zone des Niayes

Même si les chiffres ne sont pas établis avec précision, on considère les Niayes comme la deuxième région de la production de mangues au Sénégal (37% de la production totale) et la première pour la mangue exportée (70% des exportations) (Ternoy et al. 2006 ; Rey and Dia 2010). Le microclimat de la région fait

que la mangue arrive en production après toutes les autres région d'Afrique de l'Ouest, lui offrant un avantage sur le marché international (Rey et al. 2007). Dans les Niayes les variétés dominantes sont les variétés floridiennes 'Kent' et 'Keitt' dans les vergers d'exportation mais on retrouve de nombreuses variétés améliorées ('Amélie', 'Divine') et locales ('Boucodiékhal', 'Séwé') dans les petites exploitations.

Les systèmes de culture à base de manguiers ont été caractérisés par Grechi et al. (2013) qui ont construit la seule véritable typologie centrée sur les vergers à base de manguiers dans la région. Les auteurs ont distingué quatre types de systèmes. Dans cette thèse nous n'en retiendrons que trois, le quatrième système, majoritairement composé d'agrumes, n'étant pas observé dans les Niayes. Ces trois systèmes (Tableau 1 et Fig. 36) se distinguent par leur composition spécifique et les pratiques culturales (fertilisation, produits phytosanitaires, irrigation et taille) :

- Les vergers 'extensifs' représentent les vergers traditionnels (de cueillette) à niveau d'intrants très faible à nul. Ces vergers, de petite taille, sont constitués majoritairement de manguiers, mais également d'autres espèces plantées de manière sporadique (agrume, anacardier, goyavier). Leur densité de plantation est faible mais l'arrangement des arbres est aligné. Les variétés de manguier sont majoritairement la 'Kent' et les variétés locales.
- Les vergers 'diversifiés' regroupent les vergers avec une forte diversité arborée et +/- du maraîchage. Les arbres sont taillés chaque année et les niveaux d'intrants sont moyens : des fertilisants organiques sont employés, et l'irrigation peut être apportée par l'intermédiaire du maraîchage ou des agrumes. Ce type de verger correspond à un système agroforestier multi-stratifié (Jagoret et al. 2017). Les manguiers sont plantés avec d'autres espèces (agrume majoritairement mais également papayer, corossolier et sapotillier). Ces espèces associées au manguier constituent une strate arborée intermédiaire, le maraîchage constituant une dernière couche de végétation au sol. Le nombre de variétés de manguier est le plus important dans ce type de verger.
- Les vergers 'intensifs' représentent les vergers commerciaux d'export. Ces vergers sont en réalité des parcelles de 1 à 2 ha dans des exploitations plus larges. Dans ces parcelles le manguier est cultivé en monoculture, de 'Kent' et parfois de 'Keitt', à très haute densité (de 200 à 400 arbres.ha^{-1}). Les niveaux d'intrants sont les plus élevés avec une à deux fertilisations minérales par an, l'irrigation des arbres de la floraison à la récolte, l'application d'au minimum 2 traitements phytosanitaires et une taille est appliquée une à deux fois par an.

Les variabilités de densité de plantation et de pratiques culturales aboutissent à des rendements différents selon le système de culture. Ainsi, à dire de producteur, les vergers extensifs sont les moins productifs (maximum 5 t.ha^{-1}) suivi des vergers diversifiés (5 à 10 t.ha^{-1}) et des vergers intensifs aux rendements les plus élevés (entre 20 et 30 t.ha^{-1}). Ces chiffres de rendement ne sont cependant que des suppositions ou les

chiffres de la production vendue uniquement. L'un des objectifs de la thèse sera justement de pouvoir estimer de manière fiable et exacte les rendements des systèmes de culture.

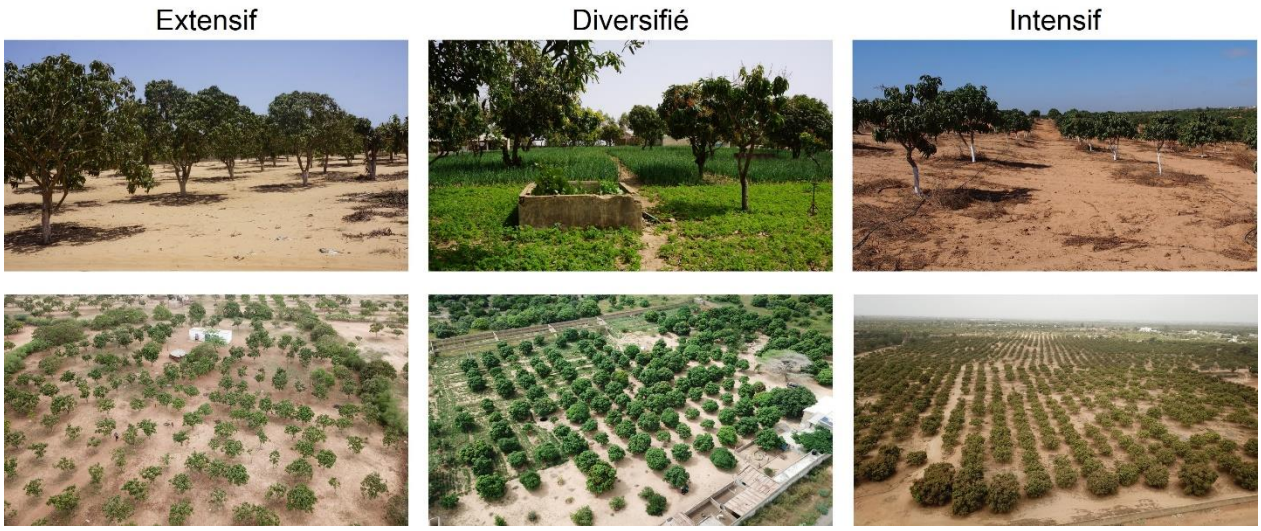


Fig. 36. Vues sols (image haut) et vues aériennes (image bas) des trois principaux types de vergers à base de manguiers dans les Niayes : vergers extensif sur sol nu, vergers diversifié avec maraîchage et verger intensif d'exportation (source photo : J. Sarron).

II.3. Echantillonnage

Les vergers ont été échantillonnés dans le sud des Niayes (500 km²) entre les communes de Sangalkam et Noto Gouye Diama (Fig. 33). Les vergers ont été choisis de manière à avoir une répartition de la diversité géographique des trois systèmes de cultures étudiés. Pour des raisons de praticité, les vergers ont été choisis au bord des routes goudronnées ou des pistes afin d'en faciliter l'accès. Au total, 29 vergers ont été sélectionnés et suivis entre 2017 et 2018. On constate une forte présence de vergers intensifs au nord de la zone d'étude (Fig. 33) du fait de la concentration des exploitations commerciales autours de la commune de Noto, place de marché importante. Afin de procéder à l'étude à l'échelle de l'arbre, 10 arbres ont été sélectionnés dans chaque verger du réseau (Fig. 37). La sélection des arbres est répartie dans le verger et représentative des proportions des trois variétés majeures de la zone : 'Kent', 'Keitt' et 'Boucodiékhal' (noté parfois BDH). De plus, les arbres ont été choisis de manière à obtenir un gradient de taille et de charge (pour l'année 2017) afin d'avoir une représentativité de structure et de rendement (Fig. 37b). Un des vergers ayant une superficie de 14 ha, bien supérieure aux autres, l'effort d'échantillonnage a été doublé. Au total : 200 'Kent', 57 'Keitt' et 43 'Boucodiékhal' ont été choisis.

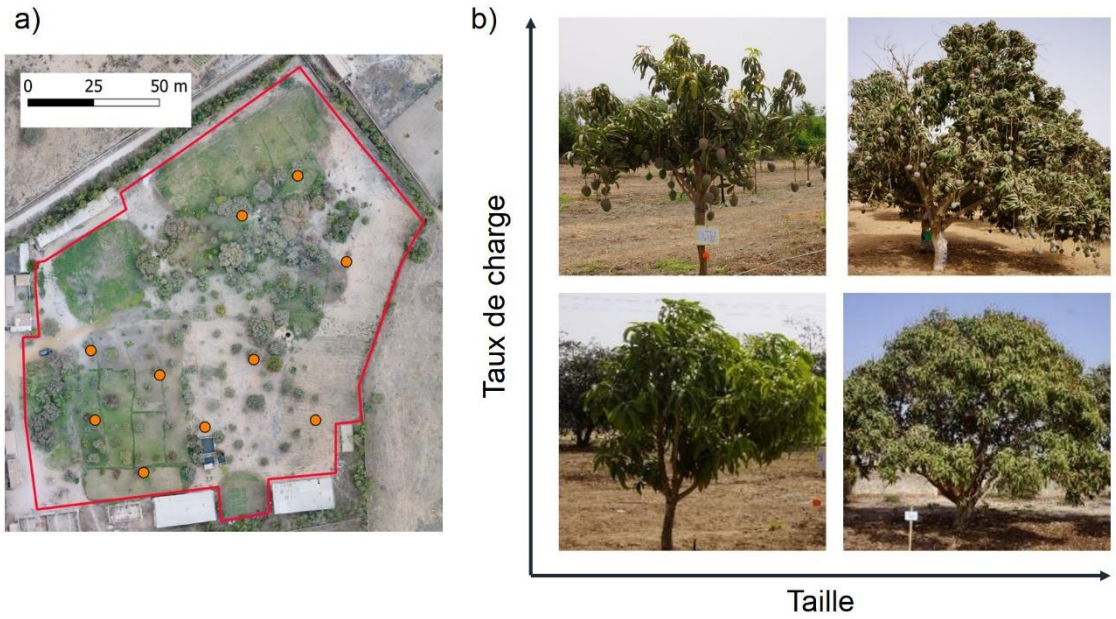


Fig. 37. Echantillonnage des arbres dans les vergers. Les arbres sont choisis de manière à être réparti dans le verger (a) et à être représentatifs de la taille et de la charge en fruit observées dans la zone (b).



Fig. 38. Protocole d'acquisition des images pour le suivi de la production des arbres (D = distance entre l'arbre et l'opérateur).

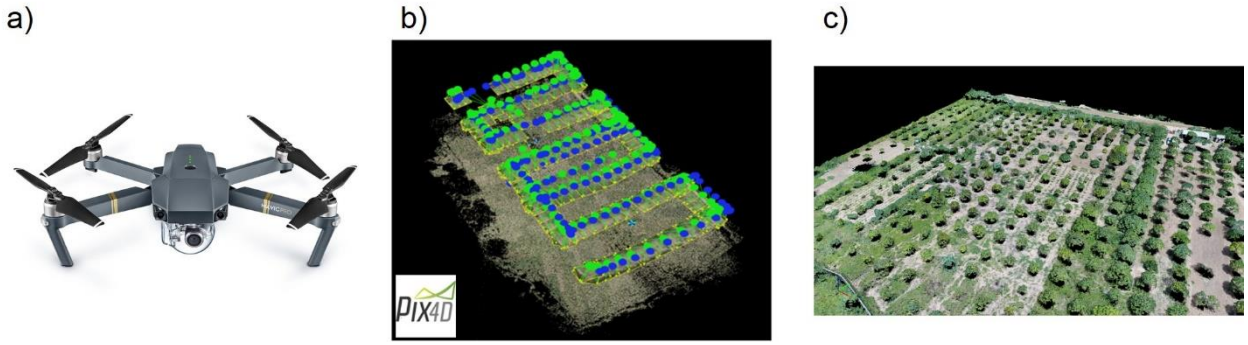


Fig. 39. Vol de drone réalisé pour cartographier les vergers : (a) drone Mavic Pro avec capteur RGB (12 Mpixel), (b) exemple de plan de vol, (c) exemple de nuage de points 3D obtenu après triangulation des images drones et servant à calculer les modèles numériques.

III. Acquisition des données

III.1. Suivi photographique pour estimer le nombre de fruits

Les méthodes employées pour estimer le nombre de fruits par arbre se basent sur l'analyse d'images. Un dispositif d'acquisition d'images a été mis en place (Fig. 38). Les arbres ont été photographiés sur deux faces opposées et à une distance fixée (2, 5, 10 ou 15 m) afin de capter l'intégralité du houppier. La prise de vue a été réalisée avec un appareil photo numérique classique (Sony Nex-7, Sony Corporation, New York, USA) prenant des images visuelles RVB de résolution 24 Mpixel. Sur les 300 arbres du réseau, les deux images ont été prises au stade fruit mature soit 1 mois à 2 semaines avant les récoltes de 2017 et 2018. L'orientation des faces à photographier n'a pas été contrainte de manière à choisir l'angle où l'arbre était le moins mêlé aux branches des arbres voisins. Pour les deux années, le même angle de prise de vue et la même distance arbre/opérateur ont été utilisés.

Afin de décrire la variabilité intra verger, trois vergers ont été spécifiquement sélectionnés de manière à représenter les trois systèmes de culture. Dans ces vergers, tous les manguiers ont été pris en photo avant la récolte de 2018 et en utilisant le même protocole d'acquisition.

III.2. Calibration des modèles d'estimation de la production

Les images d'arbre au stade mature ont été traitées par un réseau de neurones permettant de détecter la majeure partie des mangues (voir Chapitre 1 – Partie I). A partir du nombre de fruits détectés, le nombre de fruits réels est estimé par des modèles calibrés sur 60 arbres de calibration. Les arbres de calibrations ont été répartis dans les trois variétés (20 arbres de chaque) et dans les types de verger. Le nombre de fruits dans les arbres de calibration est estimé par un comptage visuel au champ à l'aide de trois experts. Afin d'être au plus juste, l'écart de comptage entre les trois évaluateurs ne devait pas dépasser les 10% sinon l'arbre devait être systématiquement recompté. Afin de convertir le nombre de fruits estimé en production (exprimée en kg), le poids moyen de chaque variété a été mesuré sur 66 et 140 fruits en 2017 et 2018 respectivement.

III.3. Structure des arbres et carte d'occupation des sols

Pour caractériser chaque verger, un vol de drone a été réalisé en 2017 en utilisant un drone commercial (quadricoptère) avec un capteur RGB (DJI Mavic Pro DJI Inc., Shenzhen,China). Un plan de vol a été mis en place afin d'acquérir des images très hautes résolution (THR) avec un taux de chevauchement ('overlap') de 80% afin d'obtenir une résolution 3D suffisante (Fig. 39). Les images permettent d'obtenir, via des algorithmes de *structure-from-motion* (Fig. 31, Lisein et al. 2015), deux modèles numériques : le modèle numérique de surface (ou digital surface model, DSM) et le modèle numérique de terrain (ou digital terrain model, DTM). Une orthomosaïque RVB est également construite à une résolution de 1.30 cm.pixel^{-1}. Pour

obtenir les modèles numériques et l'orthomosaïque, les images ont été traitées grâce au logiciel Pix4Dmapper Pro (Pix4Dmapper 1.3; Pix4D SA, Lausanne, Switzerland).

Les modèles numériques permettent de calculer un 'Canopy Height Model' (CHM) qui est en fait la hauteur de tous les points au-dessus de la surface avec une résolution de 6.40 cm.pixel^{-1} (Birdal et al. 2017). Pour chaque verger nous avons ainsi obtenu une carte d'occupation des sols (Fig. 40 et voir Chapitre 2 – Partie I) contenant les manguiers et leur variété (3 variétés étudiées + 1 classe pour les autres) ainsi que les principales autres espèces arborées (anacardier, agrume, papaye, palmier, mimosae, et autre). Les surfaces du verger et en culture maraîchère à l'intérieur des vergers ont pu être mesurées avec précision. Sur la carte d'occupation des sols, les houppiers des arbres sont délimités individuellement. En couplant la carte et l'information de hauteur contenue dans le CHM, trois paramètres de structures d'arbres sont calculés : la hauteur de l'arbre, la surface et le volume projetés du houppier.

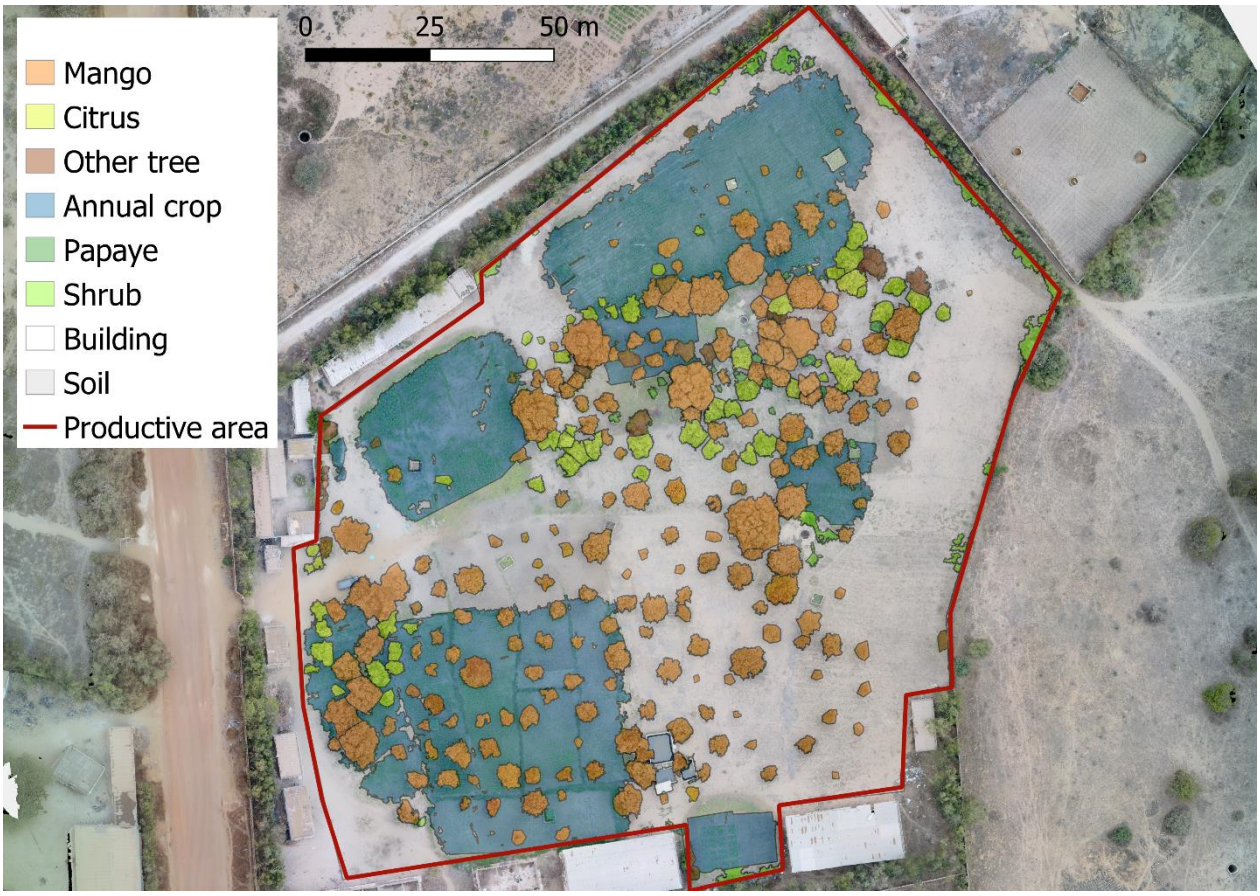


Fig. 40. Exemple d'une carte d'occupation des sols obtenue par drone (verger diversifié).

III.4. Indice de charge pour l'estimation du rendement du verger

L'outil drone permet d'acquérir des données de structure sur tous les arbres. À l'aide de ces informations et des données de production, obtenues par analyse d'images, des modèles d'estimation du rendement ont été calibrés. Or la structure ne suffit pas pour estimer la production de l'arbre du fait de l'effet des pratiques et des phénomènes d'alternance. Pour prendre en compte ces effets, nous avons introduit une variable, nommée indice de charge ('load index') dans les modèles. Cette variable est estimée visuellement, à l'aide d'un abaque, et correspond au pourcentage de la surface du houppier occupée par des fruits par rapport à la surface totale (Fig. 41). Pour chaque année, l'indice de charge est estimé par trois évaluateurs sur 50 arbres sélectionnés sur un transect traversant les diagonales du verger (Fig. 42). L'indice de charge permet ainsi d'évaluer la charge moyenne de chacune des trois variétés dans le verger.



Fig. 41. Variable de l'indice de charge ('load index') estimée visuellement sur des transects. Cette variable est introduite pour la modélisation de la production des arbres en utilisant le drone. Elle permet de prendre en compte les variabilités de charge entre les arbres.



Fig. 42. Echantillonnage des arbres par transect (50 arbres par verger) pour l'extrapolation du rendement

III.5. Données pédoclimatiques

Afin de prendre en compte l'effet température dans nos analyses, des capteurs de température ont été disposés dans chaque verger. Ces capteurs (HOBO U23 Pro v2, Onset Computer Corporation, USA) ont été protégés de la lumière directe et placé à une hauteur de 2 m. Le pas de mesure a été fixé à une mesure toutes les 10 minutes. Une station climatique (HOBO U30, Onset computer Corporation, USA) localisée au centre ISRA de Sangalkam (14°47'22.0"N, 17°13'36.0"W) a permis d'enregistrer la pluviométrie, la vitesse et l'orientation du vent ainsi que le rayonnement solaire.

S'ajoutent à ces mesures météorologiques annuelles des données pédoclimatiques prises à partir de bases de données disponibles en ligne. Le contexte climatique (radiation solaire, température et précipitation) de long terme a été caractérisé par les données du WorldClim 2 (Fick and Hijmans 2017) à la résolution spatiale 1 km. Les informations sur la structure et la composition du sol (contenu en azote, matière organique, etc.) ont été obtenues sur la base de données AfSoilGrids250m (Hengl et al. 2015) à la résolution spatiale 250 m.

III.6. Données d'enquêtes

Les informations relatives au verger (âge, porte greffe, conduite du maraîchage, etc.) et à la conduite du manguier ont été demandées au producteur sous forme de questionnaires avec questions ouvertes ou fermées (Annexe I). Les informations ont été demandées avec le plus de précision possible. Malheureusement de nombreux producteurs n'ont pas de suivi précis des produits et des pratiques appliqués sur les vergers. Ceci nous a conduits à simplifier certaines variables en créant des catégories d'intensité par exemple.

CHAPITRE 1

Estimation et étude de la production du manguier

Introduction

Ce chapitre présente les résultats de la thèse concernant l'estimation de la production à l'échelle de l'arbre et l'étude des facteurs de son hétérogénéité. Dans les deux premières parties, nous abordons le développement et la validation de deux méthodes basées sur l'analyse d'images. Ces méthodes ont été mises en place pour estimer le nombre de mangues dans le houppier pour différentes variétés et dans différentes conditions environnementales et pratiques culturales. Elles combinent des algorithmes de *machine learning* avec des modèles d'estimation. Dans un premier temps (Partie I), la classification k-nearest neighbours (KNN), orientée sur le pixel, a été utilisée pour détecter les mangues. Son étude a permis de démontrer les limites de cet algorithme pour la détection de fruits dans des milieux hétérogènes (plusieurs variétés et arbres de taille variable). Dans un second temps (Partie II), un réseau de neurones convolutif (CNN), le Faster-RCNN, a été entraîné et déployé dans notre contexte d'étude. Son utilisation a permis d'améliorer grandement le pourcentage de mangues détectées comparé au KNN. L'emploi du CNN combiné à des modèles d'estimation, calibré pour chaque variété, donne ainsi une mesure précise de la production réelle de l'arbre.

La troisième partie de ce chapitre (Partie III) aborde le concept de '*production gap'*. La production des 300 arbres du réseau de vergers, mis en place dans la zone des Niayes, a été estimée pendant deux années consécutives. La méthodologie des '*yield gap*' (van Ittersum et al. 2013) a été adaptée à l'échelle de l'arbre pour permettre l'étude des facteurs explicatifs de l'hétérogénéité de production dans la zone d'étude. Par le biais de la modélisation, la production atteignable de chaque variété est estimée en fonction de la structure des arbres que l'on considère comme un facteur limitant (Annexe II). Le production gap se défini alors comme la différence entre la production atteignable et la production réelle. La méthode statistique employée permet d'identifier les facteurs augmentant ou diminuant l'écart de production parmi les pratiques culturales, la structure du verger et le pédoclimat.

Ces premiers travaux constituent une phase nécessaire et importante puisque la production du manguier en Afrique de l'Ouest n'avait jamais été estimée jusqu'à maintenant. L'outil d'estimation de la production, combinant le CNN et les modèles, fonderont la base d'un outil plus large, déployé auprès des acteurs de la filière (producteurs, acheteurs, exportateurs et institutions) pour leur permettre d'obtenir des informations fiables sur la production de mangues à différentes échelles (Annexe III). L'étude de la variabilité de production à l'échelle de l'arbre permet de mieux comprendre les facteurs contrôlant cette dernière. Ce type d'approche permet de mieux comprendre l'établissement de la production dans les différents systèmes de culture et également de comparer les effets des pratiques culturales et de l'environnement.

I. Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions?

Cette partie du Chapitre 1 présente un premier outil pour la détection de mangues basé sur un algorithme de *machine learning* : k-nearest neighbours (KNN). Cet algorithme classifie chaque pixel et nécessite une phase de calcul d'indices de couleur et de texture ainsi qu'une phase de post-traitement. Ces travaux ont permis de montrer que le KNN est moins performant pour détecter des fruits issus de conditions hétérogènes (plusieurs variétés et plusieurs systèmes de culture). Pour ces raisons il a été remplacé par un algorithme plus performant de *deep learning* (voir Chapitre 1 – Partie II).

Cette partie est constituée d'un acte de conférence à comité de lecture qui a été présenté au *XXX. International Horticultural Congress* (IHC) à Istanbul (Turquie) en août 2018, dans le *II. International Symposium on Mechanization, Precision Horticulture, and Robotics*, et a remporté le prix du Young Minds Award de la meilleure présentation orale du symposium. Cet acte est actuellement en cours de publication.

- Sarron J.; Sané C.A.B.; Borianne P.; Malézieux É.; Nordey T.; Normand F.; Diatta P.; Faye É. **Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions?** XXX. International Horticultural Congress, 2018, Istanbul, Turkey [*Acta Hortic., in press.*].

Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions?

Julien Sarron[1,2,3], Cheikh Amet Bassirou Sané[4], Philippe Borianne[5,3], Éric Malézieux[1,3], Thibault Nordey[1,3,6]Faye[1,2,3], Frédéric Normand[1,3], Paterne Diatta[2], Youga Niang[2], Émile Faye[1,2,3]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]Centre pour le Développement de l'Horticulture, ISRA, Dakar, Senegal
[3]Univ Montpellier, F-34090 Montpellier, France
[4]Université Cheikh Anta Diop (UCAD), Fac. Sci. Tech., BP 5005 Dakar, Senegal
[5]CIRAD, UMR AMAP, F-34398 Montpellier, France
[6]The World Vegetable Center, Eastern and Southern Africa, P.O. Box 10, Duluti, Arusha, Tanzania;

Abstract

In the last decade, image analysis using machine learning algorithms proved its potential for the detection and counting of plant organs. Numerous studies provided fruit tree yield estimates based on machine learning with high levels of efficiency. However, most of these studies were conducted under homogeneous conditions of fruit aspect. The aim of this study was to develop an efficient machine learning method for ripe mango fruit detection from RGB images and to test it under heterogeneous field conditions for tree yield estimation in Senegal. The algorithm consisted in a k-nearest neighbours classification based on colour and texture features followed by a post-treatment based on shape indices. The F1 score, which accounts for both precision and recall performances, reached 0.73 for a set of 42 images of 'Kent' trees in homogeneous conditions. When performed on a second set of 128 images representing the actual heterogeneity in tree structure (height, canopy volume) and cultivars ('Kent', 'Keitt' and 'Boucodiékhal') found in the Niayes region of Senegal, the F1 score fell to 0.48. This decrease resulted from the heterogeneous field conditions in terms of fruit size, fruit colour and light exposure caused by different tree structures among cultivars. Despite the algorithm was less efficient under these conditions, significant linear relationships were evidenced between the number of detected fruits and the actual number of fruits per tree for each cultivar ('Kent': $R^2 = 0.92$, 'Keitt': $R^2 = 0.93$, and 'Boucodiékhal': $R^2 = 0.90$). These models were cross-validated and reached a relative RMSE of 14%. Those results offer new opportunities to accurately and rapidly estimate mango yield and to further identify the parameters that drive its variability at the tree and orchard scales.

Keywords: image analysis, automated fruit counting, k-nearest neighbours, algorithm efficiency, Senegal.

I.1. Introduction

Mango (Mangifera indica L.) is a major fruit crop of the tropics and sub-tropics that guarantee incomes and food security for local and international markets (Mukherjee and Litz, 2009). In Senegal, as in many others places in the tropics, the mango sector is extending under environmental and demographic changes, stressing the need for accurate estimation of fruit production. One of the main issues that currently hampers mango crop development is the impossibility to accurately estimate crop yield at relevant stages of growth to inform farmer decisions. To date, yield estimation in tropical orchards is still based on visual inspection of a limited number of trees, a tedious and time consuming method that depends on the observer reliability and accuracy

(Linker et al., 2012; Gongal et al., 2015). Precision agriculture, through image processing and machine vision system, offers new opportunities to rapidly gather accurate information of the orchard production and assess in-field yield heterogeneities.

Over the last few years, several studies described efficient machine vision systems for fruit detection and yield estimation in trees (see Gongal et al., 2015 for a review). Advances of these systems has been triggered by development of both sensors (RGB, multispectral, LiDAR, thermal, etc.) and efficient methods for image processing (e.g. machine learning). Moreover, they were usually developed and evaluated in monocultivar commercial orchards under homogeneous conditions with low variation in light and tree structure and no technical constraint for image acquisition (i.e. in-row configuration of orchard) (Hung et al., 2015). In those conditions, Payne et al., (2014); Qureshi et al., (2016); Bargoti and Underwood, (2017), obtained relatively high performances in term of mango fruit detection. However, most of these studies focused on improving image analysis technics rather than testing the robustness of their method when used under heterogeneous field conditions in where variations in tree canopy and fruit aspect occur. Indeed, variation in tree and fruit aspects – including tree cultivars and structure, fruit colour and shape - can affect detection algorithm and lead to tree yield misestimates.

To address these issues, this study described and evaluated the efficiency of a machine learning algorithm for ripe mango fruit detection when used under homogeneous conditions (trees of a single cultivar in a commercial orchard, homogeneous fruit aspect…) and under heterogeneous conditions (3 cultivars, various tree height, canopy shape and variable fruit colour). The aim of this work was to pinpoint the features responsible for discrepancies in algorithm efficiency depending on imaging conditions. Besides seeking for the best efficiency, this work investigates more thoroughly the possibility of using machine learning to operationally estimate tree yield under heterogeneous field conditions.

I.2. Materials and methods

a. Study area

In West Senegal, the Niayes region is the country's first area for mango exports and foremost region for local market. The climate mitigated by a cool and humid trade wind during the hot season makes this region favourable for fruits and vegetables production (Diame et al. 2015). In this region, a large diversity of cropping systems co-exists from small family-based diversified orchards to large monocultivar commercial orchards (Grechi et al. 2013) (Fig. 43).

b. Image acquisition for algorithm

The study was carried out in orchards in the Niayes region of Senegal. In these orchards, mango trees have been pictured on 2 opposite sides using a Sony Nex-7 RGB camera (Sony Corporation, New York, USA) with a fixed focal length of 18 mm and operated in autofocus mode. Images were acquired at a resolution of 4000 x 6000 pixels under clear sky conditions between 10:00 AM and 15:00 PM (to minimize shadow and light variation effects). Trees were imaged at fruit maturity between 1 to 2 weeks prior to the first harvest date predicted by the farmers. Two sets of image were acquired to test the effect of the variability of field conditions on the efficiency of the detection algorithm.

Image set 1 was made of 50 images of 25 'Kent' trees planted in-rows in a commercial orchard (Fig. 43a). All images were acquired the same day, in June 2017 (7 days before harvest) at a 5 m distance from tree

trunk. In this orchard, trees were annually pruned resulting in homogeneous tree structure (height and canopy volume). Moreover, trees displayed a weak reproductive asynchronism within and between individuals (Dambreville et al., 2013). Under these conditions, fruits appeared homogeneous in size and colour (a strong blush with few green parts).

Images set 2 was made of 128 images of 64 trees acquired at the maturity stage between May and July 2017 in 30 orchards of the study area from diversified orchards with various cultivars (Fig. 43b) to monocultivar orchards (Fig. 43a). Images were taken on 22 'Kent', 22 'Keitt', and 20 'Boucodiékhal' (BDH) mango trees depicting field heterogeneity of cultivars and tree structures found in the studied area. Because trees height ranged from 1.5 m to 12 m high, the distance between the camera and the tree was 5, 7, 10 or 15 m in order to capture the entire tree. Under these conditions, trees showed strong reproductive asynchronisms within (i.e. various fruit development stages within the same tree) and between trees. Variability in tree height and asynchronisms leaded to a high heterogeneity in fruit size and colour (varying from fully green to yellow and red) within and between the images of the set.

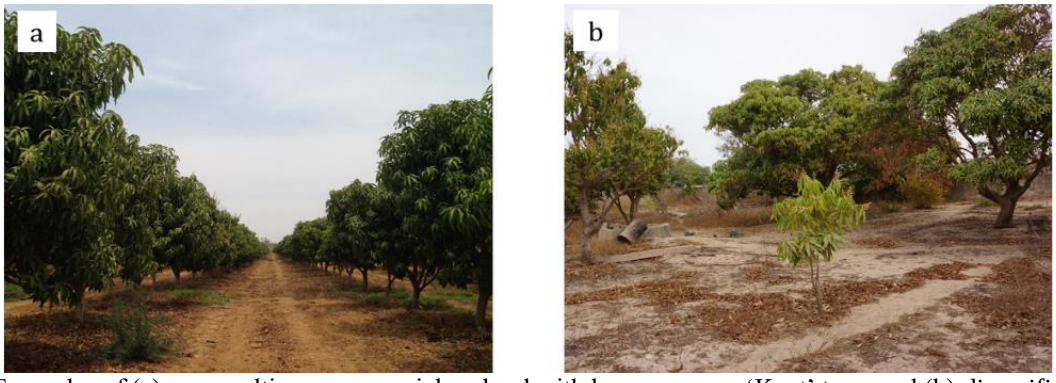


Fig. 43. Examples of (a) monocultivar commercial orchard with homogeneous 'Kent' trees and (b) diversified orchard with heterogeneous trees of various cultivars.

c. Algorithm workflow

The algorithm for mango detection and localization was developed and evaluated using the open source R software version 3.4.1 (R Core Team, 2015). The workflow of the algorithm included a k-nearest neighbours (KNN) classification followed by a post-treatment based on shape indices (Fig. 44). KNN is a pixel-based machine learning classification that defines the class of a test pixel according to the class of the k-nearest training pixels described by N features (for details on KNN see Gongal et al. 2015). Two training datasets were built using 8 and 24 representative images of Image set 1 and 2, respectively. For each training image, pixels of fruit and non-fruit were manually selected. Then, non-fruit pixels were randomly resampled to represent 3 times the number of fruit pixels to reduce dataset size and then computing time. Each training pixel had 6 input variables for KNN classification; values for R, G and B bands, and 3 feature variables: the Cr band of YCbCr colour space (ITU standard), the VARI index (Gitelson et al., 2002) and the variance layer issued from a 3x3 texture filter (Payne et al., 2014). To reduce the computing time, these 3 feature variables have been chosen among 20 commonly used features in the literature (based on colour spaces RGB, HSL or YCbCr, and on texture) using a Learning Vector Quantification for feature selection (data not shown).

The KNN classification resulted in binary images with fruit and background pixels including a lot of noise. Consequently, we used a post-treatment in order to: 1) reduce the noise, 2) localize patches of fruit pixels,

and 3) filter patches by using shape indices (Fig. 44). First, noise reduction involved holes filling in groups of fruit pixels followed by a median filter to delete isolated pixels. Second, disjoint groups of fruit pixels were identified as patches. A double thresholding was applied on patches size to delete small patches (e.g. residual noise or background fruits) and big patches corresponding to misclassification of large objects (soil, branches, etc.). The kept sizes depended on the distance between the camera and the tree. Third, patches were filtered by their shape to select the elliptical and circular ones. Shape.index and perim.area.ratio calculated by the PatchStat function of SDMTools R package (VanDerWal et al., 2014) and the Related Circumscribing Circle (RCC) described by Baker and Cai (1992) were used (Baker and Cai, 1992). Patches with values of shape.index below 2.5, perim.area.ratio above 0.05 and RCC below 0.60 were considered as mangoes. Threshold values for patches size and shape indices were manually fine-tuned by visual comparison of resulted image and the same values were used for all cultivars.

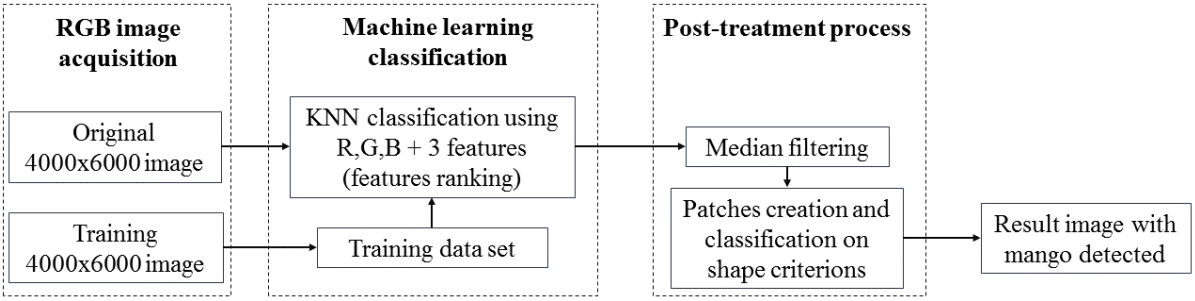


Fig. 44. Flowchart of the KNN-based algorithm.

d. Algorithm evaluation

The performance of the algorithm was evaluated by computing the precision, recall and F1 score variables (Qureshi et al., 2016). Two validation sets were built from the remaining images not used for training: Validation set 1 was composed by 42 images of Image set 1; Validation set 2 was composed by 104 images of Image set 2. In each validation image, three independent mango experts labelled all the mature fruits with bounding boxes to provide an "image-truth". Centroid position of each patches in output image were compared to all bounding box of image-truth to estimate true positives (TP), false positives (FP) and false negatives (FN). A true positive was counted when a centroid of a patch was contained in one of the bounding box. For each validation image compared with image-truth, the total number of TP, FP and FN were used to compute Precision, Recall, and F1 score as followed:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN} \qquad \text{F1 score} = \frac{2 \times P \times R}{P + R}$$

The effect of cultivars on performance variables was investigated using one-way analysis of variance (ANOVA) at a confidence level of 95%. In the case of significant effect, multiple comparisons of mean between modality were performed using Tukey's test.

e. Yield estimation

To estimate tree yield (defined as the number of fruits per tree) from algorithm detection result, the number of mature fruits was counted by field visual counting on 104 trees. To minimize estimation errors, three experts measured the field count which varied from 7 to 650 fruits per tree. These trees were photographed and images were analysed using the algorithm described above. The automatic count, was calculated as the

sum of the number of detection from both images of the tree. Half of the 104 trees (18 ‘Kent’, 18 ‘Keitt’ and 16 ‘BDH’) were used for the calibration of linear yield model for each cultivar. These models were evaluated by determining the coefficient of determination (R^2). The remaining 52 trees (17 ‘Kent’, 17 ‘Keitt’ and 18 ‘BDH’), not included in models building, were used to compute the root mean squared error (RMSE) and relative RMSE (RMSE%) of validation.

I.3. Results

a. Algorithm performances

Under homogeneous conditions of Image set 1, the KNN algorithm average precision, recall and F1 score were respectively 0.85, 0.64 and 0.73. When used under heterogeneous conditions of Image set 2, F1 score fell to 0.48. This fall is related to the decrease of both precision and recall to 0.70 and 0.36 respectively. Thus, the KNN algorithm was twice less efficient in detecting mango fruit under heterogeneous conditions than under homogeneous ones. For Image set 2, tree cultivar had highly significant effect on precision, recall and F1 score ($p < 0.001$). A Tukey’s test showed that the F1 score was significantly different between the 3 cultivars. F1 score was on average (mean + SD) 0.53 ± 0.15, 0.47 ± 0.15, and 0.28 ± 0.21 for ‘Kent’, ‘Keitt’, and ‘BDH’, respectively.

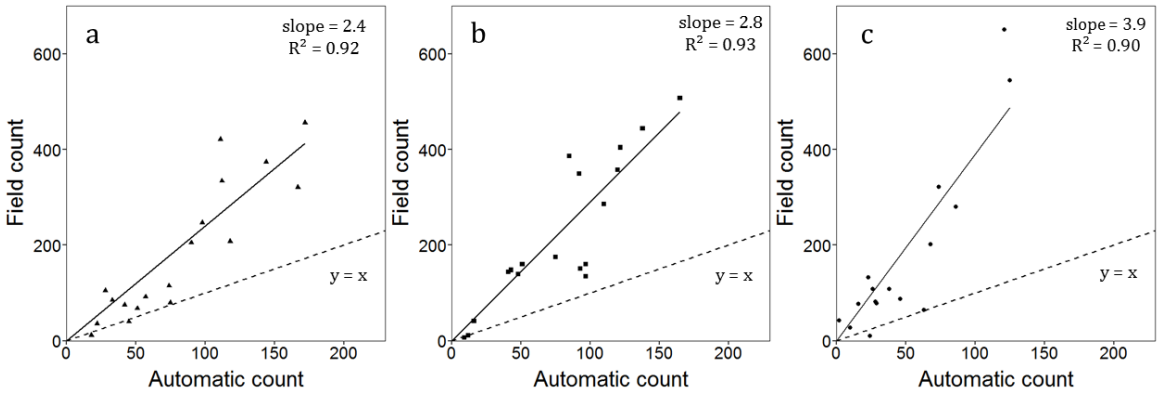


Fig. 45. Field count versus automatic count of mango fruits for (a) ‘Kent’, (b) ‘Keitt’, and (c) ‘BDH’ cultivars. The intercept of linear regression line was forced to 0.

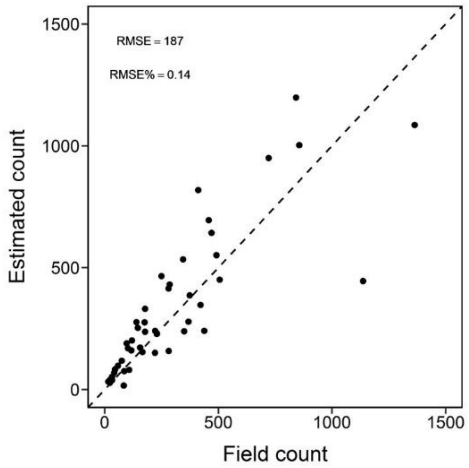


Fig. 46. Relationship between field and automatic count of mango fruits for 52 validation trees. The dotted line is the 1:1 line.

b. Linear models for yield estimation

Linear regression between automatic count from the KNN algorithm and field count reached a R^2 of 0.89 for all trees. Despite the strong relationship, RMSE showed a level of error of 83.64. By fitting a linear model for each cultivar, R^2 values reached 0.92, 0.93 and 0.90 (Fig. 45). As shown in Fig. 45, the intercept was fixed to 0 in order to avoid any negative yield estimation. For each cultivar, automatic count underestimated field count: slope values were similar for 'Kent' and 'Keitt' while it was about 1.5 times higher for 'BDH' (Fig. 45). These models were used to estimate yield of the 52 validation trees using their automatic count. Relationships between estimated yield count and field count of estimated by each model are displayed in Fig. 46 and show no significant difference of means. The RMSE reached 187 fruits and RMSE% of validation was of 14%.

I.4. Discussion

a. Fruit detection issues

Algorithm performance for the homogeneous Image set 1 (F1 score of 0.73) was similar to those found in the recent literature. Indeed, the method closest to ours was the recent study of Qureshi et al. (2016) that reached F1 score of 0.54 and 0.83 according to the image set. Similarly, they used a KNN classifier and a contour segmentation based on a geometric approach of mango shape (detected object compared to an elliptic model). The highest performance on mango detection was obtained by Bargoti and Underwood in 2017, who used a more complex algorithm involving a CNN with a F1 score of 0.90 (Bargoti and Underwood, 2017).

Compared with Image set 1 results, the lower performances obtained with the algorithm under heterogeneous Image set 2 (F1 score of 0.43) could be explained by various image related factors. Examining the outputs of the algorithm steps allowed us to identify sources of false negative and false positive. Firstly, parts of errors came from the post-treatment step which sometimes deleted mango occluded by leaves, branches or the peduncle (smaller in size or non-conventional shape) or kept false positive (object close to mango shape). This observation could be explained by the fact that threshold values for shape indices were set empirically for all cultivars and they may not have been optimal in all situation. Moreover, clustered mangoes were not accurately detected by the algorithm and contributed to increase false negative since the algorithm considered a group of mangoes as a unique fruit. Development of a special post-treatment for big patches might be envisaged to solve this issue (Linker et al. 2012).

Secondly, misclassification of fruit pixels could result in a decrease of efficiency between image sets. Indeed, we used KNN classification as it is one of the simplest machine learning method to setup. But KNN classifier performances depend on the quality of the training dataset and the input variables used, and decrease when pixel heterogeneity within class increases (Linker et al., 2012). In our case, heterogeneity in fruit colour and texture resulted from high variability in structure (size, cultivar, leave/fruit ratio…) and phenological stage of trees. For instance, reproductive asynchronism spread out fruit phenological stages in time and space, resulting in a strong heterogeneity in fruit colours on the image; orchards with various cultivar increased differences in fruit aspects compared to monocultivar orchards. This heterogeneity made the task challenging to build a training dataset good enough to capture the entire variability of fruit aspect while discriminating objects close to mango (e.g. light exposed leaves). This was probably the main raison responsible for the differences in algorithm performances between Image set 1 and 2. This assertion is especially true for 'BDH' images which have lower F1 score. Such poor result could be explained by the

natural fruit color and aspect (less blushed and slender) that make fruits more confusing with the leaves. Improvement of the training dataset by increasing pixel quantity and/or adding more input variables could have led to better results. However, applying such solutions would increase the computation time and the number of training image needed. Increasing the number of classes (e.g. add a class for leaves) could also be considered but it has been shown that it does not systematically increase classification performances (Hung et al. 2015). Recently, the use of other classification methods such as CNN (Bargoti and Underwood, 2017) have proven to be more effective tools for increasing detection performances while reducing issues related to training dataset.

b. Accuracy of yield estimation: is the image-truth true?

Generally, and as described by Linker (2017), the performance of vision-based yield estimation depends on two factors: (1) how accurately fruits are detected by the algorithm and (2) how accurately the number of fruits actually detected in the images correlates with yield. In our case, assessment of algorithm accuracy relied on image-truth built by expert annotations. However, discrepancies in fruit aspect, shape and colour resulting from the mango-related issues made the boundary between immature and mature fruits difficult to visually identify on RGB images by the annotators. Consequently, algorithm performance is highly correlated with human subjective representation and perception of a mango. For instance, improving the F1 score is possible by imposing experts to annotate only the fruits that are easily detectable (i.e. fruits blushed and clearly visible).

On the other hand, this study permitted to build three reliable linear models ($R^2 > 0.90$, RMSE% = 14%) for tree yield estimation. The slightly higher slope for 'BDH' cultivars compared to the 2 others (Fig. 45) could be explain by its particular canopy structure. Indeed, 'BDH' trees have higher proportion of fruits inside the canopy affecting the ratio between external fruits - detectable - and internal fruits. This different canopy structure may be due to physiological traits of the cultivar combined to the fact that 'BDH' trees are exclusively grown in extensive non-pruned orchards which contribute to densify canopy.

Hence, despite its low performances, the machine vision system revealed to be an efficient tool for yield estimation when combined with linear models. This tools is adapted for yield estimation under heterogeneous conditions and for the three main cultivars cultivated in the Niayes region. The advantages of this machine vision system is that it relies only on two opposite images of the tree taken with a commercial camera. Additionally, the proposed tool is implemented in the open source R software, and requires low computing resources. One way to improve yield estimation models would be to reduce RMSE mainly by reducing detection errors of the algorithm (e.g. improve the training dataset, use a more performant classifier, etc.). These results discussed the assertion made by Linker (2017) as it raises the following question: is it mandatory to reach the highest performances for machine vision system to accurately estimate tree yield?

I.5. Conclusion

By combining a machine learning algorithm trained on heterogeneous field conditions with manual fruit count on a large number of trees, this work provided an effective and reliable tool for mango yield estimation from RGB images. This tool is based on strong models ($R^2 > 0.90$, RMSE% = 14%) which allow us to quantify mango yields of the three main cultivars in the Niayes region of Senegal. However, yield estimation using colour imaging still remains a challenging task when using KNN machine learning for heterogeneous

field conditions (algorithm construction, variability of fruit aspect, and annotator effects). Further analysis will now require to investigate the relationship between yield estimates and easily measurable tree parameters of structure (tree height, canopy volume and cultivar…) to being able to extrapolate yield models at the orchard scale.

Acknowledgements

This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

References

Baker, W.L., and Cai, Y. (1992). The r. le programs for multiscale analysis of landscape structure using the GRASS geographical information system. Landsc. Ecol. 7, 291–302.

Bargoti, S., and Underwood, J. (2017). Deep fruit detection in orchards. In Robotics and Automation (ICRA), 2017 IEEE International Conference On, (IEEE), pp. 3626–3633.

Dambreville, A., Lauri, P.-E., Trottier, C., Guedon, Y., and Normand, F. (2013). Deciphering structural and temporal interplays during the architectural development of mango trees. J. Exp. Bot. 64, 2467–2480.

Diame, L., Blatrix, R., Grechi, I., Rey, J.-Y., Sane, C.A.B., Vayssieres, J.-F., de Bon, H., and Diarra, K. (2015). Relations between the design and management of Senegalese orchards and ant diversity and community composition. Agric. Ecosyst. Environ. 212, 94–105.

Gitelson, A.A., Kaufman, Y.J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87.

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., and Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. Comput. Electron. Agric. 116, 8–19.

Grechi, I., Sane, C.A.B., Diame, L., De Bon, H., Benneveau, A., Michels, T., Huguenin, V., Malezieux, E., Diarra, K., and Rey, J.-Y. (2013). Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68, 447–466.

Hung, C., Underwood, J., Nieto, J., and Sukkarieh, S. (2015). A feature learning based approach for automated fruit yield estimation. In Field and Service Robotics, (Springer), pp. 485–498.

Linker, R. (2017). A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation. Precis. Agric. 18, 59–75.

Linker, R., Cohen, O., and Naor, A. (2012). Determination of the number of green apples in RGB images recorded in orchards. Comput. Electron. Agric. 81, 45–57.

Mukherjee, S.K., and Litz, R.E. (2009). Botany and Importance. In The Mango: Botany, Production and Uses, R.E. Litz, ed. (Cambridge, USA: CABI), pp. 1–18.

Payne, A.B., Walsh, K.B., Subedi, P.P., and Jarvis, D. (2014). Estimating mango crop yield using image analysis using fruit at ‘stone hardening’ stage and night time imaging. Comput. Electron. Agric. 100, 160–167.

Qureshi, W.S., Payne, A.B., Walsh, K.B., Linker, R., Cohen, O., and Dailey, M.N. (2016). Machine vision for counting fruit on mango tree canopies. Precis. Agric. 18, 224–244.

R Core Team (2015). R : A language and environment for statistical computing (Vienna, Austria).

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., and Storlie, C. (2014). SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises.

II. Deep Mangoes: from fruit detection to cultivar identification in colour images of mango trees

Les travaux sur le KNN (Chapitre 1 – Partie I) ont permis de démontrer que cet algorithme simple de *machine learning* s'avérait inefficace dans nos conditions hétérogènes (plusieurs systèmes de culture, plusieurs cultivars, structure des arbres). Cette Partie II présente les travaux sur l'algorithme de *deep-learning*¸ le Faster R-CNN, utilisé pour la détection du nombre de fruits dans l'arbre. Cet algorithme nécessite la construction d'une base de données labellisées (fruits dans des boîtes englobantes) utilisée dans une phase d'entrainement. Malgré des contraintes liées au temps nécessaire pour entrainer l'algorithme, ses performances supérieures au KNN, en termes de temps et d'efficacité de détection, font qu'il a été utilisé pour la suite des travaux de thèse.

Cette partie est composée d'un second acte de conférence présenté à l'*International Conference on Digital Image and Signal Processing* (DISP'19) à Oxford en 2019 et publié dans les archives ouvertes arXiv (Cornell University).

- Borianne P.; Sarron J.; Borne F.; Faye É. **Deep Mangoes: from fruit detection to cultivar identification in color images of mango trees**. DISP'19, 2019, Oxford, UK, arXiv:1909.10939.

Se trouve également en annexe un poster de synthèse sur l'outil d'estimation de la production de mangues à l'échelle de l'arbre (Annexe III) et présenté lors des journées du CIRAD en juillet 2018.

- Sarron J.; Sané C.A.B.; Borianne P.; Diatta P.; Faye É. **PixYield – L'analyse d'images photographiques comme outil pour l'estimation de rendement : le cas du manguier au Sénégal**

Deep Mangoes: from fruit detection to cultivar identification in colour images of mango trees

Philippe Borianne[1,2], Frédéric Borne[1,2], Julien Sarron[3,4,5], Émile Faye[3,4,5]

[1]CIRAD, UMR AMAP, F-34398 Montpellier, France
[2]AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France
[3]CIRAD, UPR HortSys, F-34398 Montpellier, France
[4]HortSys, Univ Montpellier, CIRAD, Montpellier, France
[5]Centre pour le Développement de l'Horticulture, ISRA, Dakar, Senegal

Abstract

This paper presents results on the detection and identification mango fruits from colour images of trees. We evaluate the behaviour and the performances of the Faster R-CNN network to determine whether it is robust enough to "detect and classify" fruits under particularly heterogeneous conditions in terms of plant cultivars, plantation scheme, and visual information acquisition contexts. The network is trained to distinguish the 'Kent', 'Keitt', and 'Boucodiekhal' mango cultivars from 3,000 representative labelled fruit annotations. The validation set composed of about 7,000 annotations was then tested with a confidence threshold of 0.7 and a Non-Maximal-Suppression threshold of 0.25. With a F1-score of 0.90, the Faster R-CNN is well suitable to the simple fruit detection in tiles of 500x500 pixels. We then combine a multi-tiling approach with a Jaccard matrix to merge the different parts of objects detected several times, and thus report the detections made at the tile scale to the native 6,000x4,000 pixel size images. Nonetheless with a F1- score of 0.56, the cultivar identification Faster R-CNN network presents some limitations for simultaneously detecting the mango fruits and identifying their respective cultivars. Despite the proven errors in fruit detection, the cultivar identification rates of the detected mango fruits are in the order of 80%. The ideal solution could combine a Mask R-CNN for the image pre-segmentation of trees and a double-stream Faster R-CNN for detecting the mango fruits and identifying their respective cultivar to provide predictions more relevant to users' expectations.

Keywords: Faster R-CNN, mango fruit detection, mango cultivar identification, neural network

II.1. Introduction

The estimation of pre-harvest agricultural production is essential to meet development challenges and reduce the vulnerability of populations to global changes. Indeed, one of the main issues hindering the development of perennial crops is the impossibility of early, easily, and accurately estimating crop yields in order to guide farm management effectively. To date, yield estimation in tropical orchards is still based on a visual inspection of a limited sample of trees, a tedious, time- and cost-consuming method that depends on the reliability and accuracy of the observer (Linker et al. 2012). The cultivation of mango (*Mangifera indica L.*) in West Africa, and more particularly in Senegal, needs particular attention because of the physiological specificities of trees, especially the reproductive asynchronism and the inter-tree heterogeneity (Dambreville et al. 2013). Moreover, West African farmers require tools adapted to the conditions of their family-sized and diversified cropping systems that can provide them with production information in a simple, inexpensive and convenient way.

Over the last few years, several studies described efficient machine vision systems for fruit detection and fruit yield estimation (Gongal et al. 2015). These systems ranged from simple colour pixel segmentation to more advanced machine learning method using combination of colour, shape and texture features acquired by multiple sensors. Moreover, they were usually developed and evaluated in orchards under homogeneous conditions in terms of plant cultivars, plantation schemes (density, row, etc.), cultivation practices and visual information acquisition methods (Payne et al. 2014; Hung et al. 2015; Qureshi et al. 2016). Under heterogeneous conditions, the performance of machine learning approaches drops significantly (Sarron et al. sub.).

One of the major advances in the object localization and detection was the multi-scale sliding window algorithm (Sermanet et al. 2014; Guo et al. 2016) for using Convolutional Neural Networks (CNNs). Regions with CNN features or R-CNN (Girshick et al. 2014) improved by almost 50% the detection performance by combining the object extraction using a Selective Search (Uijlings et al. 2013) and the region classification by SVMs (Adankon and Cheriet 2015). This key step was published under the name Fast R-CNN (Girshick 2015). As the R-CNN, it used Selective Search to extract possible objects, but instead of typing them using SVM classifiers, it applied the CNN on all the image and then used both Region of Interest (RoI) and Pooling on the feature map with a final feed forward network for classification and regression. The RoI Pooling layer and the fully connected layers allowed the model to be end-to-end differentiable and easier to train. Faster R-CNN (Ren et al. 2015; Chen and Gupta 2017) added a Region Proposal Network (RPN), in an attempt to do without the Selective Search algorithm and make the model completely trainable end-to-end.

Different works used deep CNN for yield estimation (Hou et al. 2016) and/or fruit counting (Rahnemoonfar and Sheppard 2017), and especially based on the Faster R-CNN (Sa et al. 2016; Dore et al. 2017). Despite the asynchronous production cycles of mango trees, this neural network obtained high performances for fruit detection and counting at plot scale in homogeneous field conditions (Stein et al. 2016; Bargoti and Underwood 2017).To our knowledge, no work has been done to identify the stage of development of the fruits and their cultivar, which is assumed to be known. However, these two aspects are essential for estimating economic yields and forecasts of fruit farms, particularly on family farms that are increasingly using multiple grafts: several fruit cultivars are simultaneously grafted on the same tree. One of the major interests of this type of practice is to be able to diversify the production on a small area in order to make the farmer less dependent on the fluctuations of economic markets while preserving biodiversity and the health balance of the crops.

Besides mango detection in heterogeneous mango orchards, this study presents the results of fruits classification among mango cultivars. Indeed, visual images of mango trees often encompassed fruits carried by other nearby trees (adjacent or in background), and fruits too heavily occluded so that their status is difficult to assess. In order to provide farmers with the most accurate results of the number of fruits that will actually be produced per tree, we investigated the capacities of a Faster R- CNN network to classify the fruit condition. The evaluation of the fruit detection and identification is usually done using the standard Mean Average Precision (Powers 2011). Among these indicators, we used the F1-score well suited to the statistical comparison with a real "image truth". Different notions were introduced, for example the global Jaccard indicator for qualifying the geometric fitting between expert annotations and network predictions or the multi-tiling for aggregating multiple detections. The detection reporting from tiles to images is also described and discussed.

II.2. Materials and methods

a. The Faster R-CNN

Faster R-CNN started with 500x500 pixel pre-cut images called tiles and provided a list of labelled bounding boxes with prediction probabilities. Its architecture is based on three components (Fig. 47). First, a Region Proposal Network (RPN) is used to find up to a predefined number of regions that may contain objects. Fixed sized reference bounding boxes are placed uniformly throughout the original image. They give a list of possible relevant objects and their locations in the original image. Second, a Region of Interest Pooling (RoIP) is applied using the features extracted by the CNN in the previously given bounding boxes to extract those features which would correspond to the relevant objects. This step significantly accelerates the identification of the interest areas. Third, the R-CNN module uses that information to classify the content in the bounding box (or discard it using the "background" label) and adjust the bounding box coordinates (so it better fits the object).

This method involves using a CNN pre-trained for the task of classification and then re-trained for the recognition of mangoes. No real consensus exists on the best network architecture for the pre-training step, i.e. between different networks with a varying number of weights. Even if the ZF (Zeiler and Fergus 2014), MobileNet (Howard et al. 2017) and DenseNet (Huang et al. 2017) pre-training seem attractive, we used the VGG (Simonyan and Zisserman 2015) pre-trained on ImageNet (Deng et al. 2009) because it was dedicated to the large-scale image recognition. The network has been fine-tuned from a learning set of annotated tiles, where the annotations are bounding boxes surrounding visible (part of) mango fruits and labelled according to the defined classes.

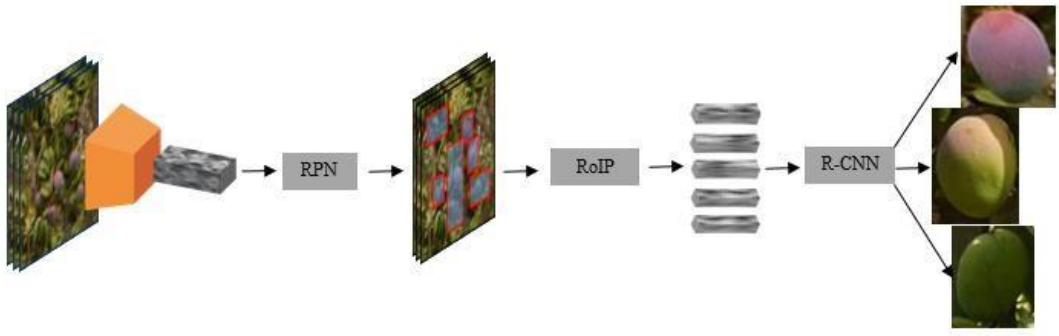


Fig. 47. Complete Faster R-CNN architecture

b. Key network parameters

The confidence score. The network gives each box the probability of containing an object. The confidence-threshold is set to detect only boxes with high certainty.

The non-maximum suppression (Bodla et al. 2017). The Non-Maximum Suppression (NMS) pre-vents multiple detection of the same object. The NMS-threshold is set to define the non-significant intersections of boxes.

The iteration number. The main mechanism of networks, called gradient descent, is based on an iterative minimum search algorithm: the algorithm needs to be repeated to converge to the best possible solution. The number of iterations (given as a training parameter) indirectly defines the number of times the network observes the data.

c. Matching Function

Matching function consists in associating two by two the expert annotation and network prediction boxes. The optimum matching between the expert annotations A_i and the network predictions P_j was given by the Jaccard matrix (Bouchard et al. 2013); the $J_{i,j}$ Jaccard index was defined as the ratio of areas between both the intersection and the reunion of two boxes (Eq. 3). Matches were made in descending order of Jaccard indices as long as greater than 0.25. Once fitted, the boxes were removed from the list of possible candidates and the process was repeated until all the boxes have been matched or no box could be matched.

$$J_{i,j} = \frac{area(A_i \cap P_j)}{area(A_i \cup P_j)} \quad (3)$$

The algorithmic complexity of the matching was in O(IJ) where I and J respectively represented the maximum numbers of annotations and predictions per tile (which visibly does not exceed 20 objects).

d. Merging Function

Image tiling, i.e. partitioning the image into tiles which can be processed by the neuronal network, is necessary for the detection of "small" objects in "large" images. Without tiling, small objects "disappear" during the geometric transformations applied for data formatting. If the recomposition of the image is simply achieved by juxtaposing the tiles, the aggregation of detected objects is potentially more complex, especially for objects "cut" into several parts by the tiling. When sufficiently significant i.e. quite close to the examples given in the training data set, these different parts are respectively detected, and the object is counted several times (Fig. 48). When not significant enough, no part of the object is detected and the object is not recorded. The multiple detections of the same object appearing on several tiles must be merged to adjust as well as possible the count and the shape of the boxes characterizing this last one.

We proposed an approach based on the following assumption: “The size of the objects to be detected is significantly smaller than those of the tiles used. Consequently, there should be at least one tile in which the object appears integrally”. We used so different tiling, i.e. ways to partition the native image. For a given tiling, the objects validated as detected (or detections) were those whose bounding box was not incident at any edges of the tile that contained it. Objects not checking this condition were not retained: they were detected in another tiling, more exactly in a more appropriate tile containing them integrally. The problem was thus to merge detections coming from different tiling while taking care not to count more than once a given object.

The matching function previously described, based on the use of a Jaccard matrix, was used here to identify matches between objects detected in two different tiling. In case of matching, the smaller bounding box was removed. Objects that find no homologue correspond to the identification of "integral" objects in one of the tiling and "cut" objects in the other.

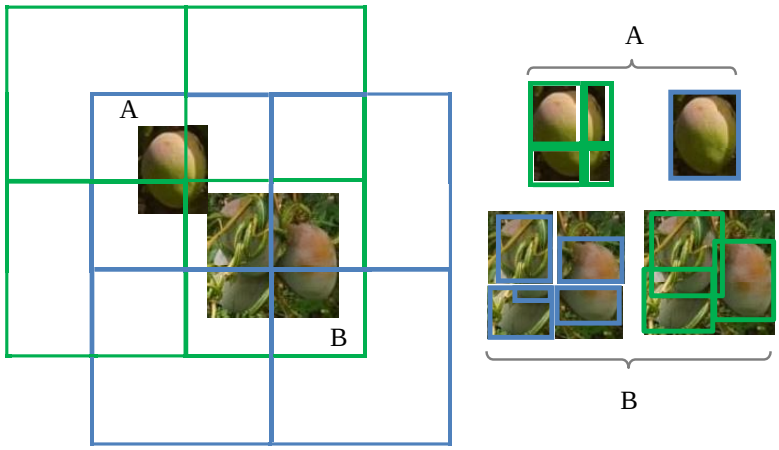


Fig. 48. Image tiling and object detection. A- The green tiling splits the object into 4 parts which are each detected by the network; in the blue tiling, the object is detected only once in its entirety. B- The blue tiling breaks 3 objects into 4 parts leading to 5 detections; in the green tiling, the 3 objects are correctly detected.

e. Indicators

The statistical indicator. We used the F1-score (Eq. 4) as test's accuracy between 'image truth' and network predictions; it is defined as the best compromise between precision (Eq. 5) and recall (Eq. 6), two statistical indicators taking into account respectively the numbers of "True Positive", "False Positive" and "False Negative" classes where :

- "True Positives" (TP) are mangoes annotated by the expert and correctly detected by the network,
- "False Negatives" (FN) are mangoes annotated by the expert and not detected by the network,
- "False Positives" (FP) are mangoes detected by the network but not considered as such by the expert.

$$F1score = \frac{2 \times precision \times recall}{precision + recall} \quad (4)$$

$$precision = \frac{number\ of\ mangoes\ correctly\ detected}{total\ number\ of\ detected\ mangoes} = \frac{VP}{VP + FP} \quad (5)$$

$$recall = \frac{number\ of\ mangoes\ correctly\ detected}{total\ number\ of\ expert\ annotated\ mangoes} = \frac{VP}{VP + FN} \quad (6)$$

The geometric fitting. A global Jaccard index (Eq. 7) was introduced to for quantifying the difference between the geometry of the A_i expert annotations and the P_i network predictions, i.e. the precision of the geometric fitting of the bounding boxes of mangoes.

$$J_\alpha = \frac{\sum_i area(A_i \cap P_j)}{\sum_i area(A_i \cup P_i) + \alpha \sum_j area(A_j) + \alpha \sum_k area(P_k)} \quad (7)$$

where $i \in \{VP\}, j \in \{FN\}, k \in \{FP\}$. An index of 1 will be the perfect superposition of annotations and prediction boxes.

II.3. Results and discussion

The main question is whether the network is sensitive enough to identifying the cultivar of the mango fruits it detects. We used 150 native colour images of 'Kent', 'Keitt', and 'Boucodiekhal' (BDH) mango cultivars taken at a distance of 5 meters by a Sony Nex-7 RGB camera with a fixed focal length of 18 mm. The 4000x6000 pixel native images were cut in 500x500 pixel tiles which were manually annotated and labelled under ImageJ (Schneider et al. 2012) to form a representative set of about 10,000 annotations in terms of shape, colour, sunlight conditions or occlusions. The labelled annotations were rectangular boxes, from 10 to 80 pixels side, including as closely as possible the visible fruits on the images: labels specified mango cultivar. These annotated data have been distributed as follows: 3,000 heterogeneous annotations for network training, 7,000 for cross validation.

a. Parameter setting

Various experiments were conducted to determine the best settings, i.e. the ones that maximizes the performances of the network in detection mango fruit. The same set-tings were used for fruit cultivar identification.

The confidence and NMS levels. The fruit detection network was trained with an arbitrarily number of 50,000 iterations. The validation set was then tested by simultaneously varying the confidence and NMS thresholds from 0.35 to 0.9 and 0.005 to 0.5 respectively in steps of 0.05 (Fig. 49). The F1-score (Eq. 4) ranged from 0.66 for the threshold- couple (0.35, 0.5) to 0.90 – in fact 0.899687 – for (0.7, 0.25). Different settings led to relatively close performances. We chose the setting resulting in the highest performances, acknowledging that this result certainly depended on both the training and the validation sets used. Experimentation could have been further developed by studying the performance of the cultivar identification network to find the optimum threshold pair values, but we considered that the optimum parameter setting of the fruit detection network could be transposed to the cultivar identification network.

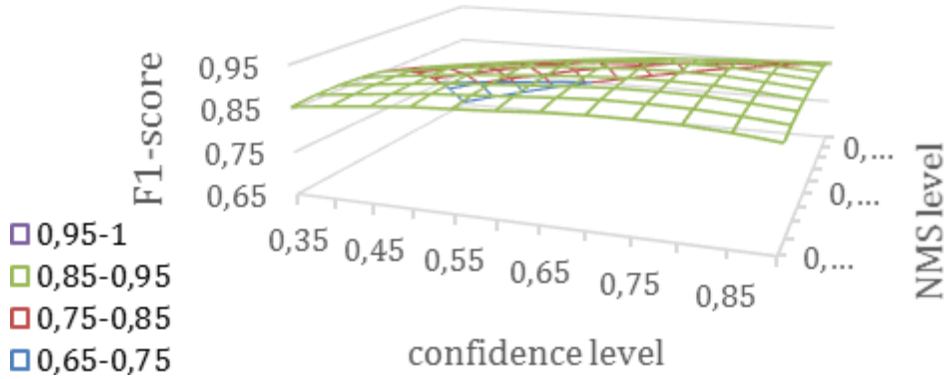


Fig. 49. Variation in the network performance according to the confidence and NMS thresholds.

The number of iteration. The network was trained with numbers of iterations ranging from 500 to 30,000 in steps of 500. The validation set was then tested with a confidence threshold of 0.7 and a NMS threshold of 0.25 (Fig. 50).

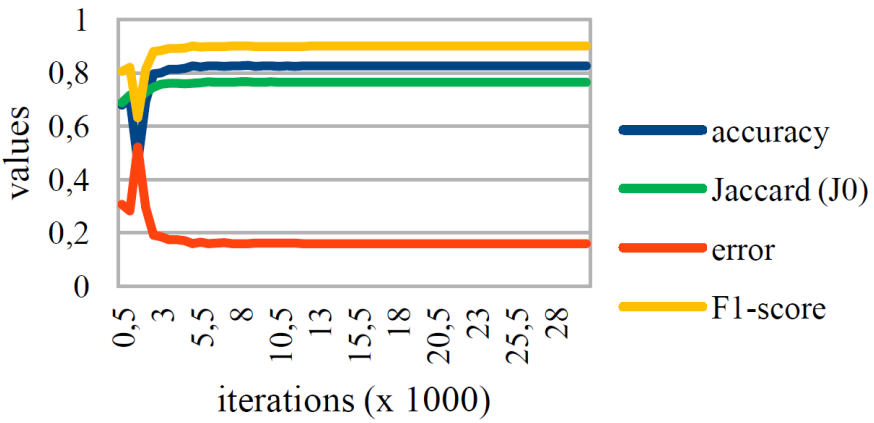


Fig. 50. Variation in the network performances according to the number of iterations.

Error and accuracy were indicative: they represented respectively the percentage of 'correct' and 'incorrect' predictions by the network. The functions characterizing the network behaviour were strongly correlated and were stabilized from 15,000 iterations.

The size of the training data set. The network was trained with an iteration number of 15,000 and a size of the training data set ranging from 200 to 5,000 by steps of 200. The validation set was then tested with a confidence threshold of 0.7 and a NMS threshold of 0.25 (Fig. 51). The network performance reached its maximum after 2,650 training annotations: the accuracy of the geometric fitting did not exceed 77%, even when significantly increasing the size of the training data set. A training set of 3,000 annotations reinforced the stability of the detection process.



Fig. 51. Variation in the network performances according to the size of the training data set.

b. Localization and identification of fruits

In the orchards of Western Africa, fruit plots often mix different cultivars of fruit for a better use of soils and to spread the fruit production along the year. Estimating mango yield at the orchard scale requires assessing the production of each cultivar grown, i.e. the number of fruits produced (or to be produced) per cultivar and tree. For all the experiments below, the network was fine-tuned with 3,000 annotations and 15,000 iterations. The validation set was then tested with a confidence threshold of 0.7 and a NMS threshold of 0.25.

The mango detection at the tile scale. This experiment aimed to ensure that the network can satisfactorily locate mangoes of various cultivars and stages of development in images of full trees. The F1-score on the validation set was 0.90 for a J0 fitting precision of 0.77 (Fig. 50): the geometric offset between annotations

and predictions could appear high, but it was a very satisfactory result considering the heterogeneity of the data set. In Fig. 52, the expert annotations (in red) were almost superimposed on the network predictions (in blue).

The in-depth study of False Negatives pointed out the network's difficulties in detecting partially occulted fruits, essentially in fruit clusters (Fig. 53.A) and fruits inside canopy (Fig. 53.B). In the first case, the NMS threshold prevented 'correct' detection of background fruits partially masked by foreground fruits. In the second case, fruits were rather hidden by leaves and wood, causing a significant change in the signature of the mango. On the False Positive side, most of the residual confusion concerned leaf arrangements with curvatures similar to those of mangoes (Fig. 53.C). Interestingly, the network detected fruits not annotated by the expert (Fig. 52.B and 53.D): in absolute terms, they were false False-Positives which penalized the network performance.

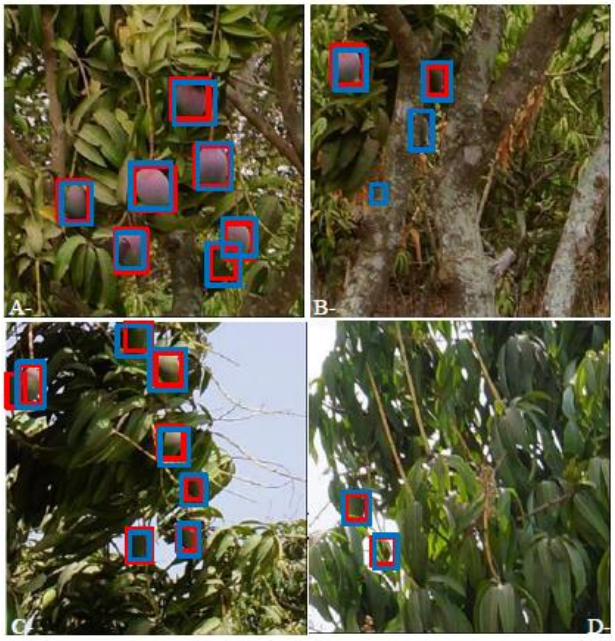


Fig. 52. Detection of mangoes in tree tiles. Red boxes were expert annotated fruits, and blue boxes network detected fruits. The network detected visible mangoes well, regardless of light exposure or fruit growing stage. The pictures shown tiles of mango trees seen at 5 meters distance for 'Keitt' (A, B) and 'Kent' (C, D) cultivars.

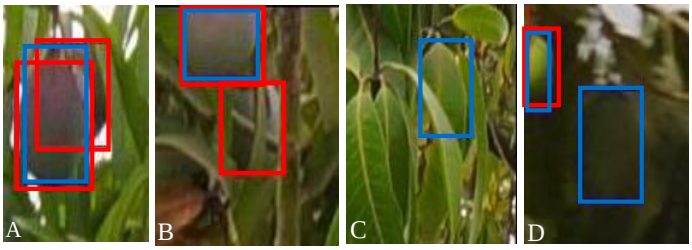


Fig. 53. Detection errors. False Negatives are red boxes, False Positives blue boxes. A- a background mango was not detected due to the application of the NMS threshold; B- mango partially hidden by foliage was not detected because this situation is not sufficiently represented in the training data set. C- a leaf was confused with a mango by the network. D- a mango inside the tree was detected by the network even though it had not been annotated by expert.

The mango detection at the full image scale. Four tiling were used to cover a maxi-mum of cutting configurations, without however being able to guarantee that they are all well considered: the 1st is a

partition in tiles of 500x500 pixels starting from the image point coordinates (0,0), the second from the point (0,250), the third from the point (250,0) and the fourth from the point (250,250).

The F1-score was 0.71 for a J0 fitting precision of 0.67. This drop in performance is explained by the significant increase in False Positives (Fig. 54) compared to the previous experiment: the expert limited his count to the foreground tree while the network processed all the trees in the image. Previously, only tiles in which the expert had observed a fruit were used, which partially excluded fruits from adjacent trees; here, all tiles in the images were used.

The algorithmic complexity of the detection merging is in $O(n^2)$ where n represents the number of fruits visible in the image: it is fixed by the calculation of the Jaccard matrix. It can look high even if it only takes 30" to process a native image on a Dell Studio XPS 8 x i7 3 GHz with a Nvidia Quadro M2000 graphics card. Further study to determine the optimal tiles to use for the fusion of detected fruits would be of definite theoretical interest, even if the best way to reduce computation times would be to use adaptive tiles, i.e. tiles centred on "clipped" objects: but such an approach would require two instantiations of the neural network. The major problem is to reduce the exploration area of the network to the foreground tree. The YOLO network (Redmon et al. 2016) is currently under study to pre-zone the image and exclude background trees: to refine the results, it will probably be necessary to use (in addition or not) segmentation networks such as SegNet (Badrinarayanan et al. 2016) or similar (Garcia-Garcia et al. 2017). For example, Mask-RCNN (He et al. 2017) customizes the Faster-RCNN with an additional parallel branch that makes it powerful for both the bounding-box object detection and the object segmentation.



Fig. 54. Detection of mangoes in ‘Keitt’ tree image. Red boxes were expert annotated fruits, blue boxes network detected fruits. Unlike the network, the expert did not consider fruits of the background trees.

Identification of mango cultivars. For this experiment, the network was trained to identify the 'Kent', 'Keitt' and 'Bdh' cultivars composing the data set (Fig. 55).For the learning step, each cultivar class was described by about 1,000 labelled annotations. In identifying cultivars, the F1-score was to 0.56 for a J0 fitting precision of 0.71. This significant drop in performances is explained by the high increase in False Negatives, i.e. mangoes not detected: barely 4,000 mangoes were detected by the network during the cross-validation while the data set contains nearly 7,000 expert annotations.

When the tested fruit looks alike fruits of at least two variety classes, the network assigns predictive probabilities (by class) to it that are quite close and finally lower than the confidence threshold used by the

network for the detection phase. Consequently, the object was not considered as a fruit and is therefore not detected. (Fig. 56, blue rectangles). Lowering the confidence threshold may in some cases be an option, but it might significantly increase the number of False Positives. The best option would probably be to use a two-stream faster R-CNN (Peng and Schmid 2016) to firstly detected all the fruits of the image and then qualify them in classes.

Fig. 56 shown the behaviour of the identification network: mangoes correctly detected and identified by the cultivar identification network were represented by green rectangles, mangoes correctly detected but incorrectly identified by red rectangles, and mangoes not detected by the identification network although counted by a simple detection network by blue rectangles.

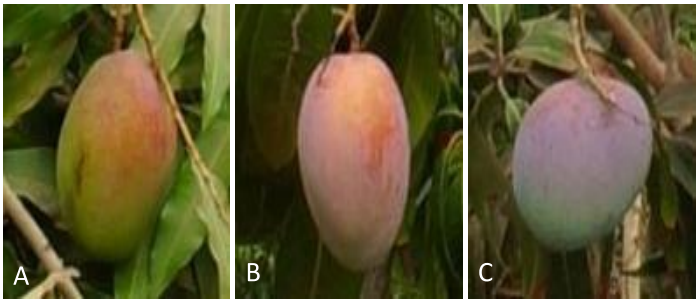


Fig. 55. Mango cultivars at the pre-harvest stage. A- 'BDH' mango fruit is ovoid-flat with a skin col-our ranging from red to green and an average weight of 450 grams. B- 'Keitt' mango fruit is ovoid-oblong; skin color is pink with less than 30% red and the average weight is 500 to 600 grams. C- 'Kent' mango fruit is rather ovoid-wide with yellow skin and red spot; the average weight does not exceed 550 grams.

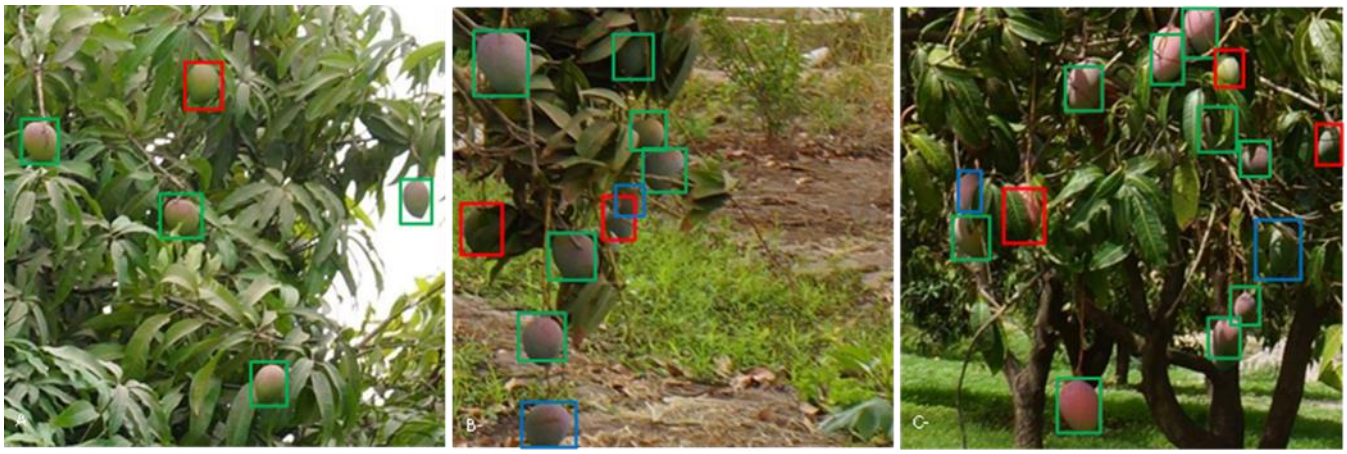


Fig. 56. Identification errors of 'BDH' (A), 'Keitt' (B) and 'Kent' (C) cultivars.

Errors in identification probably have several causes. Among them, the physiological features are not negligible. Since the expression of the really discriminating features of fruits only appears after the juvenile stages, there would therefore be a link between the cultivar identification certainty and the fruit development stages. It is therefore necessary to know / be able to identify the stage of development of the fruits to certify the detected fruit cultivars. This is all the more important in the case of grafted trees that carry several fruit cultivars. Moreover, angles of view or partial occlusions of fruits probably have a greater impact on the responses of an identification network than on those of a detection network.

Beyond this observation, the question was how many correctly detected mangoes were well identified (classified). Table 2 gave the identification percentages, class by class, of correct detected fruits (900 'BDH' mangoes detected on 2,000 annotated, 1,200 'Keitt' on 2,500 and 1,600 'Kent' on 2,500). The identification

rates without error; i.e. the detected fruits attributed to the right cultivar class, ranged from 80 to 90% (see bold values of the Table 2). The identification errors were caused by visual similarities in mangoes, especially when partially masked, viewed from irrelevant angles or at too early stages of development.

Table 2. Multi-class identification rate. *Each column indicates how the neural network identifies the fruits of a given cultivar.*

		Expert labelling		
		Bdh	**Keitt**	**Kent**
Network labelling	**Bdh**	**86.9**	2.4	1.0
	Keitt	2.6	**78.2**	6.6
	Kent	10.5	19.4	**92.4**

Some questions remain open: for example, why is there 3 times more confusion between the Keitt and Kent fruits than between Kent and Keitt fruits: a simple conjuncture or a deeper network learning problem? It is obvious that, unlike adult or mature fruits, young fruits have very similar aspects, perhaps too similar, so the network can unambiguously identified their respective cultivar. But if that were possible, the solution would surely based on a Ensemble Neuronal Network (Kumar et al. 2017; Zameer et al. 2017) - i.e. on the simultaneous use of several different networks fine-tuned on a same dataset or on several identical networks trained on different data for significantly increasing the relevance of fruit cultivar identification unless this qualification is limited to sufficiently mature fruit.

II.4. Conclusion

We evaluated the behaviour of the Faster R-CNN network to determine whether it was robust enough to "detect and identify" fruits under particularly heterogeneous conditions in terms of tree cultivars, plantation scheme, and visual information acquisition contexts. The network was trained using 3,000 representative labelled annotations of 'Kent', 'Keitt', and 'Boucodiekhal' mango cultivars. The validation set composed of about 7,000 labelled annotations was tested with a confidence threshold of 0.7 and a Non- Maximal-Suppression threshold of 0.25.

The network accuracy (F1-score) was 90% for fruit detection, but fell to 56% for fruit cultivar identification. When the tested fruit looks alike fruits of at least two variety classes, the network assigns predictive probabilities (by class) to it that are quite close and finally lower than the confidence threshold used by the network for the detection phase. Consequently, the object was not considered as a fruit and is therefore not detected.

The cultivar identification rates of the detected mango fruits were in the order of 80%. Errors were caused by visual similarities in mangoes, especially when partially masked, viewed from irrelevant angles or at too early stages of development.

Image tiling, i.e. partitioning the image into tiles which can be processed by the neuronal network, is necessary for the detection of "small" objects in "large" images. Tiling may lead to "artificially" clipped objects, which will result in multiple detection of these objects. We combined a multi-tiling approach with a Jaccard matrix to identify and merge multiple detections and thus report the detections made at the tile scale to the native image.

The cross validations showed the need to undertake additional reflexion to make the predictions of the R-CNN Faster network more relevant to users' expectations. Future works could focused on (i) the image pre-processing to separate neighbouring trees from each other (using convolutional segmentation networks such

as SegNet or Mask R-CNN) and, (ii) the two-stream Faster R-CNN developing to identify the cultivar of all detected fruits on each segmented tree.

Acknowledgments: This work was supported by the French National Research Agency under the Investments for the Future Program (ANR-16- CONV-0004) and by the PixYield Creativity and Scientific Innovation action funded by CIRAD, the French agricultural research and international cooperation organization. The authors wish to thank Dr. Sané C.A.B for its expert work and active participation in the annotation of visual data.

References

Adankon MM, Cheriet M (2015) Support Vector Machine. In: Li SZ, Jain AK (eds) Encyclopedia of Biometrics. Springer US, Boston, MA, pp 1504–1511

Badrinarayanan V, Kendall A, Cipolla R (2016) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. ArXiv151100561 Cs

Bargoti S, Underwood J (2017) Deep fruit detection in orchards. In: Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, pp 3626–3633

Bodla N, Singh B, Chellappa R, Davis LS (2017) Soft-NMS — Improving Object Detection with One Line of Code. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, pp 5562–5570

Bouchard M, Jousselme A-L, Doré P-E (2013) A proof for the positive definiteness of the Jaccard index matrix. Int J Approx Reason 54:615–626. https://doi.org/10.1016/j.ijar.2013.01.006

Chen X, Gupta A (2017) An Implementation of Faster RCNN with Study for Region Sampling. ArXiv170202138 Cs

Dambreville A, Lauri P-E, Trottier C, et al (2013) Deciphering structural and temporal interplays during the architectural development of mango trees. J Exp Bot 64:2467–2480. https://doi.org/10.1093/jxb/ert105

Deng J, Dong W, Socher R, et al (2009) ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, FL, pp 248–255

Dore A, Devy M, Herbulot A (2017) Détection d'objets en milieu naturel: application à l'arboriculture. In: 16ième Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur (ORASIS 2017). Colleville-sur-Mer, France, p 8

Garcia-Garcia A, Orts-Escolano S, Oprea S, et al (2017) A Review on Deep Learning Techniques Applied to Semantic Segmentation. ArXiv170406857 Cs

Girshick R (2015) Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, pp 1440–1448

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Columbus, OH, USA, pp 580–587

Gongal A, Amatya S, Karkee M, et al (2015) Sensors and systems for fruit detection and localization: A review. Comput Electron Agric 116:8–19. https://doi.org/10.1016/j.compag.2015.05.021

Guo Y, Liu Y, Oerlemans A, et al (2016) Deep learning for visual understanding: A review. Neurocomputing 187:27–48. https://doi.org/10.1016/j.neucom.2015.09.116

He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, pp 2980–2988

Hou L, Wu Q, Sun Q, et al (2016) Fruit recognition based on convolution neural network. In: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE, Changsha, China, pp 18–22

Howard AG, Zhu M, Chen B, et al (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv170404861 Cs

Huang G, Liu Z, Maaten L van der, Weinberger KQ (2017) Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp 2261–2269

Hung C, Underwood J, Nieto J, Sukkarieh S (2015) A feature learning based approach for automated fruit yield estimation. In: Field and Service Robotics. Springer, pp 485–498

Kumar A, Kim J, Lyndon D, et al (2017) An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. IEEE J Biomed Health Inform 21:31–40. https://doi.org/10.1109/JBHI.2016.2635663

Linker R, Cohen O, Naor A (2012) Determination of the number of green apples in RGB images recorded in orchards. Comput Electron Agric 81:45–57. https://doi.org/10.1016/j.compag.2011.11.007

Payne A, Walsh K, Subedi P, Jarvis D (2014) Estimating mango crop yield using image analysis using fruit at 'stone hardening' stage and night time imaging. Comput Electron Agric 100:160–167. https://doi.org/10.1016/j.compag.2013.11.011

Peng X, Schmid C (2016) Multi-region Two-Stream R-CNN for Action Detection. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp 744–759

Powers DMW (2011) Evaluation: from precision, recall and F-mesure to ROC, informedness, markedness and correlation. J Mach Learn Technol 2:37–63

Qureshi WS, Payne A, Walsh KB, et al (2016) Machine vision for counting fruit on mango tree canopies. Precis Agric. https://doi.org/10.1007/s11119-016-9458-5

Rahnemoonfar M, Sheppard C (2017) Deep Count: Fruit Counting Based on Deep Simulated Learning. Sensors 17:905. https://doi.org/10.3390/s17040905

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp 779–788

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp 91–99

Sa I, Ge Z, Dayoub F, et al (2016) DeepFruits: A Fruit Detection System Using Deep Neural Networks. Sensors 16:1222. https://doi.org/10.3390/s16081222

Sarron J, Sané CAB, Borianne P, et al (2020) Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions? Acta Hortic 8

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089

Sermanet P, Eigen D, Zhang X, et al (2014) OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. ArXiv13126229 Cs

Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs

Stein M, Bargoti S, Underwood J (2016) Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors 16:1915. https://doi.org/10.3390/s16111915

Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM (2013) Selective Search for Object Recognition. Int J Comput Vis 104:154–171. https://doi.org/10.1007/s11263-013-0620-5

Zameer A, Arshad J, Khan A, Raja MAZ (2017) Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. Energy Convers Manag 134:361–372. https://doi.org/10.1016/j.enconman.2016.12.032

Zeiler MD, Fergus R (2014) Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp 818–833

III. Production gaps at the tree scale: concept and application to mango in West Africa

L'utilisation du CNN pour estimer le nombre de fruits par arbre a permis d'obtenir la production (en kg) de chacun des 300 arbres de l'échantillon et sur les deux années d'étude. Cette base de donnée permet d'étudier les facteurs impactant la production à l'échelle de l'arbre (Fig. 32, p. 66). La Partie III de ce chapitre est un article scientifique décrivant le concept de '*production gap*' qui est l'adaptation du concept de *yield gap* à l'échelle de l'arbre. La définition de différents niveaux de production (production atteignable et production réelle) permet de prendre en compte plusieurs facteurs impactant le rendement. L'analyse des frontières permet d'estimer les production gaps sur notre échantillon de 300 arbres et de comparer les performances des trois variétés dans différents contextes.

Cet article a été soumis et en révision en juillet 2019 à une revue à comité de lecture : *Agronomy for Sustainable Development*.

- Sarron J.; Faye É.; Nordey T.; Diatta J.; Malézieux É. **Production gaps at the tree scale: concept and application to mango in West Africa.** *Agron. Sustain. Dev.* [*accepted*]

Se trouve également en annexe un poster de synthèse sur l'effet des variables endogènes de l'arbre et sur la notion d'indice de charge (Annexe II). Ce poster a été présenté lors du *XV European Society for Agronomy Congress* (ESA) à Genève en août 2018.

- Sarron J.; Sané C.A.B.; Diatta J.; Faye É.; Malézieux É. **Estimating and modelling productivity of mango trees from endogenous factors. A methodology set up and validated for African Orchards.** XV European Society for Agronomy Congress, 2018, Genève, Suisse.

Production gaps at the tree scale: concept and application to mango in West Africa

Julien Sarron[1,2,3,*], Emile Faye[1,2,3], Thibault Nordey[1,2,4], Damien Beillouin[1,2], Jeanne Diatta[3], Frédéric Normand[1,2], Eric Malézieux[1,2]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France

[2]HortSys, University Montpellier, CIRAD, F-34090 Montpellier, France

[3]Centre pour le Développement de l'Horticulture, ISRA, Dakar 14000, Senegal

[4]The World Vegetable Center, Eastern, and Southern Africa, P.O. Box 10, Duluti, Arusha, Tanzania

*Corresponding author: julien.sarron@cirad.fr

Abstract

In developing countries, fruit tree production and its driving factors are largely unknown compared to annual crops. This study adapts the concept of yield gap to the tree scale and defines the production gap. The production gap of the tree is calculated as the difference between attainable production, defined by 'production-limiting' factors (i.e., tree cultivar, structure, and age), and actual production. The 'production-reducing' factors (i.e., annual climate, management practices, and orchard structure) are identified as the drivers of the production gap. We applied this approach to estimate and analyze the production gap of 300 mango trees consisting of three cultivars sampled in the three main cropping systems (extensive, diversified, and intensive) found in Senegal, West Africa. For two years, individual tree structure parameters and actual production were precisely measured using airborne and ground image analysis. Through a stochastic frontier analysis, we estimated the attainable production per tree and the production gap and studied the relative importance of production-limiting and production-reducing factors. We found that the mean production gaps were 0.52, 0.41, and 0.67-fold the attainable production for 'Kent', 'Keitt', and 'Boucodiékhal' cultivars, respectively. Mango trees in diversified orchards showed the highest average actual production (217 fruits). In addition, the production gaps of trees in diversified orchards reached 0.43, a similar value compared to intensive orchards but much lower than extensive orchards (0.72). Irrigation and fertilization are effective in reducing tree production gaps. Based on the production gap concept, we clearly showed that large margin exists to increase mango production. This study made it possible to estimate mango tree production and, to our knowledge, represents the first application of yield gap analysis at the tree scale, helping to bridge the knowledge gap concerning fruit trees in West Africa.

Keywords: fruit production, yield gap, stochastic frontier analysis, Senegal

III.1. Introduction

Agriculture faces the challenge of increasing crop yields while reducing their environmental impacts under a changing climate and a limited amount of water resources (van Ittersum et al. 2013). Fruit trees are of key importance for food security (nutrient intake and improved health) and for the income resilience of growers, especially in developing countries under subtropical and tropical climates (Malézieux et al. 2009). In these regions, fruit trees are mostly grown in multispecies cropping systems with their inherent variabilities in tree cultivar and structure (Malézieux et al. 2009; Grechi et al. 2013).

Mango (*Mangifera indica L.*) represents the fifth most widely produced fruit in the world (51 Mt in 2017, FAOSTAT 2017) and ranks among the most cultivated fruit tree species in the subtropics and tropics. In West Africa, mango production has continuously increased over the last decades to reach approximately 5% of the global production (FAOSTAT 2017). In this region, the mango tree is of key importance for food security and is mostly grown in smallholder orchards (< 5 ha) under different cropping systems: from extensive orchards with no management practices, to highly diversified orchards (Vannière et al. 2007; Grechi et al. 2013). Some growers produce mango for the export market in large monospecific orchards with high levels of inputs (i.e., fertilizers, phytosanitary products, and water).

Each tree has its own potential and actual production (expressed in number of fruits or in kg), and orchard yield, expressed in unit of product per unit of land, is the sum of the production of all of the trees. The tree cultivar and structure (i.e., tree crown dimensions) appear as the primary determinants that define the production potential and the upper limit of the number of fruits, through modulation of photosynthesis efficiency and light interception (Schaffer et al. 2009). Many studies have highlighted a positive and cultivar-dependent correlation between the structure and the productivity of the tree (see Menzel and Le Lagadec (2017) for a review). Nonetheless, as for most perennial crops, mango tree endogenous factors (i.e., tree genetics, age, and structure) also interact with exogenous (i.e., external factors such as the environment and management practices) factors (Schaffer et al. 2009; Boudon et al. 2017). These complex interactions result in largely variable flowering intensity or fruit development between trees, and lead to reproductive asynchronisms (i.e., phenological stages spread out over time between trees and even within trees) and irregular bearing (Léchaudel and Joas 2007; Ramírez and Davenport 2010; Dambreville et al. 2013). As a consequence, the actual tree production observed in a given year may be lower than the tree potential, consequently causing production gaps and disparities between trees at the orchard scale. This is especially the case in smallholder orchards where management practices are not optimal (Vannière et al. 2007). Such observations motivates the study of the drivers of mango production at the tree scale.

Despite the importance of mango production in West Africa, studies analyzing the drivers of mango tree production remain scarce. Some previous studies focused on the relationship between tree features (structure, age, and cultivar) and the water use (e.g., Oguntunde et al. 2011). However, they did not focus on tree production and were conducted in experimental conditions. Only a few authors studied mango tree production in field conditions, e.g., focusing on pests like fruit flies (Vayssières et al. 2008). Thus, field data on fruit tree production and growth conditions (environment, management practices, etc.) remain sparse and inaccurate in West Africa (Carletto et al. 2015). At the tree scale, the absence of tools to accurately estimate tree production in the field hampers the development of a statistical methodology for the investigation of impacting factors. Consequently, studies on tree production variability are still inexistent to our knowledge, especially in smallholder and complex orchards in which the within-field variations in tree structure and management appear to be exacerbated (Grechi et al. 2013).

Yield gap methodologies have been widely used in the literature to decipher the cause of crop yield limitations due to the environment, genetic traits, or management practices factors (van Ittersum et al. 2013). Yield gap is commonly defined as the difference between the potential yield (Y_p) – the yield of a specific crop genotype in a well-defined biophysical environment with no limitation in nutrients or water resources, and no pest and disease pressures (i.e., the 'growth-defining' factors) – and the actual yield (Y_a) (Neumann et al. 2010). Studies focusing on yield gaps at field, farm, regional, and global scales mainly concern annual crops like cereals (e.g., Neumann et al. 2010; Silva et al. 2017). While yield gap analysis provides a robust methodology for identifying the most important factors limiting crop yields (van Ittersum et al. 2013), few references exist on fruit crops such as banana (Chopin and Blazy 2013) or mango (de Bie 2004; Mensah and Brümmer 2016). To our knowledge, no yield gap analyses have been performed at the tree scale. The application of this methodology to fruit crops is partly hindered by the lack of data and the absence of crop models for the estimation of potential tree production (Boudon et al. 2017).

The aim of this paper was to bridge the knowledge gap concerning the factors that lead to variability in mango tree production in West Africa. We adapted the yield gap concept to define the production gap (P_g) as the difference between the attainable production (P_{att}) and the actual production (P_a). Tree cultivar, structure, and age were considered as production-limiting factors that determine P_{att}. Using the analysis of ground and unmanned aerial vehicle (UAV) images, Pa and tree structure parameters were precisely measured on a large set of mango trees representative of mango trees found in the main West African cropping systems. Applying stochastic frontier analysis (SFA) allowed us to estimate P_{att} and P_g and to reveal the effects of production-reducing factors including annual climate, mango management practices, and orchard structure. Finally, the tree performances (P_a, P_{att}, and P_g) under the three main cropping systems found in the study area were compared. This study highlighted the effects of tree structure and of some management practices (irrigation and fertilization) on P_{att} and P_g, respectively.

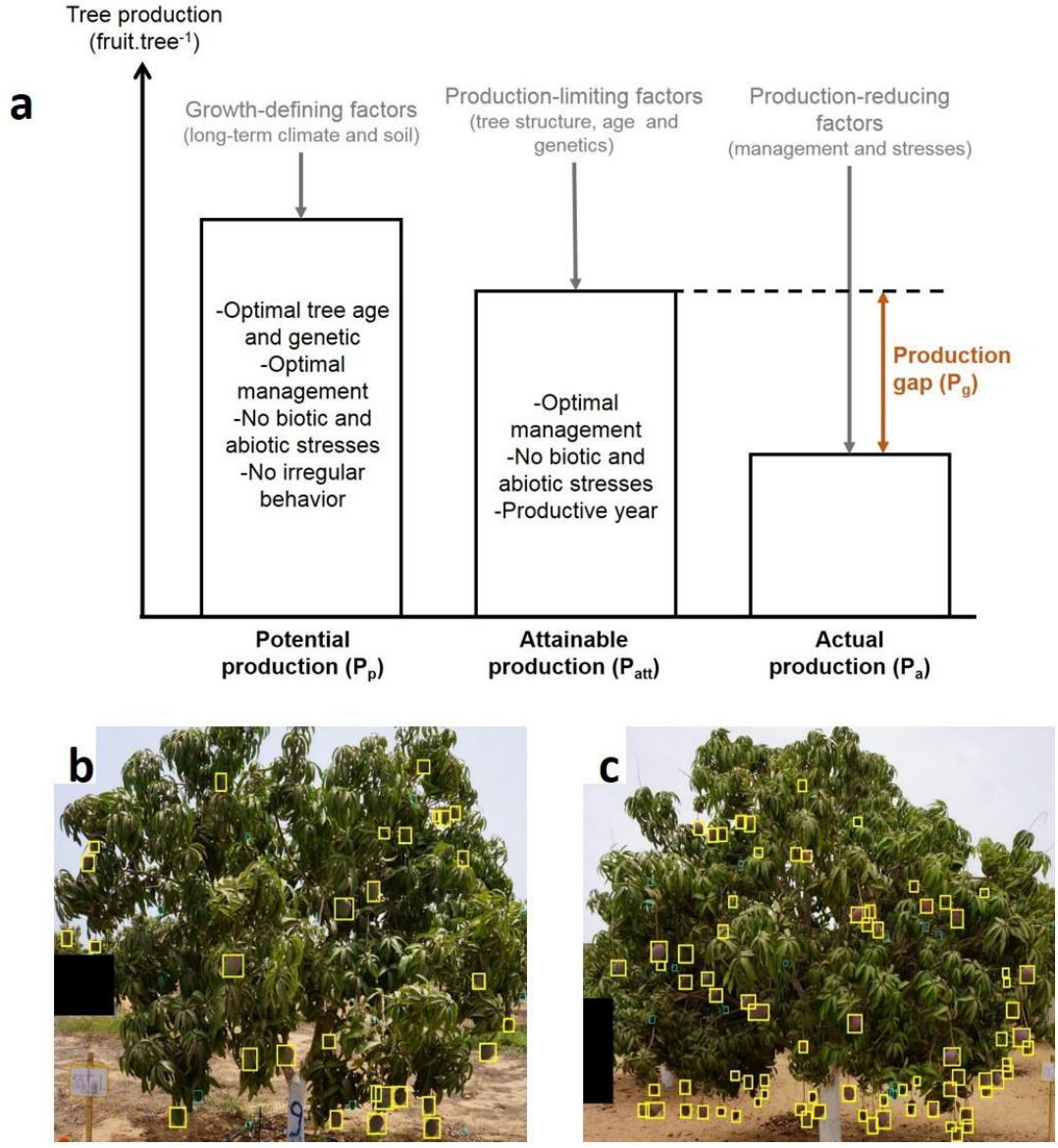


Fig. 57. (a) Conceptual framework of the production gap concept. Actual production (P_a) estimated from machine vision detection for two Kent trees with similar crown volume with (b) low fruit load; and (c) high fruit load. Production gap (P_g) in fruit number could possibly be explained by production-reducing factors (management and stresses) or a non-productive year.

III.2. Materials and methods

a. The concept of tree production gap

As a result of the objective to quantify the drivers of mango production at the tree scale, as well as the perennial status of trees, the common definition of yield gap had to be adapted. We defined 'potential production' (P_p) as the tree production obtained by a given cultivar grown with optimized pruning practices, no limitation on water and nutrients, and control of biotic and abiotic stress (Fig. 57a). Under such conditions, tree production was determined by long-term climate (temperature, precipitation, and solar radiation), soil features (N and organic content), and by genetic traits. These factors, referred to as 'growth-defining' factors, govern the tree structure (i.e., tree crown dimension) and photosynthesis efficiency (Schaffer et al. 2009). Like potential yield (Y_p), P_p was considered as theoretical (van Ittersum et al. 2013) since it assumes optimal conditions of tree structure and age and no irregular behavior (asynchronism and alternate bearing).

In fact, such optimal conditions are unobserved and tree production is actually limited by the interaction between tree age and structure (Dambreville et al. 2013; Boudon et al. 2017), referred to as 'production-limiting' factors in this study. We then defined the 'attainable production' (P_{att}) as the maximum production that a tree could reach given its current structure and age (Fig. 57a). For a given year and modality of growth-defining and production-limiting factors, P_{att} was reached when inputs were correctly used to control abiotic and biotic stresses and to maximize the number of fruits per unit of crown volume (i.e., the fruit load) (Léchaudel and Joas 2007; Ramírez and Davenport 2010).

The actual production (P_a) was defined by the 'production-reducing' factors, which were annual climatic (water and temperature) and biotic stresses during the growing cycle. In addition, some orchard structure parameters were also considered as production-reducing factors because they could possibly have an impact on P_a during the growing cycle. For example, tree planting density could conceivably influence individual tree production by impacting between-tree competition for resources (Menzel and Le Lagadec 2017). In this study, we focused on the P_{att} and the P_a, since P_p was difficult to estimate for fruit trees, and we defined the production gap (P_g) as the difference between P_{att} and P_a.

b. Statistical stochastic frontier modeling

In order to determine P_g, and because no functional model exists on mango to determine P_p and P_{att}, we used the statistical stochastic frontier analysis (SFA). This methodology arose from economic studies (Battese and Coelli 1995) and has been used to estimate grain yield gaps (Neumann et al. 2010; Silva et al. 2017). SFA is a parametric method that fits a function (i.e., stochastic frontier) representing the attainable output for a given set of factors referred to as 'input factors' (Battese and Coelli 1995). In our case, the function measured P_{att} as the maximum P_a obtained for specified levels of inputs, which were the production-limiting factors (Section III.2.a). Stochastic frontier modeling considers two independent random errors: statistical noise (e.g., measurement errors, data uncertainties) and inefficiency. Different functions exist to model the stochastic frontier. In this study, the Cobb-Douglas function was chosen because of its simplicity and common use in yield gap studies (Neumann et al. 2010; Mensah and Brümmer 2016; Silva et al. 2017). Thus, the frontier function is specified as follows:

$$ln(Pa_i) = \beta_0 + \sum_{k=1}^{k=N} \beta_k \, ln(x_{ki}) + v_i - u_i = ln(Patt_i) - u_i \qquad (8)$$

where Pa_i and $Patt_i$ denote the actual production (P_a) and the attainable production (P_{att}) of the ith tree respectively; x_{ki} (k = 1,2,...,N) is the vector of the N input factors (i.e., production-limiting factors) for the

i^{th} tree; β_0 and β_k (k = 1,2,...,N) are the parameters to be estimated; and v_i and u_i are the statistical noise and the inefficiency, respectively, both normally distributed. The statistical noise v_i is an inherent random stochastic error that takes the deviation from the production frontier caused by measurement errors into account (Neumann et al. 2010). The inefficiency u_i is a non-negative random variable representing the technical inefficiency effects and is assumed to be independent of v_i. The function of u_i is specified as:

$$u_i = \delta_0 + \sum_{j=1}^{j=Z} \delta_j z_{ji} + \varepsilon_i \qquad (9)$$

where z_{ji} (j = 1,2,...,Z) denotes the vector of the Z inefficiency factors (i.e., production-reducing factors) for the ith tree; δ_0 and δ_j (j = 1,2,...,Z) are the parameters to be estimated; and ε_i is the residuals. The inefficiency u_i quantifies the deviation of the P_a from the frontier that is not caused by statistical noise.

A variable calculated from the model is the 'production efficiency' (P_{eff}), which is related to u_i and can be defined as the ratio of the actual production (P_a) to its corresponding attainable production (P_{att}):

$$Peff_i = \exp(-u_i) = \frac{Pa_i}{Patt_i} \qquad (10)$$

Hence:

$$Pg_i = 1 - Peff_i = \frac{Patt_i - Pa_i}{Patt_i} \qquad (11)$$

where $Peff_i$ and Pg_i denote the production efficiency and the relative production gap of the ith tree, respectively.

Thus, SFA is comparable to two combined regressions: the first one (Eq. 8) allows to compute the frontier function that integrates the production-limiting factors; the second (Eq. 9) allows to compute the effects of the production-reducing factors on P_g. The relative production gap (expressed as a ratio of P_{att}) was preferentially used rather than the absolute production gap (expressed in fruit) for more consistency in the comparison between cultivars and cropping systems. The following sections describe the variables, the origin of data (Table 3), and the SFA model development for analysis of the factors driving mango P_g in the Niayes region, a production basin in Senegal.

c. Study site and data collection

Study site and cropping systems

The study was carried out in the Niayes region, a major mango production basin in western Senegal that covers around 500 km² and is located between 14°4' to 15°0' N and 16°6' to 17°2' W. The region is part of the Sudano-Sahelian zone and is characterized by unimodal rainfall from July to September (mean rainfall of 430 mm per year) and a relatively cool and humid climate generated by the proximity to the Atlantic Ocean (with daily mean temperatures ranging from 16°C to 35°C). The landscape is characterized by a succession of dune, with sandy and unleached soils, and depressions, with clay and peaty soils. The groundwater level is not deep in these depressions and can reach the surface during the rainy season (Fare et al. 2017). This topography, combined with favorable climatic conditions, make this region one of the major fruit, including mango, and vegetable production areas in Senegal.

Table 3. Description of the variables used in the stochastic frontier analysis of mango tree production.

Variable name	Unit	Definition and source of data
P_a	fruits	Actual production estimated from ground image analysis
Tree structure and age		
tree height	m	Individual tree height estimated by UAV (Sarron et al. 2018)
crown area	m^2	Individual tree projected crown area estimated by UAV (Sarron et al. 2018)
crown volume	m^3	Individual tree crown volume estimated by UAV (Sarron et al. 2018)
DBH	cm	Diameter at breast height
Annual climate		
T_flower	hour	Cumulative hours below degree stimulating flowering (18°C day/10°C night)
T_stress	°C	Cumulative hours above degree stressing development (40°C)
T_mean	°C	Mean temperature during cycle (fruiting and flowering)
previous rain	mm	Annual rainfall of the study area in 2016 and 2017
Mango management		
phyto	categorical	Frequency of phytosanitary product application obtained from the survey (*occas* = no application or not applied every year; *annual* = 1 or more applications per year)
irrig	categorical	Mango irrigation (*no*; *yes*)
fert	categorical	Frequency of fertilizer application obtained from the survey (*occas* = no application or not applied every year; *annual* = 1 or more applications per year)
prun	categorical	Frequency of mango tree pruning obtained from the survey (*occas* = no application or not applied every year; *annual* = 1 or more applications per year)
Orchard structure		
annual_perc	%	Ratio of annual crop area compared to productive area measured by UAV
mango_density	$trees.ha^{-1}$	Mango tree density measured by UAV
citrus_density	$trees.ha^{-1}$	Citrus tree density measured by UAV

In this area, 30 mango-based orchards were selected to be spatially distributed in the study area and to represent the diversity of cropping systems and environmental conditions. Each orchard was classified into one of the three main types of mango cropping systems identified by Grechi et al. (2013) in the Niayes region. This typology was mainly based on the crop composition (tree species and/or annual crops) and mango management practices in the orchard. The three mango cropping systems were: (i) extensive smallholder orchards mainly composed of mango trees planted on bare ground with none or a low level of management practices (pruning, irrigation, chemicals, etc.) (n = 10 orchards in our sample); (ii) diversified smallholder orchards composed of mango trees along with other fruit tree species (citrus, papaya, etc.) occasionally with market gardening underneath the trees (n = 10); and (iii) monospecific commercial-based orchards composed of one or two Floridian export cultivars (‘Kent’ and ‘Keitt’; Rey et al. 2007) with a high level of inputs (n = 10). The area of the sampled orchards varied from 0.90 to 8.70 ha, with an average of 2.2 ha. Information about the performance of these cropping systems was inaccurate and based on expert knowledge. Monospecific orchards are known to reach higher mango yields (from 20 to 40 $t.ha^{-1}$) than smallholder orchards (i.e., extensive and diversified orchards), which produce between 1 to 10 $t.ha^{-1}$ of mangoes and are predominant in the Niayes region.

Ten trees were sampled in each selected orchard among the three main cultivars found in the Niayes region: 'Kent', 'Keitt', and 'Boucodiékhal', a local cultivar. For each orchard, the ten productive trees (at least 5 years old) were selected to ensure a spatial distribution and to depict the entire variability of tree structure, age, and fruit load found within the orchard. Thus, the total number of selected trees was 300 for the study.

Estimation of actual tree production (P_a)

A deep-learning image analysis algorithm (Borianne et al. 2019) was used to estimate the actual tree production P_a (expressed in number of fruits) of the sampled trees for both 2017 and 2018. Briefly, each tree was imaged on two opposite sides using a hand-held Sony Nex-7 RGB camera (Sony Corporation, New York, NY, USA) at a fixed focal length of 18 mm and at a resolution of 6000x4000 pixels. Images were taken a few days (maximum two weeks) before harvest. A region-based convolutional neural network (CNN) algorithm, called Faster R-CNN (Ren et al. 2015), wasin then trained to accurately detect mango fruits in the crown, as illustrated in Fig. 57b-c. CNNs turned out to be very effective for fruit detection and Faster R-CNN has already been used for mango (Koirala et al. 2019). In 2017, the algorithm was trained to detect the mango fruits of 'Kent', 'Keitt', and 'Boucodiékhal' cultivars using 3,000 labeled fruits annotated by three experts on 150 images. The algorithm reached a detection accuracy of 92% when evaluated against 7,000 additional fruit annotations. Faster R-CNN functioning, parameterization, and evaluation are described in Borianne et al. (2019).

In this study, we ran this neural network algorithm on 1,200 images of mango trees (2 sides x 300 trees x 2 years) to achieve the automatic count of visible fruits per tree, computed as the number of detections from both images of the tree (Fig. 57b-c). Then, to estimate actual tree production (P_a) from the results of algorithm detection, the actual number of fruits was visually counted in the field (field count) by three experts on a total of 41 'Kent', 42 'Keitt', and 33 'Boucodiékhal' trees over both years. For each cultivar, the relationship between the automatic and the field count was fitted using a linear or a polynomial regression with the intercept forced to 0. These models were evaluated by computing the coefficient of determination (R^2), and the normalized root-mean-squared error (NRMSE) calculated as the RMSE divided by the range (max - min) of field count values. These models were used to estimate the P_a of the 300 sampled trees for both years and expressed in number of fruits. Finally, since the trees displayed different structures, the fruit load was also computed as P_a divided by the crown volume (fruits.m^{-3}) measured by UAV.

Characterization of the production-limiting factors: tree structure and age

For a given cultivar, the attainable production (P_{att}) is limited by the production-limiting factors, mainly tree structure and age (Section III.2.a). In this study, tree structure parameters (tree height, crown area, and crown volume) were measured using UAV photogrammetry and object-based image analysis following the method described in Sarron et al. (2018). Briefly, in 2017, each of the 30 orchards was overflown using a UAV (DJI Mavic Pro, DJI Inc., China) in order to compute a high-resolution RGB orthomosaic (1.30 cm.pixel^{-1}) and a Canopy Height Model (CHM) using photogrammetry procedures. The two layers (CHM, orthomosaic) were analyzed through an object-based workflow to obtain a land cover map on which individual tree crowns were delineated and tree species and mango cultivars identified. A land cover map combined with the CHM made it possible to compute three tree structure parameters: tree height (in m), the projected crown area (in m^2), and the crown volume (in m^3) (Table 3). The tree structure parameters could possibly be strongly correlated with each other and with the tree production. Thus, correlations between tree structure parameters and P_a were explored by computing Spearman's rho coefficient. For each cultivar, we kept the structure parameter that had the highest correlation coefficient with P_a.

Since, tree age was not directly accessible in most cases, we added the diameter at breast height (DBH) as a production-limiting factor and proxy of age. Indeed, the relationships between tree age, vigor and structure parameters in mango were demonstrated. Consequently, tree age can be depicted by combining DBH and

tree structure (Menzel and Le Lagadec 2017). DBH was derived from trunk perimeter directly assessed on each tree with a measuring tape. Thus, the selected structure parameter (i.e., the most correlated with P_a) and the DBH both constituted the production-limiting factors to be used in the SFA model (Eq. 8). These factors were only estimated the first year of the study (2017) since we assumed that tree structure and DBH did not greatly vary between the two consecutive years.

Characterization of the production-reducing factors: annual climate, management practices, and orchard structure

In order to characterize the annual climate of each orchard, we used the temperature and rainfall variables as production-reducing factors in the SFA model. To describe the annual climate for 2017 and 2018 that might influence the annual tree production (P_a), temperature loggers (HOBO U23 Pro v2, Onset Computer Corporation, USA) recorded air temperature every 10 minutes at a height of 2 m from January 2017 to December 2018 in each orchard. Temperature was summarized at hour intervals (mean, minimum and maximum). Flowering and fruiting periods are cultivar-specific: flowering occurs from January to April, from mid-January to mid-April, and from February to mid-April for 'Boucodiékhal' (early cultivar), 'Kent', and 'Keitt' (late cultivar), respectively; fruiting occurs from mid-February to mid-June, from mid-February to July, and from mid-March to August for 'Boucodiékhal', 'Kent', and 'Keitt', respectively. Thus, temperature variables were computed during the specific flowering and fruiting periods of the cultivar (Table 1). Three variables were used: '*T_flower*' as the cumulative degree hours stimulating flowering (i.e., number of hours with T < 18°C day and < 10°C night; Ramírez and Davenport 2010) '*T_stress*' as the cumulative degree hours of temperature stress (T > 40°C; Schaffer et al. 2009); and '*T_mean*' as the mean temperature during the reproductive cycle (from the beginning of flowering to the end of fruiting).

In Senegal, mango flowering and fruiting stages take place during the dry season, from January to mid-June. Therefore, the rainfall of the previous rainy season (i.e., of 2016 and 2017), occurring the year before harvest, impacts the water availability during the current fruit production season (Fare et al. 2017). The rainfall of the previous rainy season was measured by the climatic station (HOBO U30, Onset Computer Corporation, USA) located at the Sangalkam experimental station of the Senegalese Research Institute on Agriculture (ISRA) within the study area (14°47'22.0"N, 17°13'36.0"W).

Management practices on mango trees were surveyed during 2017 so as to be used as production-reducing factors related to the cropping system. Information (date and frequency, quantity of input, tool used, etc.) on tree pruning, chemical applications (phytosanitary product, fertilizer), and irrigation were collected by interviewers by means of a questionnaire. Pruning practice was included because we distinguished effect of pruning on the tree structure (included as production-limiting factors) from the direct effect of pruning on tree production and alternate behavior (Léchaudel and Joas 2007; Dambreville et al. 2013). Flower and fruit thinning practices, which could be applied to ensure bigger fruits, were not observed among the sampled orchards. Rootstocks were unknown in most of the cases and this variable was not considered in the study. Unfortunately, data were obtained with not much details in most cases since most growers do not record their activities. Thus, information on management was summarized into categorical variables describing levels of frequency of each practice (Table 3). In addition, the ratio between the area of annual crops and the total productive area was computed from the land cover map because intercropping practices have the potential to increase water availability for the trees (Malézieux et al. 2009; Rathore et al. 2013).

Variables describing orchard structure were obtained from the UAV land cover map. These variables were the mango planting density (i.e., number of mango trees per ha) and the citrus planting density. They characterized the competition for light interception that could reduce production at the tree scale (Menzel and Le Lagadec 2017).

d. SFA modeling and tree production gap analysis

For each cultivar, an SFA model was developed using the Cobb-Douglas function (Eq. 8 and Eq. 9) and the P_a values for both years. Simultaneous estimations of the maximum likelihood of the model parameters and error terms (β, δ, v, and ε in Eqs. 8 and 9) were processed using the R package 'frontier' (Coelli and Henningsen 2017). The SFA model makes it possible to directly identify significant effects of production-limiting and production-reducing factors on P_{att} and production gap P_g respectively.

In order to analyze the effect of the cultivar and the cropping system on P_a, fruit load, P_{att}, and P_g, ANOVAs were performed with no distinction of the year. After each significant ANOVA, Tukey's test was performed for comparison between cultivar or orchard type. For each cultivar, correlations between tree structure parameters on P_a or fruit load were revealed by Spearman's test. Linear regression was fitted between the fruit load and tree crown volume to show the possible link between both variables and the estimated percentage of fruit load variance explained by the tree structure. Yearly variation of P_g was revealed using the non-parametric Wilcoxon signed-rank test because data were not normally distributed. Potential alternate behavior was established by computing the regression slope of P_g between both years: a slope close to -1 would indicate an alternate bearing behavior of the trees. Finally, effects of some management practices (irrigation and fertilization) on P_g were explored by means of ANOVA and Tukey's tests.

III.3. Results and discussion

a. Estimation of the actual tree production (Pa) and fruit load

Combining image analysis with a regression model (Fig. 57) predicted the P_a with an R^2 of 0.94, 0.92, and 0.95 for the 'Kent', 'Keitt', and 'Boucodiékhal' cultivars, respectively. The NRMSE varied from 6.9 to 7.2%, meaning that the field count was estimated from the automatic count with approximately 7% of error. The best relationship between automatic counts and field counts was obtained with a linear regression for 'Kent', and with a 2nd order polynomial regression for 'Keitt' and 'Boucodiékhal'. The model reliabilities were similar to Koirala et al. (2019) who reported an R^2 between 0.94 and 0.98 when estimating the field count of trees in different (five) commercial orchards, and also by using two images per tree analyzed with the Faster R-CNN algorithm. These authors only reported RMSE (expressed in fruit per tree), but values could be averaged by the mean number of fruits per tree (thus, NRMSE ranged from 11 to 19%). This confirms the accuracy of image analysis algorithms combined with regression models to rapidly estimate individual tree production. Such a tool improves yield estimation compared to manual estimation that is inaccurate, fastidious and that increases the cost of data acquisition (Koirala et al. 2019).

The P_a of the 300 sampled trees for two years showed a wide distribution, from 0 (no fruit in the tree) to 1,238 fruits, with an average (av.) of 157 fruits and a standard deviation (sd.) of 159 fruits (Fig. 58a). For the analysis of the fruit load, 18 trees with a crown volume lower than 1 m^3 were removed since such a low volume gave abnormal fruit load values. Thus, fruit load varied from 0 to 127 fruits.m^{-3}. 'Keitt' trees (av. = 262 fruits) produced significantly more fruits than both 'Boucodiékhal' (av. = 154 fruits) and 'Kent' (av. = 127 fruits) trees. Even though 'Boucodiékhal' had a high P_a, it had the lowest fruit load (av. = 1.5 fruits.m^{-3}) compared to 'Keitt' (av. = 7 fruits.m^{-3}) and 'Kent' (av. = 5 fruits.m^{-3}). However, 95% of the trees had a P_a lower than 450 fruits and only six trees had a P_a value greater than 800 fruits. Although the highest estimates of P_a (1,238 fruits) might be doubtful, they are comparable to the range of the field count measured for the calibration of models, which varied from 0 to 1,364 fruits. Fourteen trees belonging to extensive and diversified orchards had no fruit (five in 2017 and nine in 2018). This was not surprising since non-productive trees are common in West Africa, generally due to environmental stresses, inadequate management practices, and alternate behavior (Vannière et al. 2007). These trees were removed for SFA modeling.

These calculations of P_a and fruit load represent the first accurate estimates of production from a broad panel of mango trees from different cropping systems in West Africa. Data on fruit tree production in this region remain scarce and, to our knowledge, Vayssières et al. (2008) proposed the only study reporting individual mango tree production in a field context. In central and northern Benin, the authors manually counted the fruits of 350 mango trees of seven cultivars (including 'Keitt' and 'Kent') but from only five orchards with an intermediate level of inputs. In Benin, the output of the export cultivars was 120 and 157 fruits for 'Kent' and 'Keitt', respectively (Vayssières et al. 2008), comparable to our estimates. Nonetheless, to our knowledge, no study has yet focused on the main mango-based cropping systems existing in West Africa, especially the extensive system that predominates in the region (Vannière et al. 2007).

b. Deciphering mango production gaps in the Niayes area

The three SFA models developed to compute P_g and P_{att} from the P_a estimates are presented in Table 4. In the first part of the model (i.e., production-limiting factors) a positive estimate indicated a positive effect of the parameter on P_{att}, while in the second part (i.e., production-reducing factors) a positive estimate indicated an increase of P_g, negatively impacting the production (Table 4).

Impacts of production-limiting factors on actual and attainable production and fruit load

For the three cultivars, strong positive correlations were found between the production limiting factors, the tree structure parameters (tree height, crown area, and crown volume) and DBH (Spearman's rho > 0.61; p-value = 0). Correlations between P_a and tree structure parameters were significant but weaker with positive coefficients between 0.31 and 0.62, depending on the cultivar and the structure variable. The selected tree structure variable for the frontier model was the crown area for 'Kent' and 'Keitt', and the tree height for 'Boucodiékhal' since they had the highest correlation coefficient with P_a. On average, 'Boucodiékhal' trees had a greater crown volume than the two other cultivars (164 m^3 vs. 109 m^3 for 'Keitt' and 46 m^3 for 'Kent'), but a lower P_a (Fig. 58a), thus explaining the low fruit load observed for the local cultivar (Section III.3.a).

Linear regressions pointed out a significant negative effect of the tree crown volume on the fruit load for all cultivars. Nonetheless, tree crown volume only explained from 3.3% to 12.6% of the fruit load variance, suggesting that exogenous production-reducing factors predominantly explained tree production efficiency and gap.

The tree structure variable positively impacted P_{att} for all cultivars (all p-value < 0.008; Table 2). In addition, an increase in DBH induced an increase in P_{att} for 'Kent' (p-value < 0.05). These results corroborated a positive relationship of tree structure and age with P_a and P_{att}, and supported previous conclusions reported by Menzel and Le Lagadec (2017). These studies established allometric relationships between tree growth, age and production, and the effect of the cultivar on these relationships. Although it was not clearly understood, rootstock and scion affected tree growth and productivity. In the present case, the rootstock of each tree was unknown. However, most growers use a local dwarf cultivar ('Sewe') as rootstock for export cultivars 'Kent' and 'Keitt' to control tree size, while 'Boucodiékhal' is not usually grafted (Rey et al. 2007).

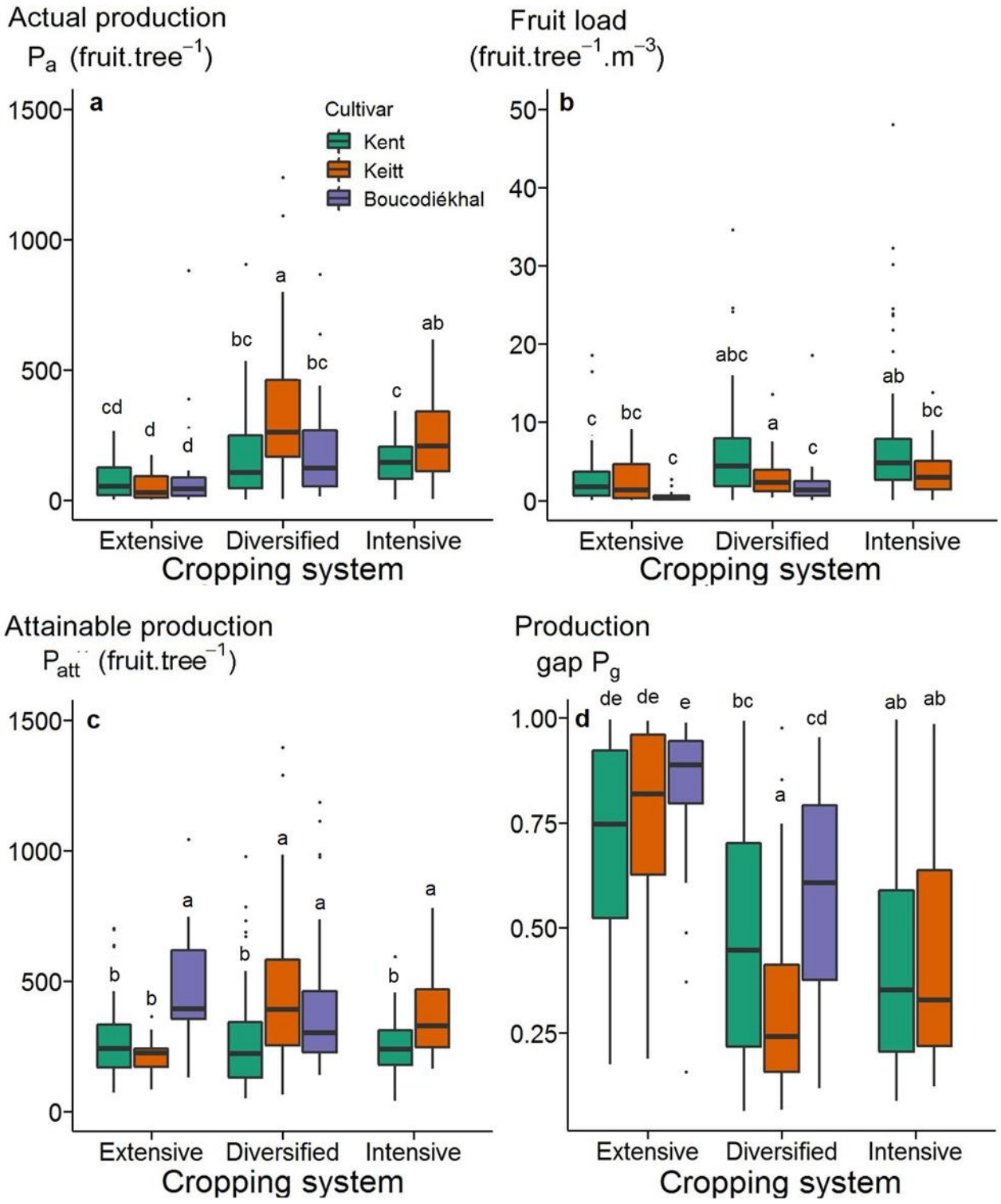


Fig. 58. Breakdown of attainable production for the three cultivars ('Kent', 'Keitt', and 'Boucodiékhal') under the different cropping systems: (**a**) distribution of actual production P_a; (**b**) distribution of fruit load (=P_a/crown volume); (**c**) distribution of attainable production P_{att} calculated by frontier modeling; (**d**) distribution of the production gap P_g (= (P_{att} – P_a)/P_{att}). Letters indicate groups according to Tukey's test.

Table 4. Parameter estimates, standard error (SE) and p-value (significant at the 0.05 level) of the stochastic frontier models estimated for the three mango cultivars ('Kent', 'Keitt', and 'Boucodiékhal').***: 0.001; **: 0.01; *: 0.05.

Variable	Kent		Keitt		Boucodiékhal	
	Estimates (SE)	p-value	Estimates (SE)	p-value	Estimates (SE)	p-value
Production-limiting factors(log-transformed)						
Intercept	5.74 (0.04)	<0.001***	6.12 (0.13)	<0.001***	6.14 (0.22)	<0.001***
tree height					1.14 (0.43)	0.008*
crown area	0.44 (0.08)	<0.001***	0.53 (0.14)	<0.001***		
DBH	0.50 (0.13)	<0.001***	0.18 (0.35)	0.616	0.63 (0.44)	0.159
Production-reducing factors						
Intercept	-8.08 (28.93)	0.780	94.00 (53.67)	0.080	-43.47 (28.86)	0.132
irrig [yes]	-3.79 (1.73)	0.028*	10.38 (6.00)	0.084	3.03 (6.76)	0.655
phyto [annually]	2.25 (1.20)	0.060	-2.37 (1.67)	0.156	2.96 (0.92)	0.001**
fert [annually]	-1.84 (1.79)	<0.001***	7.07 (4.94)	0.153	2.57 (1.28)	0.046*
prun [annually]	0.01 (1.88)	<0.001***	-20.52 (11.40)	0.072	-2.50 (1.08)	0.021*
annual_perc	-1.98 (0.88)	0.024*	3.81 (1.52)	0.012*	-0.75 (0.44)	0.089
mango_density	-8.05 (2.87)	0.005**	-2.25 (1.98)	0.255	0.49 (0.82)	0.547
citrus_density	-0.44 (0.60)	0.470	-13.02 (5.71)	0.023*	-3.18 (1.11)	0.004**
T_flower	4.19 (3.60)	0.245	-12.24 (7.44)	0.100	9.63 (5.27)	0.068
T_stress	0.12 (0.81)	0.879	1.63 (1.15)	0.157	-0.35 (0.68)	0.611
T_mean	5.95 (24.08)	0.805	-52.78 (35.33)	0.135	30.29 (22.12)	0.171
previous rain	3.20 (5.97)	0.592	-24.82 (13.77)	0.071	5.85 (4.08)	0.151
sigma squared	5.96	0.005**	2.20	0.012*	1.12	<0.001***
gamma	0.99	<0.001***	0.90	<0.001***	0.93	<0.001***
log-likelihood	-423.44		-114.02		-85.80	
mean Peff	0.48		0.59		0.33	

Year variability of the production gap and effects of annual climate factors

The P_g was not significantly different across years for 'Keitt', but lower in 2018 compared to 2017 for 'Kent' (Wilcoxon's p-value = 0.002) and 'Boucodiékhal' (p-value = 0.006) (Fig. 59a). Logically, 'Kent' and 'Boucodiékhal' both showed significantly higher P_a and fruit load in 2018 than in 2017 since their P_g decreased. Nonetheless, the regression slope between P_g in 2017 and P_g in 2018 were positive and not close to -1, which revealed a non-strict alternate behavior but, instead, rather an annual variation. SFA models showed that annual climate factors had no effect on P_g and showed a high standard error of the estimate, especially *T_mean*, *T_flower*, and previous rain (standard error > 4; Table 4). Thus, the studied annual climatic factors could not explain the yearly variation of the tree performances. However, several studies demonstrated the cultivar-specific effect of environment factors on mango tree production with a major effect of water and temperature (Léchaudel and Joas 2007; Schaffer et al. 2009), and on alternate bearing (Dambreville et al. 2013). Hence, additional studies are needed to better understand environmental effects on tree phenology and production gaps over several years (more than two years) in West Africa. In addition, biotic stresses (diseases and pests), which also depended on environmental conditions, could impact the tree production at several stage of the growing cycle (Léchaudel and Joas 2007). However, biotic factors were

not included in the present study. Finer factor assessment at the tree scale and the consideration of the biotic stresses would improve the quality of the analysis. For example, measuring the water status at the tree scale might replace the data collected at a coarser resolution by the climatic station.

Effect of management practices and orchard structure on the production gap

Management practices and orchard structure factors had contrasting effects on P_g, and depended on the cultivars (Table 4). Due to the poor quality and resolution of the data on management practices, it appeared complicated to strictly decipher effects of each management practice on P_g. For instance, the annual application of fertilizers reduced the P_g of 'Kent' trees, whereas it increased the P_g for 'Boucodiékhal'. This could be explained by the fact that we compared trees from different cropping systems, which might create correlations between modalities of the management practice variables. For example, intensive orchards encompassed 83% of the irrigated trees and 61% of the pruned trees in this study. Consequently, 96% of the irrigated trees in the sample were also pruned. Correlations might also explain the positive effect of tree density on tree production. Indeed, orchards with higher mango density are intensive orchards (density of 233 trees.ha^{-1} compared to 170 trees.ha^{-1} on average for the sample) and in which inputs are applied annually. Thus, these correlations limited the use of SFA regression to estimate management effects on P_g.

Nonetheless, it was possible to separately demonstrate the effect of irrigation and fertilization practices that both appeared to be highly significant in the SFA model (Table 4). Irrigated trees showed significantly lower P_g than non-irrigated (0.41 vs. 0.60). This was especially true for the 'Kent' cultivar (P_g = 0.43 for irrigated tress compared to 0.62 for non-irrigated). However, no irrigation effect could be evidenced for the 'Keitt' trees in the sample (Fig. 59b). Only two 'Boucodiékhal' trees were irrigated in this study, resulting in a representation that was too small for statistical analysis of this cultivar. As for fertilization, trees with an annual application of fertilizer generally had a reduced P_g compared to trees with no/occasional fertilization (0.44 vs. 0.70; Fig. 59c). Consequently, when fertilizers were applied annually, P_g decreased by 43%, 38%, and 25% for 'Kent', 'Keitt', and 'Boucodiékhal', respectively compared to occasional (e.g., every two years) or no application.

Cultivar and cropping system discrepancies in tree performances

The two-way ANOVAs highlighted significant effects of cultivar, cropping system, and interaction on P_a and fruit load, P_{att}, and P_g (all p-values < 0.02; Fig. 58). The average P_g of the studied sample was of 0.52, meaning that mango trees generally reached 52% of the P_{att}. Considering the cultivars, both 'Keitt' and 'Boucodiékhal' trees reached the highest P_{att}, with an average of 408 fruits, which was around 60% higher than the 'Kent' trees (P_{att} of 258 fruits). However, due to high P_{att} but low P_a, 'Boucodiékhal' trees reached the highest P_g of 0.67 compared to 0.52 for 'Kent' and 0.41 'Keitt'. Hence, despite their high production potential, most of the trees were not very productive and almost 35% of these trees produced less than 50 fruits. Results on 'Boucodiékhal' completed the conclusions of Grechi et al. (2013) since they revealed that local cultivars were over-represented in orchards with no and low input. This indicated a low level of care for such cultivars in the Niayes region as growers take less care of local cultivars because their fruit production is usually sold on the local market (Grechi et al. 2013). After planting, these trees are not pruned (only old or falling branches are occasionally removed).

Considering the cropping systems, diversified orchards reached the highest average P_a of 217 fruits. Trees in diversified orchards had a similar fruit load of 6 fruits.m^{-3} compared to trees in intensive orchards (Fig.

58a-b). The extensive system reached the lowest production, with an average P_a of 82 fruits and a fruit load of 2.5 fruits.m^{-3}. Trees in diversified orchards had the highest P_{att} of 357 fruits compared to intensive and extensive orchards (an average of 280 fruits). Diversified and intensive orchards reached the same P_g of 0.43, while the P_g of the extensive orchards was of 0.72 (Fig. 58c-d). In addition to an equivalent P_{att} observed between extensive and intensive orchards, no significant difference in the average tree crown volume was observed between both systems (av. = 66 m^3). Thus, the low level of management practices might explain the low productivity in extensive orchards.

Under the extensive system, all cultivars showed a large P_g of 0.70, 0.74, and 0.82 for 'Kent', 'Keitt', and 'Boucodiékhal', respectively. Similarly to the intensive system, the extensive system was characterized by a predominance of 'Kent' trees (av. = 70%), more than for the diversified system (av. = 49%). Nonetheless, none of the orchards in this system were irrigated and 72% of them were not fertilized. As a consequence, despite the relative importance of the more productive 'Kent' cultivar in extensive orchards, low tree care levels impeded tree productivity. For example, 'Kent' trees in the extensive system reached a P_a of 78 fruits, while they reached a P_a of 142 fruits in the intensive system (Fig. 58a). Growers of the diversified system managed trees at an intermediate level: they applied phytosanitary product annually in 90% of the orchards and pruned trees annually, but only 22% of them directly irrigated mango trees. However, the diversified system contained large areas shared with annual cropping (av. = 9% compared to 5% in extensive and 0% in intensive systems). Annual intercrops are related to water availability for mango trees and could have a positive impact on mango tree production (Rathore et al. 2013). In the Niayes area, growers who cultivate annual crops (mostly vegetables or peanuts) are motivated by a continuous income throughout the year and the awareness that trees will benefit from annual crop irrigation (Malézieux et al. 2009). However, few researchers have studied the effect of intercropping on mango yields in West Africa.

c. Measuring the production gap at the tree scale: application and limits

In this study, we measured production as the number of fruits contained on the tree before harvest, a variable that included the harvested and non-harvested fruits. At harvest, the growers usually pick marketable fruits that are not damaged or that conform to certain criteria (in size, color, and shape), especially for the export market (Vannière et al. 2007). For example, biotic stresses like fruit flies damage the fruits and have a great impact on the exportability of the production, especially in West Africa (Vayssières et al. 2008). The distinction between marketable and non-marketable fruits was not made in our study because we only focused on factors that impacted the potential production of the tree by limiting or reducing the fruit load. Thus, harvesting practices were not studied since they had no effect on the elaboration of tree production. Finally, we are aware that counting the number of fruits did not provide complete information about tree production, which also depends on individual fruit weight. The fruit load was demonstrated to be correlated with the fruit weight (Léchaudel and Joas 2007). However, measurement of fruit weight before harvest is still a complex task, even by means of image analysis (Wang et al. 2017). Recent research will help to develop such tools but they still require improvement regarding their cost and adaptation to the context of complex orchards.

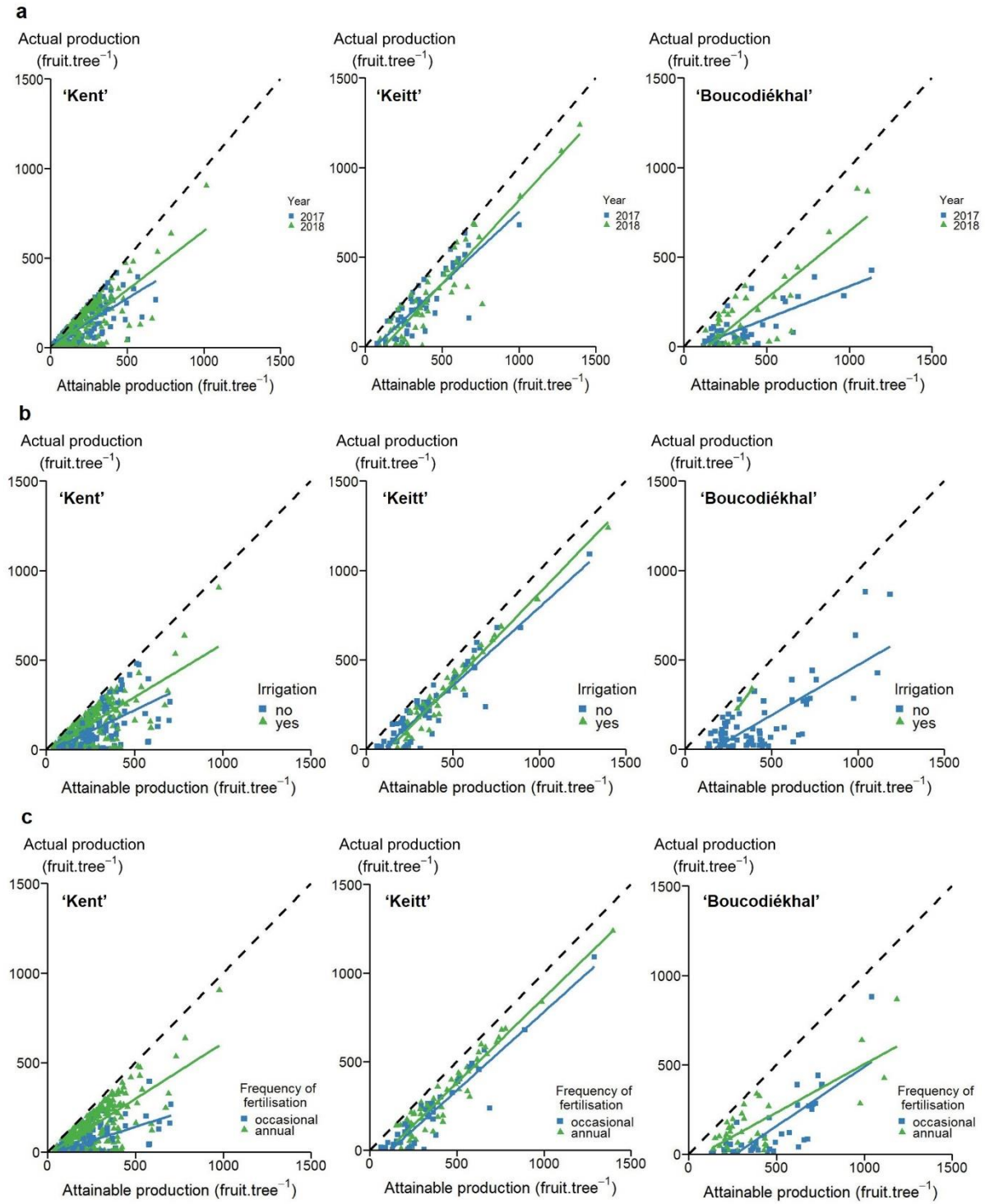


Fig. 59. Attainable production (P_{att}) compared to actual production (P_a) for the three studied cultivars ('Kent', 'Keitt', and 'Boucodiékhal') with emphasis on (**a**) year; (**b**) irrigation; and (**c**) fertilization. P_{att} is estimated from stochastic frontier functions. The dotted line represents the 1:1 equation.

Nonetheless, the present study provides the first estimates of attainable mango production and production gaps at the tree scale. At the farm or field scale, several studies have estimated mango yield gaps, but by analyzing different orchard performances and using different methods. For instance, Mensah and Brümmer (2016) computed the stochastic frontier of the mango quantity (in kg) produced per orchard in Ghana and highlighted that socio-economic and management variables were responsible for a yield gap of 21 to 52%, depending on the studied region. Another study, carried out by de Bie (2004), analyzed several effects on mango growers' income (economic profit per hectare) in Thai orchards using regression models. They established that yield gaps at the orchard scale were firstly explained by management factors (49% of yield gap explained), followed by environmental factors (30%), and a year effect (21%). However, neither study provided estimates of attainable mango production in the local context and nor of the effect of endogenous and exogenous factors on production.

Compared to previous studies, the present study made it possible to estimate attainable production and production gaps for trees from three different cropping systems. SFA models allowed us to separate the different types of factors impacting attainable tree production (i.e., production-limiting factors) from those impacting the production gap (i.e., production-reducing factors). This adaptation of the yield gap concept at the tree scale allowed us to integrate the endogenous tree factors (i.e., tree cultivar, structure, and age) and their variabilities. Indeed, in West Africa, tree structure, cultivar, and age can greatly vary within and between orchards. These production-limiting factors are endogenous to the tree and appear difficult or even impossible (age) to improve through by the growers' practices. Thus, production-limiting factors have to be studied separately from production-reducing factors that can be managed by the grower.

This study allowed us to identify potential ways to improve production improvements for growers in West Africa. Introducing tree management (irrigation, pruning, etc.) on local cultivars in the orchard would be valuable for mango productivity (Fig. 59b-c). More generally, improving management practices in smallholder orchards would reduce the production gap and increase orchards yields (Vannière et al. 2007). Diversified orchards are characterized by a higher Patt compared to intensive orchards (Fig. 58). Thus, an equivalent reduction of the production gap will be more profitable in diversified orchards than in intensive orchards. For example, we can establish that reducing the Pg of 'Keitt' trees by one point would increase the production of 45 fruits in diversified orchards, compared to an increase of 38 fruits in intensive orchards. However, further studies must analyze the costs (e.g., additional inputs) and benefits for the grower when reducing the mango production gap (van Ittersum et al. 2013). One limitation of this study was the poor quality and resolution of the data on annual climate and management practices, and the absence of consideration of the biotic stresses. A better assessment of management and climatic factors by measuring the inputs at the tree scale would improve the accuracy of the SFA estimates. More generally, fruit crops in West Africa require a long-term effort for accurate data collection to better understand the productivity variability among cropping systems and to improve agricultural practices (Carletto et al. 2015). Finally, using the production gap concept developed in would allow us to assess the efficiency of all the trees in an orchard. This would make it possible to compute the production gap at the orchard scale and to map the within-field variability of the gap as has already been done with tree production (Sarron et al. 2018).

III.4. Conclusion

This study presents the development of yield gap methodology to include perennial specificities for studying the production gap at the tree scale. Production gap (P_g) was defined as the difference between the attainable production (P_{att}), defined by the production-limiting factors (i.e., tree cultivar, age, and structure), and the

actual production (P_a, in number of fruits). Stochastic frontier analysis facilitated the estimation of both P_{att} and P_g at the individual tree scale and identification of the production-reducing factors (i.e., annual climate, management practices, and orchard structure) affecting P_g. This method was applied on 300 mango trees of three cultivars representative of the tree structure and age of three cropping systems (extensive, diversified, and intensive) in West Africa. 'Keitt' trees had the lowest P_g (0.41), compared to 'Kent' (P_g = 0.52), whereas the local cultivar ('Boucodhiékhal') had the highest P_g (0.67), even if it displayed the highest production ability (P_{att} of 426 fruits). These estimates also made it possible to draw conclusion about the performances of the three studied cropping systems: the extensive system had the lowest efficiency (P_g = 0.72) due to low management intensity and, conversely, the diversified system had an efficiency similar to the one of the intensive system (P_g = 0.43). Results suggested that fruit load variations and production gaps were mainly caused by production-reducing factors rather than production-limiting factors. In addition, irrigation and fertilization were shown to have positive effects on tree production, thus contributing to a reduction in P_g, especially for 'Kent' trees. In conclusion, this study achieved the double goal of accurately estimating tree productivity under different cropping systems and analyzing the effect of the production-limiting and production-reducing factors driving the tree productivity.

Acknowledgments: This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empir Econ 20:325–332. https://doi.org/10.1007/BF01205442

Borianne P, Sarron J, Borne F, Faye É (2019) Deep Mangoes: from fruit detection to cultivar identification in color images of mango trees. In: DISP'19, International Conference on Digital Image and Signal Processing. Oxford, UK, p 7

Boudon F, Jestin A, Briand A-S, et al (2017) The role of structural and temporal factors in the architectural development of the mango tree: evidences from simulation. Acta Hortic 83–90. https://doi.org/10.17660/ActaHortic.2017.1160.12

Carletto C, Jolliffe D, Banerjee R (2015) From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. J Dev Stud 51:133–148. https://doi.org/10.1080/00220388.2014.968140

Chopin P, Blazy J-M (2013) Assessment of regional variability in crop yields with spatial autocorrelation: Banana farms and policy implications in Martinique. Agric Ecosyst Environ 181:12–21. https://doi.org/10.1016/j.agee.2013.09.001

Coelli T, Henningsen A (2017) frontier: Stochastic Frontier Analysis. R package version 1.1-2

Dambreville A, Lauri P-E, Trottier C, et al (2013) Deciphering structural and temporal interplays during the architectural development of mango trees. J Exp Bot 64:2467–2480. https://doi.org/10.1093/jxb/ert105

de Bie CAJM (2004) The yield gap of mango in Phrao, Thailand, as investigated through comparative performance evaluation. Sci Hortic 102:37–52. https://doi.org/10.1016/j.scienta.2003.11.017

FAOSTAT (2017) FAO, Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/faostat/en/#home

Fare Y, Dufumier M, Loloum M, et al (2017) Analysis and Diagnosis of the Agrarian System in the Niayes Region, Northwest Senegal (West Africa). Agriculture 7:59. https://doi.org/10.3390/agriculture7070059

Grechi I, Sane CAB, Diame L, et al (2013) Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68:447–466. https://doi.org/10.1051/fruits/2013094

Koirala A, Walsh KB, Wang Z, McCarthy C (2019) Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO.' Precis Agric. https://doi.org/10.1007/s11119-019-09642-0

Léchaudel M, Joas J (2007) An overview of preharvest factors influencing mango fruit growth, quality and postharvest behaviour. Braz J Plant Physiol 19:287–298. https://doi.org/10.1590/S1677-04202007000400004

Malézieux E, Crozat Y, Dupraz C, et al (2009) Mixing plant species in cropping systems: concepts, tools and models. A review. Agron Sustain Dev 29:43–62. https://doi.org/10.1051/agro:2007057

Mensah A, Brümmer B (2016) Drivers of technical efficiency and technology gaps in Ghana's mango production sector: A stochastic metafrontier approach. Afr J Agric Resour Econ 11:17

Menzel CM, Le Lagadec MD (2017) Can the productivity of mango orchards be increased by using high-density plantings? Sci Hortic 219:222–263. https://doi.org/10.1016/j.scienta.2016.11.041

Neumann K, Verburg PH, Stehfest E, Müller C (2010) The yield gap of global grain production: A spatial analysis. Agric Syst 103:316–326. https://doi.org/10.1016/j.agsy.2010.02.004

Oguntunde PG, Fasinmirin JT, van de Giesen N (2011) Influence of Tree Age and Variety on Allometric Characteristics and Water Use of Mangifera indica L. Growing in Plantation. J Bot 2011:1–8. https://doi.org/10.1155/2011/824201

Ramírez F, Davenport TL (2010) Mango (Mangifera indica L.) flowering physiology. Sci Hortic 126:65–72. https://doi.org/10.1016/j.scienta.2010.06.024

Rathore AC, Saroj PL, Lal H, et al (2013) Performance of mango based agri-horticultural models under rainfed situation of Western Himalaya, India. Agrofor Syst 87:1389–1404. https://doi.org/10.1007/s10457-013-9646-5

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems. pp 91–99

Rey J-Y, Diallo TM, Vannière H, et al (2007) The mango in French-speaking West Africa: varieties and varietal composition of the orchards. Fruits 62:57–73. https://doi.org/10.1051/fruits:2006051

Sarron J, Malézieux É, Sané C, Faye É (2018) Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. Remote Sens 10:1900. https://doi.org/10.3390/rs10121900

Schaffer B, Urban L, Lu P, Whiley AW (2009) Ecophysiology. In: Litz RE (ed) The mango: botany, production and uses. CABI, Cambridge, USA, pp 170–209

Silva JV, Reidsma P, Laborte AG, van Ittersum MK (2017) Explaining rice yields and yield gaps in Central Luzon, Philippines: An application of stochastic frontier analysis and crop modelling. Eur J Agron 82:223–241. https://doi.org/10.1016/j.eja.2016.06.017

van Ittersum MK, Cassman KG, Grassini P, et al (2013) Yield gap analysis with local to global relevance—A review. Field Crops Res 143:4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Vannière H, Didier C, Rey J-Y, et al (2007) The mango in French-speaking West Africa: cropping systems and agronomical practices. Fruits 62:187–201. https://doi.org/10.1051/fruits:2007014

Vayssières J-F, Korie S, Coulibaly O, et al (2008) The mango tree in central and northern Benin: cultivar inventory, yield assessment, infested stages and loss due to fruit flies (Diptera Tephritidae). Fruits 63:335–348. https://doi.org/10.1051/fruits:2008035

Wang Z, Walsh K, Verma B (2017) On-Tree Mango Fruit Size Estimation Using RGB-D Images. Sensors 17:2738. https://doi.org/10.3390/s17122738

CHAPITRE 2

Extrapolation et analyse du rendement à l'échelle du verger

Introduction

Ce chapitre présente les résultats relatifs à l'estimation du rendement de mangues depuis l'échelle de l'arbre à l'échelle du verger (voir Fig. 32, p. 66). Une première méthode (Chapitre 2 – Partie I) consiste à utiliser l'imagerie drone et la modélisation. À partir de la carte d'occupation des sols et du Canopy Height Model (CHM) des 29 vergers de l'échantillon (voir Matériels et Méthodes), trois paramètres de structure des manguiers ont été calculés : la hauteur de l'arbre ainsi que la surface et le volume projeté du houppier. Ces informations sont utilisées pour calibrer des modèles d'estimation de la production de l'arbre à partir des données de production mesurées à l'aide des outils présentés dans le Chapitre 1. Dans chaque modèle est ajoutée une variable appelée indice de charge ('*load index*') permettant de prendre en compte les variations annuelles de la production de l'arbre. En pondérant les estimations des modèles avec l'indice de charge moyen du verger, les rendements sont obtenus pour chaque verger. Cette méthodologie d'estimation et de spatialisation du rendement utilisant le drone constitue une première sur les plantes pérennes.

La seconde méthode (Chapitre 2 – Partie II) mise en place pour estimer le rendement des vergers est basée sur l'échantillonnage. Les outils d'estimation de la production ont été utilisés pour cartographier directement la production des arbres en photographiant chaque arbre dans trois vergers représentatifs des trois systèmes de cultures étudiés. Cette cartographie, plus précise que la précédente, a permis d'étudier les variabilités de production intra-verger. De plus, plusieurs indices d'écologie spatiale (indice de Moran, L(r) de Ripley) ont été utilisés pour décrire les hétérogénéités spatiales du volume, de la production et du taux de charge (production divisée par le volume) des manguiers. Par ce travail, nous avons démontré que l'hétérogénéité des caractéristiques des arbres (volume, production et taux de charge) variait d'un verger à l'autre. Nous avons testé la prise en compte de ces hétérogénéités dans la construction de méthode d'échantillonnage pour l'estimation du rendement du verger sans utiliser les modèles drone.

En utilisant les modèles drone, le rendement de chaque verger a pu être estimé sur les deux années d'étude (2017-2018). De plus, les cartes d'occupations des sols donnent des informations sur la structure des arbres et du verger (composition spécifique, densité de plantation et configuration spatiale des arbres). Nous avons utilisé ces informations, additionnées de données sur l'environnement et les pratiques culturales, pour réaliser un diagnostic agronomique régional (Doré et al. 2008) des variations de rendements entre les vergers à base de manguiers au Sénégal. Nous avons mis en évidence les différences de rendement entre les trois types de systèmes. Ces différences sont induites par un effet de la densité de plantation, tandis que la production moyenne des arbres est expliquée en premier lieu par la structure des arbres. L'analyse souligne ensuite l'effet des pratiques culturales et de la diversité spécifique sur la production, le rendement et le taux de charge (rendement divisé par le volume).

I. Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV

Les premiers travaux de la thèse à l'échelle de l'arbre permettent de s'intéresser à l'estimation du rendement à l'échelle du verger (Fig. 32, p. 66). De plus, l'utilisation du drone permet d'obtenir des informations précises sur la structure des arbres qui est corrélée à la production de l'arbre (voir État de l'art). Cette Partie I est un article présentant la méthodologie d'estimation des rendements des vergers de mangues en utilisant les informations acquises par drone (carte d'occupations des sols et CHM) et la modélisation. Ces modèles drones ont été validés sur 60 arbres et obtiennent des performances supérieures à l'imagerie satellitaire (R^2 supérieurs à 0.77 et des RMSE% compris entre 20 et 29%). Cette méthodologie est la première utilisant le drone pour l'estimation de rendement d'arbre fruitiers à l'échelle du verger.

Cet article scientifique a été publié en novembre 2018 à une revue à comité de lecture : *Remote Sensing*. Les annexes de cet article sont placés dans la partie Annexe IV en fin du mémoire.

- Sarron J.; Malézieux É.; Sané C.A.B.; Faye É. **Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV**. Remote Sens 10:1900. https://doi.org/10.3390/rs10121900

Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV

Julien Sarron [1,2,3], Éric Malézieux [1,2], Cheikh Amet Bassirou Sané [4] and Émile Faye [1,2,3,*]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]HortSys, University Montpellier, CIRAD, F-34090 Montpellier, France
[3]Centre pour le Développement de l'Horticulture, ISRA, 14000 Dakar, Senegal
[4]Université Cheikh Anta Diop (UCAD), Fac. Sci. Tech., BP 5005 Dakar, Senegal
*Correspondence: emile.faye@cirad.fr

Received: 23 October 2018; Accepted: 26 November 2018; Published: 28 November 2018

Abstract

In the value chain, yields are key information for both growers and other stakeholders in market supply and exports. However, orchard yields are often still based on an extrapolation of tree production which is visually assessed on a limited number of trees; a tedious and inaccurate task that gives no yield information at a finer scale than the orchard plot. In this work, we propose a method to accurately map individual tree production at the orchard scale by developing a trade-off methodology between mechanistic yield modelling and extensive fruit counting using machine vision systems. A methodological toolbox was developed and tested to estimate and map tree species, structure, and yields in mango orchards of various cropping systems (from monocultivar to plurispecific orchards) in the Niayes region, West Senegal. Tree structure parameters (height, crown area and volume), species, and mango cultivars were measured using unmanned aerial vehicle (UAV) photogrammetry and geographic, object-based image analysis. This procedure reached an average overall accuracy of 0.89 for classifying tree species and mango cultivars. Tree structure parameters combined with a fruit load index, which takes into account year and management effects, were implemented in predictive production models of three mango cultivars. Models reached satisfying accuracies with R2 greater than 0.77 and RMSE% ranging from 20% to 29% when evaluated with the measured production of 60 validation trees. In 2017, this methodology was applied to 15 orchards overflown by UAV, and estimated yields were compared to those measured by the growers for six of them, showing the proper efficiency of our technology. The proposed method achieved the breakthrough of rapidly and precisely mapping mango yields without detecting fruits from ground imagery, but rather, by linking yields with tree structural parameters. Such a tool will provide growers with accurate yield estimations at the orchard scale, and will permit them to study the parameters that drive yield heterogeneity within and between orchards.

Keywords: unmanned aerial vehicle; mango orchard; yield estimation; fruit detection; tree architecture; random forest; GEOBIA; structure-from-motion

I.1. Introduction

Mango (*Mangifera indica* L.) is a major fruit crop of the tropics and sub-tropics that guarantees the incomes and food security for local populations (Mukherjee and Litz 2009). In West Africa, to meet an ever-increasing fruit demand from local and international markets (van Ittersum et al. 2016), various mango cropping systems co-exist, from small, family-based, diversified orchards to large, commercial-based,

monospecific orchards (Rey et al. 2007; Grechi et al. 2013). In this region, more than 20 polyembryonic and monoembryonic cultivars were featured by Rey et al. (2007). However, the lack of accurate and appropriate measurement practices of horticultural production is still hampering the development of the mango sector, resulting in poor data reliability in this region (Carletto et al. 2015). Mango yield, defined here as the production per unit of area (e.g., $t.ha^{-1}$), estimated as early as possible before harvest, is a key to informing growers about cropping practices in order to enable them to address the physiological needs of each tree (pruning, watering, etc.) and to plan orchard management (harvest, packing, etc.), but also to advise regional agricultural policies (Carletto et al. 2015). At present, mango yield estimation is still based on the visual inspection of a limited number of trees on which fruits are manually counted, which is a tedious and time-consuming method that relies on the observer reliability and accuracy (Gongal et al. 2015). Yield estimation on the orchard scale then results from the multiplication of the average fruit number per sampled tree with the orchard tree density (i.e., number of trees per unit of area). Depending on the growers, this estimation is undertaken between two months and two weeks before harvest. At harvest time, orchard production is measured by counting the number of harvested fruit buckets and multiplying it by an average weight. Moreover, these inaccurate assessments of yield and production are mostly conducted in commercial and homogeneous orchard plots, and rarely in small, diversified orchards. Both methods provide low accuracy and give no yield information at a finer scale than the orchard plot (i.e., plot portion or tree scale), that sometimes exceeds 10 hectares. Whereas yield data may be accessible in commercial orchards with a quantifiable level of error, little and often no information is available in family-based orchards were harvest tasks are mostly performed by tradespeople. Additionally, in West Africa other, orchard characteristics such as acreage and planting density are unknown by the growers, or at best, are only roughly estimated. Hence, one of the underlying factors that explains the poor quality of data available is the lack of accurate and effective tools to estimate production and, to a larger extent, measure other orchard characteristics (acreage, planting density, etc.) (Carletto et al. 2015). In the meantime, providing growers with accurate yield maps will inform them about precise farming management, and will also help researchers to study the parameters driving yield heterogeneity within and between orchards (Carletto et al. 2013; Gongal et al. 2015).

One way to estimate yield in the field is to use crop models that evaluate plant yield responses to its genotype, environment, and cropping system (van Wart et al. 2013). However, models of tree crops are sparse, and there is currently no complete mango crop model forecasting production (Normand et al. 2015). Indeed, mango trees are characterized by strong vegetative and reproductive asynchronisms and irregular bearings explained by interplays between several endogenous (cultivar, temporal phenology, etc.) and exogenous (climate, soil, agricultural practices, etc.) factors (Dambreville et al. 2013). For instance, the relationship between flowering and the number of fruits remains unclear, and the effects of cultivars and previous growing cycles on this relationship has been evidenced (Dambreville et al. 2013). These features hamper the development of a mechanistic model to estimate mango tree production, and compel growers to measure yield empirically.

Another way to estimate orchard yields is to measure individual tree production (e.g., kg of fruits per tree) prior to harvest. This alternative allows growers to get rid of the effect of the tree physiological factors and environmental factors on yields, and gives a precise view of the actual orchard yield. In the last decade, many studies have focused on a computer vision systems for fruit detection and counting at the tree scale (Gongal et al. 2015). These systems combine an image acquisition method using one or several sensors including visible (RGB) cameras (Payne et al. 2013; Qureshi et al. 2014; Bargoti and Underwood 2017),

multispectral/hyperspectral cameras (Okamoto and Lee 2009), and LiDAR (Underwood et al. 2016; Stein et al. 2016), with an image processing algorithm. These tools make it possible to accurately count the number of fruits on the tree, albeit with some limitations. Firstly, even with the best algorithm, visual occlusions of fruits hinder the detection of all the fruits on the tree. Consequently, a labor-intensive field calibration is required to estimate the actual tree production (relationship between the image-based fruit count and the actual fruit number in the tree) (Bargoti and Underwood 2017). Secondly, most developed tools focus on tree scale, with no consideration of yield estimation or extrapolation on the entire orchard. Few state-of-the-art studies estimate orchard yields by means of a mobile platform to track and create a multi-view of all trees in the orchard. For example, Stein et al. (2016) used a mobile robot equipped with RGB and LiDAR sensors to map tree production and structure in a commercial orchard. While RGB images are used for multi-view fruit detection, the LiDAR sensor generates an accurate 3D point cloud which is able to accurately segment each tree crown. Then, each detected fruit is associated with its corresponding tree. Whereas such tools reach high accuracies for trees production estimation and mapping, they remain expensive and are not applicable to complex orchards (i.e., orchards with random tree arrangement).

Alternatively, to tackle the limitations of tree production estimation by ground remote sensing, one solution is to use above-ground imagery to assess tree features and link them to the yield. Forecasting of orchard yield based on the measurement of vegetation indices and tree crown area by multispectral satellite imagery was tested on mango plantations (Rahman et al. 2018). However, the results showed low consistency of the relationship between yield and tree features across different orchards, and the method remained dependent on expensive equipment. Several investigations have revealed the advantages of unmanned aerial vehicles (UAVs) for precise land cover and tree architecture mapping in forest stands (Torresan et al. 2017) and orchards (Díaz-Varela et al. 2015; Torres-Sánchez et al. 2015; Jiménez-Brenes et al. 2017) that could be used for tree production estimations. The miniaturization of sensors, low operation and equipment costs, and high flexibility in time scheduling and avoidance of cloud cover made UAVs appropriate for land cover and yield mapping. Recent studies have used UAVs with RGB sensors (Lisein et al. 2015; Birdal et al. 2017; Giannetti et al. 2018) to accurately estimate tree structure and map orchard land cover through structure-from-motion and photogrammetry (Lisein et al. 2015; Jiménez-Brenes et al. 2017). Indeed, by flying low with high overlaps, UAVs capture very high resolution (VHR) geotagged images that made it possible to build orthomosaics and digital surface models (DSM) at cm-level resolution. And thanks to this fine spatial resolution (finer than the object of interest), geographic object-based image analysis (GEOBIA)—that has proven to more effectively map land cover than pixel-based algorithms (Blaschke 2010) —can be used. GEOBIA also makes it possible to delineate and summarize objects following their spectral, shape, textural, and contextual information. UAVs and GEOBIA procedures have been widely used for tree species classification and delineation, and to characterize tree structure. For example, Torres-Sánchez et al. (2015) delineated olive trees with a GEOBIA procedure and accurately quantified tree heights and crown areas and volumes in olive orchards with UAVs. Regarding tree species classification, Michez et al. (2016) succeeded in classifying 5 deciduous tree species with a 95% accuracy rate using a UAV equipped with a hyperspectral sensor. However, land cover mapping and tree structure measurements are often described in separate studies. Few studies have proposed an integrative methodology to estimate both tree species (or cultivars) and structures in complex agro-ecological landscapes such as diversified orchards (Nevalainen et al. 2017). Moreover, whereas ground machine vision systems for yield estimation and mapping are effective, the efficiency of UAVs for such purposes has never been tested to our knowledge. Therefore, UAVs might be more interesting than ground machine vision systems due to their covering of the entire orchard, low cost, and high flexibility.

In what follows, we develop a comprehensive set of tools for quantifying the structure of mango trees, and compute orchard land covers in order to accurately estimate and map the production at the orchard scale. The procedure is based on the acquisition of VHR images with a commercial UAV to build an orthomosaic and a DSM of the orchard, upon which a GEOBIA procedure is run to determine tree species and mango cultivars. The produced land cover maps combined with Canopy Height Model (CHM) permitted us to compute tree structure parameters (tree height, crown area, and volume). Using a load index that integrates climate, year, and management effects, and the tree structure parameters as independent variables in tree production models, we mapped the individual mango tree production of the entire orchard. This procedure has been tested and evaluated in fifteen mango orchards of the Niayes region in West Senegal, ranging from family-based, plurispecific orchards, to monocultivar, commercial orchards. This toolbox constitutes a unique, easy-to-use, and useful methodology to semi-automatically map tree production at the orchard scale using UAV photogrammetry and remote sensing. This toolbox stands as a trade-off methodology between complex mechanistic tree production models and labor-intensive empirical measurements of fruit number at the tree level using machine vision systems.

I.2. Materials and Methods

After a description of the study sites, the following paragraphs describe the workflow for mango yield mapping at the orchard scale (Fig. 60).

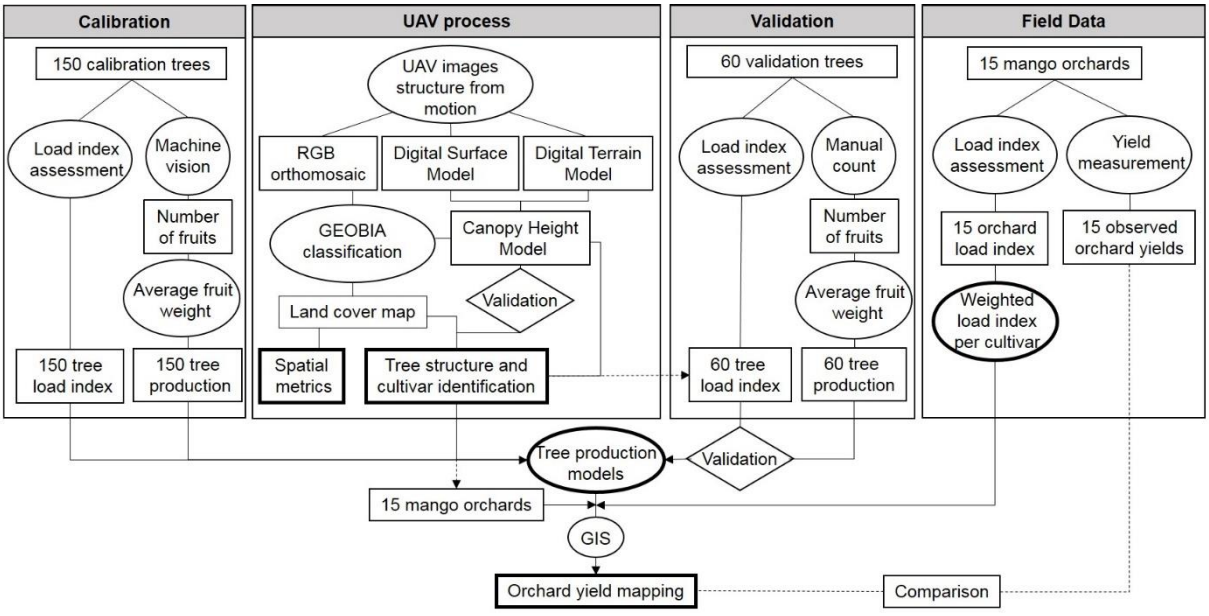


Fig. 60. Schematic workflow of the main steps for orchard yield mapping. Each step is fully detailed in Materials and Methods (I.2).

a. Study Sites

The study was carried out in the Niayes region in western Senegal, which represents a major production basin of mangoes for local markets and the first for export markets (Rey et al. 2007). The Niayes region is part of the Sudano-Sahelian zone, and is characterized by unimodal rainfall from July to September (mean rainfall of 450 mm per year) and a relatively fresh and humid climate generated by the proximity to the Atlantic Ocean (with daily mean temperatures ranging from 18 °C to 30 °C). The region is characterized by

a succession of dunes with sandy, non-lessive soils and depressions with clay and peaty soils (Fare et al. 2017). These topographic depressions—where ground water reaches the surface (shallow sedimentary aquifer retained by impermeable clay deposits)—combined with the the climatic conditions make this region the major fruit and vegetable production area in Senegal (Fare et al. 2017).

Fifteen mango-based orchards (Table 5) have been selected to represent the diversity of cropping systems found in this horticultural area, from small, family-based, diversified orchards, including other tree species (citrus, cashew, etc.) (n = 9), to large, commercial-based, monospecific orchards (n = 6) (Grechi et al. 2013). Due to the high diversity of cropping systems, the samples have shown a large variability in mango yields (as estimated by the growers). Indeed, mango yield (expressed in $t.ha^{-1}$) variability depends both on the orchard design (e.g., tree specific diversity, planting density, and pattern) and agricultural practices which partly drive the orchard production. In our sampling, tree specific diversity ranged from a unique mango cultivar to more than five cultivated tree species and at least four mango cultivars. Additionally, orchard planting pattern and density were highly variable from orchards with continuous in-row canopy to orchards with a random tree arrangement. In 2017, the actual yield of each orchard was asked of the growers.

Table 5. Description of flight missions and classification for the 15 studied orchards.

# [1]	Flight Area [ha]	Training Rate [%]	Number of Class		Latitude [2]	Longitude [2]
			Level 1	Level 2		
1	4.8	4.1	3	1	14°49′41″N	17°5′48″W
2	6.4	4.6	3	1	14°59′1″N	17°0′14″W
3	6.6	6.3	3	1	14°59′39″N	16°59′25″W
4	6.5	7.3	4	1	14°49′32″N	17°6′6″W
5	7.4	5.6	3	3	14°47′53″N	17°8′32″W
6	6.4	4.8	5	3	14°48′3″N	17°8′17″W
7	4.6	5.4	7	2	14°47′27″N	17°13′41″W
8	8.8	3.4	8	2	14°48′15″N	17°12′35″W
9	8.3	3.9	7	3	14°46′24″N	17°6′31″W
10	4.9	5.2	8	3	14°45′45″N	17°9′7″W
11	7.9	6.5	9	2	14°49′37″N	17°9′22″W
12	8.4	4.9	8	3	14°50′35″N	17°7′11″W
13–14	8.2	6.6	8	4	14°47′20″N	17°13′36″W
15	20.5	2.5	9	3	14°47′25″N	17°12′18″W
Mean	7.8	5.0				

[1] Orchards have been numbered in ascending order of number of classes. [2] UTM coordinates system (zone 28N), Datum WGS84.

b. Production Estimation on Calibration Trees

In each orchard, ten trees were selected among the three main cultivars found in the study area, i.e., 'Kent', 'Keitt' and 'Boucodiékhal' (BDH), to depict the entire variability of tree structure (trunk perimeter, height, and crown volume), age, and fruit load found in the study area. Thus, this set of 150 mango trees made of 89 'Kent', 39 'Keitt', and 22 'BDH' were used to calibrate the predictive models by measuring their production using the machine vision system described below. Then, the representativeness of the calibration trees was tested by comparing their structure assessed by UAV with the structure of all mango trees within the 15 orchards. For this purpose, independent comparisons of means between the two groups were conducted using Student t-test at a 95% confidence level.

During June 2017, the individual tree production from one month to two weeks before harvest was measured on the 150 calibration trees by means of RGB ground image analysis. The machine vision system used for this task is described in Sarron et al. (2018). Briefly, on two opposite pictures of the tree captured with a hand-held Sony Nex-7 RGB camera (Sony Corporation, New York, NY, USA), a pixel k-nearest neighbor (KNN) supervised classification allowed us to detect mango fruits under heterogeneous field conditions. The number of automatically-detected fruits on the images was linked to the actual number of fruits carried by the tree (visually counted in the field on 52 trees) with a robust linear regression for each cultivar. This method has been cross-validated on 52 additional mango trees (Sarron et al. 2018). The estimated fruit number per tree was then converted into individual production (in kg) for the 150 calibration trees using the average fruit weight measured on 90 harvested fruits for each cultivar (Table S1 in Supplementary Materials).

c. Orchard Land Cover Mapping and Tree Structure Assessment Using UAV

Individual tree detection and classification at the orchard scale were key steps in our workflow to quantify orchard land cover and tree structure (height, crown volume and area), both of which are mandatory to estimate mango yields. The procedure described below consisted of (1) UAV flights for the acquisition of VHR images, (2) generation of orthomosaics and CHM, (3) land cover mapping using advanced GEOBIA; and (4) the extraction and validation of individual tree structures using GIS.

UAV System and Flight Missions

In 2017, the 15 orchards were overflown using a DJI Mavic Pro quadricopter (DJI Inc., Shenzhen, China) equipped with a visual sensor (red, green, and blue—RGB) of 12 megapixels with a focal length of 35 mm. Flight missions were planned using Pix4Dcapture application (Pix4Dcapture 3.2; Pix4D SA, Lausanne, Switzerland) that automatically flew the UAV and triggered the camera to maintain a forward and side overlap of 80% and 70% between images, respectively. Images acquired during flights were automatically geotagged using the onboard GPS receiver. The flight height was set to 40 m, resulting in a ground sample distance of 1.30 cm.pixel^{-1} (Faye et al. 2016). The flight grids covered the entire orchards plus the surrounding landscape, leading to captured areas ranging from 4.6 to 8.8 ha (Table 5). Because of its large area (20.5 ha), orchard #15 was mapped using two successive flights. Due to the large period of UAV-image acquisition, mango tree phenological stages varied between vegetative growth, flowering, and fruit set.

Orthomosaics and CHM Generation

Geotagged VHR images were used to generate RGB orthomosaics and a digital surface model (DSM) of the 15 orchards. Orthomosaics (i.e., a georeferenced aerial image which has been geometrically corrected) of each orchard derived from structure-from-motion image reconstruction using the images acquired during the flight mission (Lisein et al. 2015). The process was fully automated under Pix4Dmapper Pro software (Pix4Dmapper 1.3; Pix4D SA, Lausanne, Switzerland). We used software parameters that enhanced the accuracy of the orthomosaic and DSM construction: that is, a minimum of two keypoint matches for 3D point setting, a high level of densification of the point cloud, and a sharp DSM filter to preserve the orientation of the surface and to keep sharp 3D features (Díaz-Varela et al. 2015). Following this process, one RGB orthomosaic, one DSM, and one digital terrain model (DTM) were generated for each orchard (UTM coordinates system zone 28N; datum WGS84). DTM was automatically generated by the software that filtered out non-ground points from the point cloud before a smoothing operation.

As the DSM does not correspond directly to tree height in relation to the ground surface, a Canopy Height Model (CHM) was computed as the subtraction of the DTM from the DSM (Birdal et al. 2017). Orthomosaic and CHM of each orchard had a spatial resolution of 1.30 and 6.40 cm.pixel^{-1}, respectively, and were exported as GeoTIFF files to be further used in remote sensing and in the GIS software.

Land Cover Mapping Using GEOBIA

The RGB orthomosaic combined with the CHM produced a four-band, multi-layer image which has been analyzed by a GEOBIA algorithm implemented in eCognition Developer 9 software (Trimble Geospatial, Munich, Germany). The aims of this algorithm were to (1) delineate objects such as individual tree crown, and (2) classify each object to obtain a precise land cover map.

Two segmentations were applied using a multiresolution segmentation algorithm (Baatz and Schäpe), which is a widely-used method for land cover classification from VHR images (Díaz-Varela et al. 2015; Ma et al. 2017). We applied segmentation on all the four bands (R, G, B, and CHM) with an equal weight for each band (Torres-Sánchez et al. 2015). The segmentation scale parameter—i.e., fixing the final size of the objects—and homogeneity criteria (shape and compactness) (Michez et al. 2016) were adjusted by expert judgment and visual interpretation. Since the orthomosaics covered the entire orchards and the surrounding areas, they showed a variety of object types from small trees or annual crop plots to large buildings (Fig. 61). This heterogeneity led to variability in object size, shape, and spectral signatures. We firstly applied a multiresolution segmentation with a coarse scale parameter of 1000, and homogeneity criterion values of 0.1 and 0.5, for shape and compactness, respectively. This coarse segmentation allowed us to delineate large and homogeneous objects of buildings and soil that were classified by thresholds on brightness, Excess Green (Torres-Sánchez et al. 2015), and Excess Red (Meyer and Neto 2008) indices. Secondly, a finer multiresolution segmentation was applied to the remaining non-classified objects with a scale parameter of 200 and homogeneity criterion values of 0.1 and 0.5 for shape and compactness, respectively.

All objects produced by the segmentations were firstly classified (level 1 classification) into ten classes (Table 6) comprising four tree species: “citrus”, “mimosae”, “cashew” and “mango”. This classification included also small heterogeneous “building” and “soil” objects that were not removed by the thresholding after the first segmentation. In order to identify mango cultivars, a second classification (level 2) was applied to classify “mango” objects into 4 cultivar classes: “Kent”, “Keitt”, “BDH” and “other cultivar”. Both classifications used the Random Forest (RF) classifier implemented in eCognition, which is a supervised machine-learning technique that is widely used in the field of remote sensing of agricultural and forest areas using satellite and UAVs (Belgiu and Drăguţ 2016; Ma et al. 2017). For a detailed description of RF functioning, see the review of Belgiu & Drăguţ (2016). In our case, eCognition was parametrized to use a max Random Tree number of 500 and the number of active variables was set to the square root of the number of features, as recommended in the literature (Belgiu and Drăguţ 2016; Lebourgeois et al. 2017). As RF performs better after feature optimization (Ma et al. 2017), a feature selection was done on one representative orthomosaic. Twenty-one active features were computed for each object: 15 spectral features corresponding to the mean and standard deviation for each of the four bands, brightness, values of hue, saturation, intensity, and Normalized Difference Index (NDI, computed on R and G bands) (Payne et al. 2013), Excess Green and Excess Red indices; and 6 texture features corresponding to entropy, standard deviation, correlation, mean, homogeneity, and contrast derivatives of Haralick’s gray level co-occurrence matrix (GLCM) (Haralick et al. 1973). Feature selection was performed using a ranking of importance

features given by two decision trees in eCognition, i.e., one for each level of classification. Selected features for RF optimization are presented in Table 6.

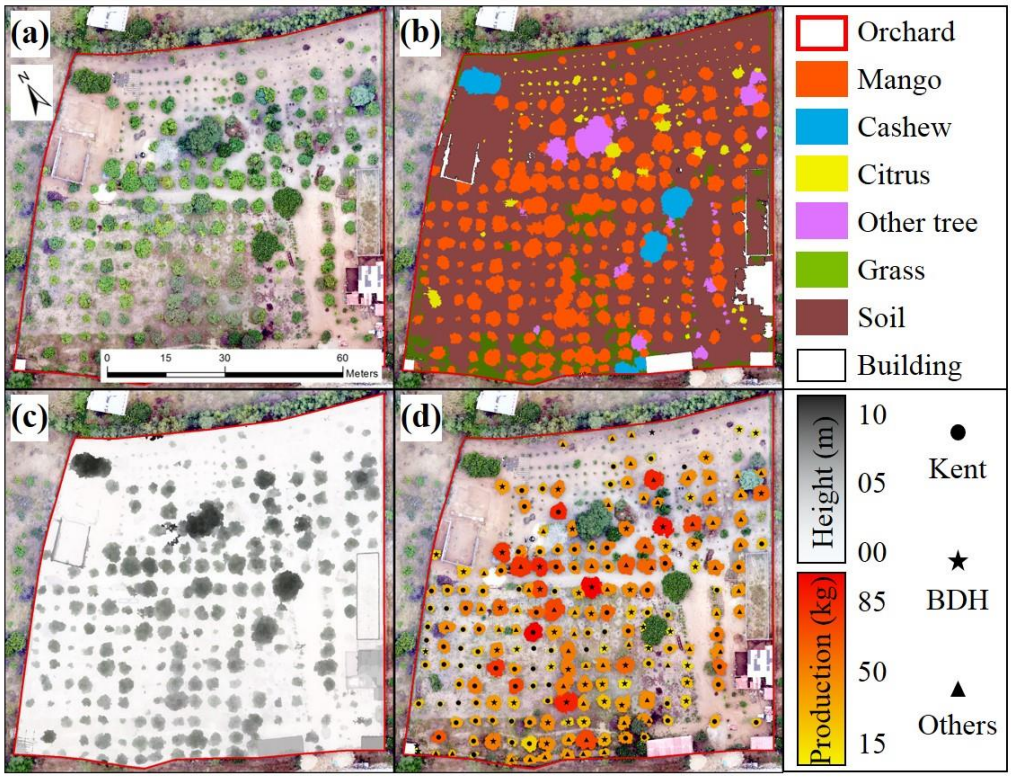


Fig. 61. Orchard yield mapping outputs. (a) UAV-acquired RGB orthomosaic; (b) GEOBIA land cover map; (c) Canopy Height Model (in meter); (d) mango cultivar yield map (in kg per tree). UTM coordinates system (zone 28N); Datum WGS84.

Because of variations in mango tree phenological stages between orchards and in environmental conditions during UAV flights, a RF classifier was trained for each classification in each orchard. Training datasets (one for each classification level) were built on manual sample objects selection thanks to the 150 calibration trees, expert photointerpretation, and knowledge of the orchard. For level 1 classification, the training dataset was class-balanced to represent classes' natural proportion, and ranged between 2.5 % and 7.3 % (Table 5) of the total number of objects in the orthomosaic (Belgiu & Drăguţ (2016) recommended a minimum of 0.25 % for high accuracy). For level 2 classification, the training dataset was made up of the ten calibration trees completed with few other mango trees for which the cultivar was known. Depending on orchard complexity, the number of classes used varied from 3 to 11 (Table 5).

Finally, the challenging task of tree crown segmentation in complex diversified systems was performed (Ke and Quackenbush 2011). Indeed, in our procedure, small trees were accurately delineated within one object, whereas large trees were over-segmented and contained several objects. After classification with the GEOBIA algorithm, a post-treatment was needed to delineate tree crowns, as most of them were over-segmented. For this purpose, eCognition "image-object-fusion" and "hierarchical classification" algorithms were applied iteratively to merge objects representing any tree crown objects that have at least 25% of shared border to another. The super-object was classified using the majority class of it sub-objects. The final land cover map obtained after the GEOBIA process is exemplified in Fig. 61.

For each orchard, classification performances (GEOBIA followed by tree crown delineation) were evaluated using 100 randomly-sampled ground control points (GCPs). The actual class of each GCP was manually checked by photointerpretation and compared to the estimated class. Confusion matrices were used to compute overall accuracy (Eq. 12) and the accuracy of the "mango" class (i.e., the fraction of known mango objects correctly classified as mango).

$$\text{Overall Accuracy (\%)} = 100 \times \frac{\text{Number of objects correctly classified}}{\text{Total number of objects}} \quad (12)$$

Table 6. Classes, spectral and texture features selected for each classification level.

Classification Level	Classes	Spectral Features Selected	Texture Features Selected
Level 1	soil building grass shrub annual crop citrus mimosae cashew mango other tree	Mean in blue, red and CHM[1]; standard deviation in blue, green and CHM [1]; brightness, hue and saturation; NDI [2], Excess Green and Excess Red	GLCM [3] homogeneity, entropy and standard deviation
Level 2 (mango cultivars)	'Kent' 'Keitt' 'BDH' other cultivar	Mean in CHM[1]and blue; standard deviation in blue and green; hue; Excess Green and Excess Red	GLCM [3] standard deviation

[1] Canopy Height Model. [2] Normalized difference index. [3] Haralick's gray level co-occurrence matrix.

Extraction of Individual Tree Structure Parameters Using GIS

We imported the land cover maps (.shapefile) and the CHM layers (.geotiff) in ArcGIS 10.3 (ESRI, Redlands, CA, USA). All orchard objects were visually checked, and the few tree crowns that were not correctly delineated during the GEOBIA procedure were manually reshaped. Then, we used the "Zonal statistics tool" of the "Spatial analyst extension" of ArcGIS to extract the maximum and sum of the pixel values of the CHM (expressed in meter) encompassed in the area of each mango tree object (Fig. 61c). Then, three structure parameters were computed for each individual mango tree: the crown area (in m^2) obtained directly from the land cover map, tree height (maximum value included in each tree object), and the crown volume (in m^3) computed as the sum of all CHM pixel values of the object multiplied by the CHM spatial resolution (i.e., 6.40 × 6.40 cm^2). For accuracy assessment of tree parameter estimation, the on-ground

height of 144 georeferenced trees was made in the field using a ruler. A linear regression was fitted between the measured and UAV-estimated tree height. We validated the stability of UAV tree height estimation by computing the root mean square error (RSME) and squared correlation coefficients (R^2). Relative RMSE (RMSE%) of validation was calculated as the percentage of the average measured tree height.

d. Model for Individual Tree Yield Estimation

Load Index

Mango tree production (i.e., kg of fruit per tree) relies on management practices (e.g., irrigation and fertilization regimes) (Crane et al. 2009) and environmental conditions (e.g., soil, climate) , which both vary in time and in space. Furthermore, mango trees display strong reproductive asynchronism and fruiting irregularity within and between the years due to several endogenous and exogenous factors (Dambreville et al. 2013). All these parameters make the task of estimating production from only cultivar and structure information quite challenging. Consequently, we introduced an integrative variable, named "load index", which encompasses management practices, sites, and years related effects. Usually, tree physiologists use fruit load, measured as a ratio between fruit number or biomass and leaves area, to analyze yield. However, fruit load is not easily assessable in the field. In this work, the load index was defined by a qualitative grading of tree bearing by visual inspection in the field: experts qualified the area of visible fruits compared to the visible overall crown area and ranked the tree into "low", "medium", and "high" load index classes. Load index was estimated on the 150 calibration trees simultaneously to image acquisition by at least two experts assisted by image abacus per cultivar (Figure S1 in Supplementary Materials).

Tree Production Models

Using the method described above, we characterized the cultivar and quantified structure parameters (tree height, crown area, and volume), load index, and production for the 150 calibration trees. The tree productions showed polynomial distributions against structure variables. We looked for the best predictive model that estimated tree production based on cultivar, load index, and tree structure. For each cultivar, an initial model relating production and the independent variables (i.e., load index factor and structure parameter co-variables) was fitted using a second-degree polynomial regression (squared values of co-variables included). From the initial model, the best predictive model was obtained by a backward elimination which iteratively deleted variables with the highest *p*-value (between 1 and 0.05) until only significant variables remained (i.e., p-value < 0.05). We fitted one model per cultivar, and an additional model was fitted for other local mango cultivars with the same method by using all trees of the three cultivars for the fittings.

These models have been validated using 60 validation trees on which the actual number of fruits was visually counted in the field by three experts using tally counters. Ground-measured production (in kg) was obtained from the conversion of the actual number of fruits using the average fruit weight (Table S1). The stability of production estimation was assessed by computing RMSE and RMSE% of validation. The significance of the differences between ground-measured and UAV-estimated tree production were tested using a Student *t*-test at a 95% confidence level.

e. Yield Mapping at the Orchard Scale

We applied the four production models to estimate mango yields at the orchard scale based on individual tree structure parameters and cultivar, spatially assessed following the workflow described above. In each orchard, the load index per cultivar was visually qualified simultaneously to UAV acquisition on 50 that were trees randomly sampled on a cross transect. The transect passed through the two orchard diagonals, with 25 trees sampled on each, ensuring the spatial distribution of the sampling. As orchard 15 was large (13.8 ha), two transects were performed, resulting in a total of 100 trees sampled. A "null" load index category was given to sampled trees carrying no fruit. Then, the weighted values $a, b, c,$ and d of "low", "medium", "high", and "null" load index categories, respectively, were computed for each cultivar. Finally, each tree production was averaged using the weighted values of load index as follows (Eq. 13):

$$P_{k/C} = aP_{k/C,low} + bP_{k/C,med} + cP_{k/C,high} + d \times 0 \quad (13)$$

where P_k is the estimated production of tree k for a given cultivar C (in kg); a, b, c, and d are the weighted values for production estimates in "low" ($P_{k/C,low}$), "medium" ($P_{k/C,med}$), "high" ($P_{k/C,high}$), and "null" (considered as 0 kg) respectively ($a + b + c + d = 1$). Estimated production was set to 0 when the predictive value given by the model was negative.

As a result, we obtained production for all mango trees in each orchard, and by summing it we calculated mango production (in tons) of the entire orchards. The orchard's productive area (in ha) was determined by delimiting cultivated areas excluding buildings on the land-cover maps. Because each tree was identified in the land cover maps, tree production has been mapped at the orchard scale in ArcGIS (Fig. 61d). We investigated for orchard yields assessed by the grower; and measured yield was opposed to the estimated yield. Additionally, the following orchard parameters were computed: estimated yield per hectare (in $t.ha^{-1}$), mango tree specific area (as a proportion of the total area), mango canopy volume (in $m^3.ha^{-1}$), and mango planting density (number of trees per hectare). Finally, based on the land cover maps, an assessment of the spatial composition was performed using the Simpson diversity index, which represents the probability that two pixels selected at random would be of different object classes (Lande 1996). Correlations between these features and orchard yields were studied using Pearson's test.

I.3. Results

a. Accuracy of Classification and Tree Structure Parameters Estimation

Land cover mapping accuracy of the 15 orchards is described in Table 7. The mean overall accuracy was of 0.89, and ranged between 0.73 and 0.99. Despite the fact that no effect of training rate on accuracy was evidenced (p-value > 0.05), an increase in the number of classes significantly decreased the accuracy (p-value < 0.005). Orchards 8, 10, and 13 had accuracies lower than 0.80. When focusing only on "mango" class accuracy, four orchards (3, 5, 7, 10 and 11) reached a score of 1 (Table 7). In these orchards, mango trees are clearly recognizable by their homogeneous crown size compared to other tree species. Only orchard 12 had a "mango" class accuracy lower than 0.80 due to a high level of confusion between mango and citrus trees.

Table 7. Classification accuracy results for the 15 studied orchards.

#	Overall Accuracy	« Mango » Accuracy
1	0.98	0.96
2	0.96	0.94
3	0.99	1.00
4	0.96	0.86
5	0.97	1.00
6	0.99	0.80
7	0.88	1.00
8	0.78	0.80
9	0.74	1.00
10	0.85	1.00
11	0.86	0.72
12	0.73	0.96
13–14	0.84	0.83
15	0.87	0.89
Mean	0.89	0.91

UAV accuracy for tree height estimation was evaluated by comparing measured and estimated heights (Fig. 62). For trees of 1.4 to 8.0 m in height, the slope of the regression line was 0.77. This suggests that CHM height values slightly underestimated the measured tree heights, especially for trees taller than 5 m. However, estimates were relatively stable (R^2 of 0.96 and RMSE% reaching 11%). We considered these results to be good enough to use the CHM values for tree height and crown volume estimates.

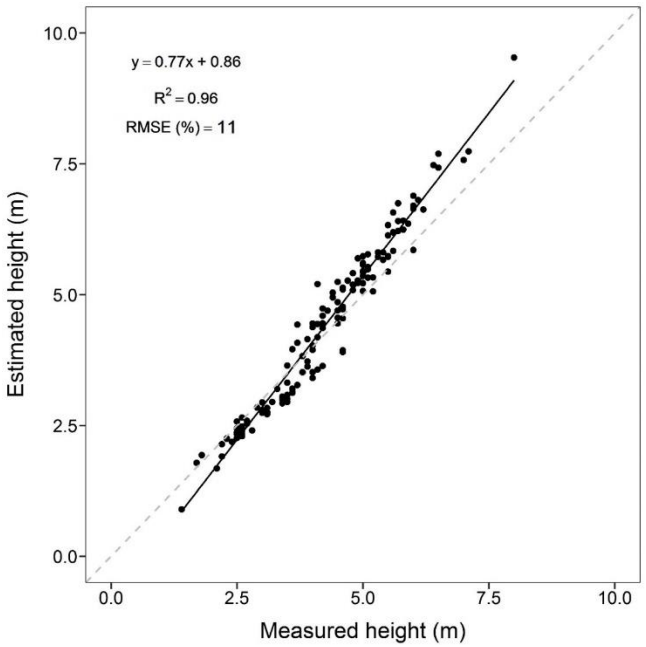


Fig. 62. Relationship between the on-ground measured (by ruler) and UAV-estimated (from Canopy Height Model) tree height values for 144 trees. The solid line is the fitted linear function and the dotted line is the 1:1 line.

b. Tree Production Models

Using a KNN-based machine vision system, we estimated the production of the 150 calibration trees. Furthermore, structure parameters (i.e., tree height, crown area and projected volume) of these trees were extracted from CHM layer. All data of each calibration tree are provided in Table S1. Comparisons of means of tree height, crown area, and volume between the calibration trees and the mango tree population reached *p*-values of 0.07, 0.35, and 0.55 respectively. As all *p*-values were greater than 0.05, the null *t*-test hypotheses that averages are equals were accepted, and calibration trees were assumed to be representative of mango heterogeneity. Mean production per calibration tree given by machine vision varied from 75 to 183 kg depending on cultivars (Table 8), and were significantly different following the cultivars (*p*-value < 0.001). These results strengthened our choice to fit different models for each cultivar instead of including a cultivar factor in a global model. The models produced appropriate R^2 values of 0.87, 0.79, 0.87, and 0.77 for 'Kent', 'Keitt', 'BDH', and other cultivars, respectively (Table 8). In each model, load index was significant, with the highest effect on production (lowest *p*-value < 0.001) compared with tree structure variables. Additionally, all the tree structure parameters were included in 'Kent' and 'Keitt' models, whereas both crown area and tree height had an effect on production for 'other cultivars', and only the crown area was significant in the 'BDH' model.

Models were validated on a set of 60 validation trees not used for fitting the models. The validation set was made of 20 'Kent', 20 'Keitt', and 20 'BDH', and their measured tree production was of 117.35, 164.94, and 155.46 kg on average, respectively (data not shown). Relationships between ground-measured production and production estimated by each model are displayed in Fig. 63, and show no significant difference of means (all *t*-test *p*-value > 0.40). The RMSE% of validation varied from 19.67% to 28.83% (Table 8 and Fig. 63). Although 'Keitt' and 'other cultivars' models showed no misestimation of production in general (Fig. 63b, d), 'Kent' and 'BDH' models seem to slightly under-estimate production for trees with high production (Fig. 63a,c).

Table 8. Tree production models per cultivars built with load index and tree structure parameters; and their respective performances. RMSE and RMSE% expressed in kg and % respectively were computed on 60 validation trees.

Cultivar	**Selected Model** [1]	**Mean Production (kg)**	**R** [2]	**RMSE**	
				kg	**%**
'Kent'	$P \sim LI + Area + Area^2 + Vol^2$	75.1	0.87	30.2	25.7
'Keitt'	$P \sim LI + Height + Area + Vol$	182.6	0.79	32.4	19.7
'BDH'	$P \sim LI + Area$	110.8	0.87	38.3	24.6
'Others'	$P \sim LI + Height + Height^2 + Area + Area^2$	104.7	0.77	42.1	28.9

[1] *P*: tree production (kg); *LI*: load index; *Height*, *Area* and *Vol*: tree height (m), crown area (m^2) and crown volume (m^3), respectively.

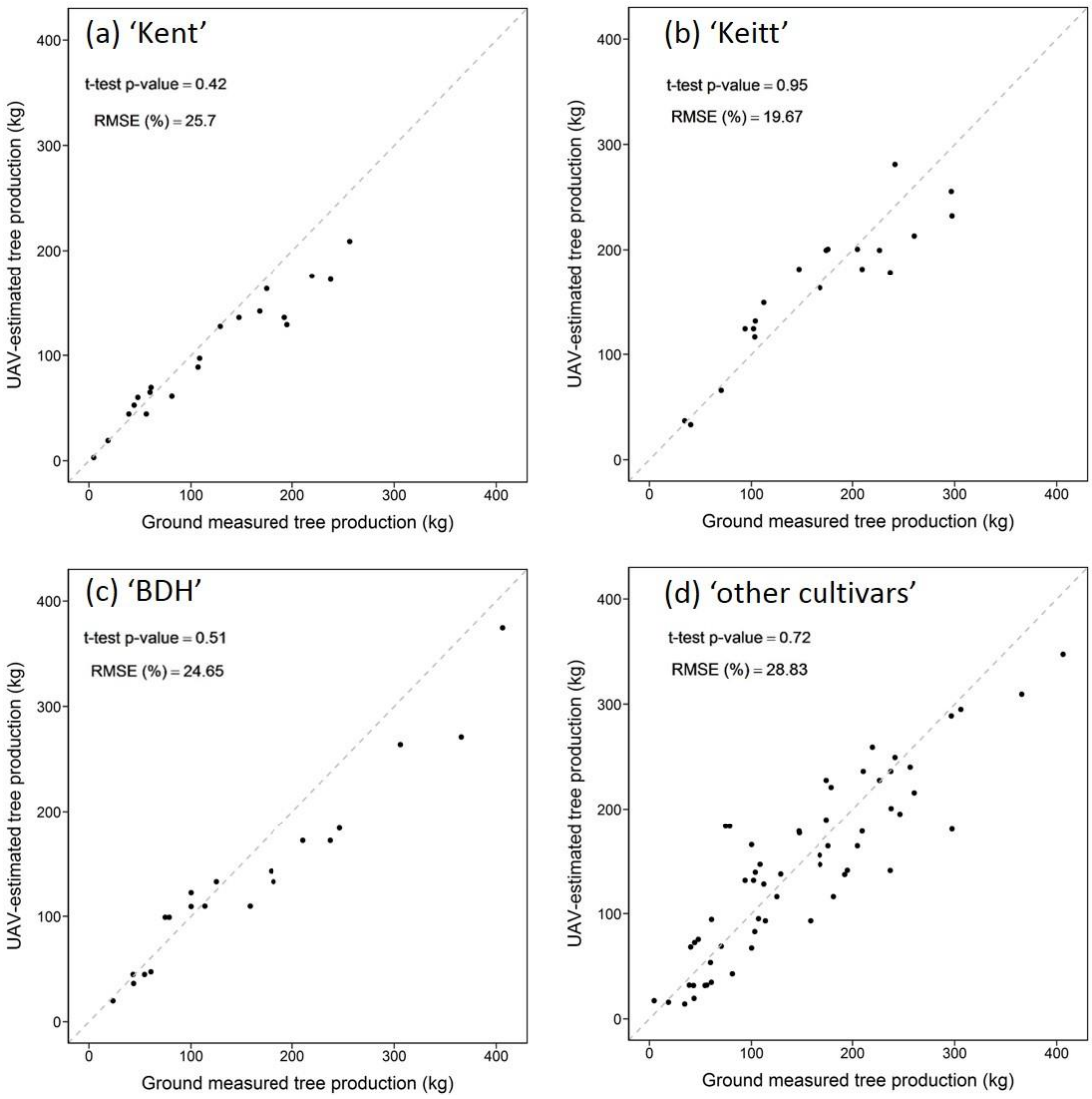


Fig. 63. Relationship between the ground-measured tree production and UAV-estimated tree production (**a**) for 'Kent' trees, (**b**) for 'Keitt' trees, (**c**) for 'BDH' trees, and (**d**) for 'other cultivars' on 60 validation trees. Relative RMSE (RMSE%) of validation and *p*-value of the *t*-test are displayed on each panel. The dotted line is the 1:1 line.

c. Yield Estimations

Tree Structures and Productions

In the 15 studied orchards, a total of 5400 mango trees have been identified (cultivars) and described (structures) following the method described above (Fig. 64a). The production was estimated on these 5400 trees by applying the models described previously (I.3.b) and using Eq. 13. The number of mango trees per cultivar per orchard was significantly different between orchards (Fig. 64a), and the frequency of load index categories depended on the orchard (Chi-squared, p-value < 0.001) (Fig. 64c). Some orchards (1, 5, 10, 12, 13) had high proportions of trees with 'high' and 'medium' load indices, while others (6, 8, 9, 11) had high proportions of 'low' and 'null' indices. Additionally, orchards displayed significant discrepancies in tree structures (Chi-squared, p-value < 0.001 for all tree structure parameters). Some orchards had an average

tree height below three meters (3, 4, 8), and in fact, corresponded to young plantations, while others (10 and 15) displayed a high variability in tree structure (tree heights ranging from 0.5 m to 16.35 m) (Fig. 64b).

As a consequence of orchard variations in load index and tree structures, mean tree production varied from 14.6 kg per tree in orchard 3 to 296.5 kg per tree in orchard 10. Comparison of these two extreme orchards showed that they have the same load indices proportion (Fig. 64c), but that trees have higher structural parameters in orchard 10 than in orchard 3 (Fig. 64b).

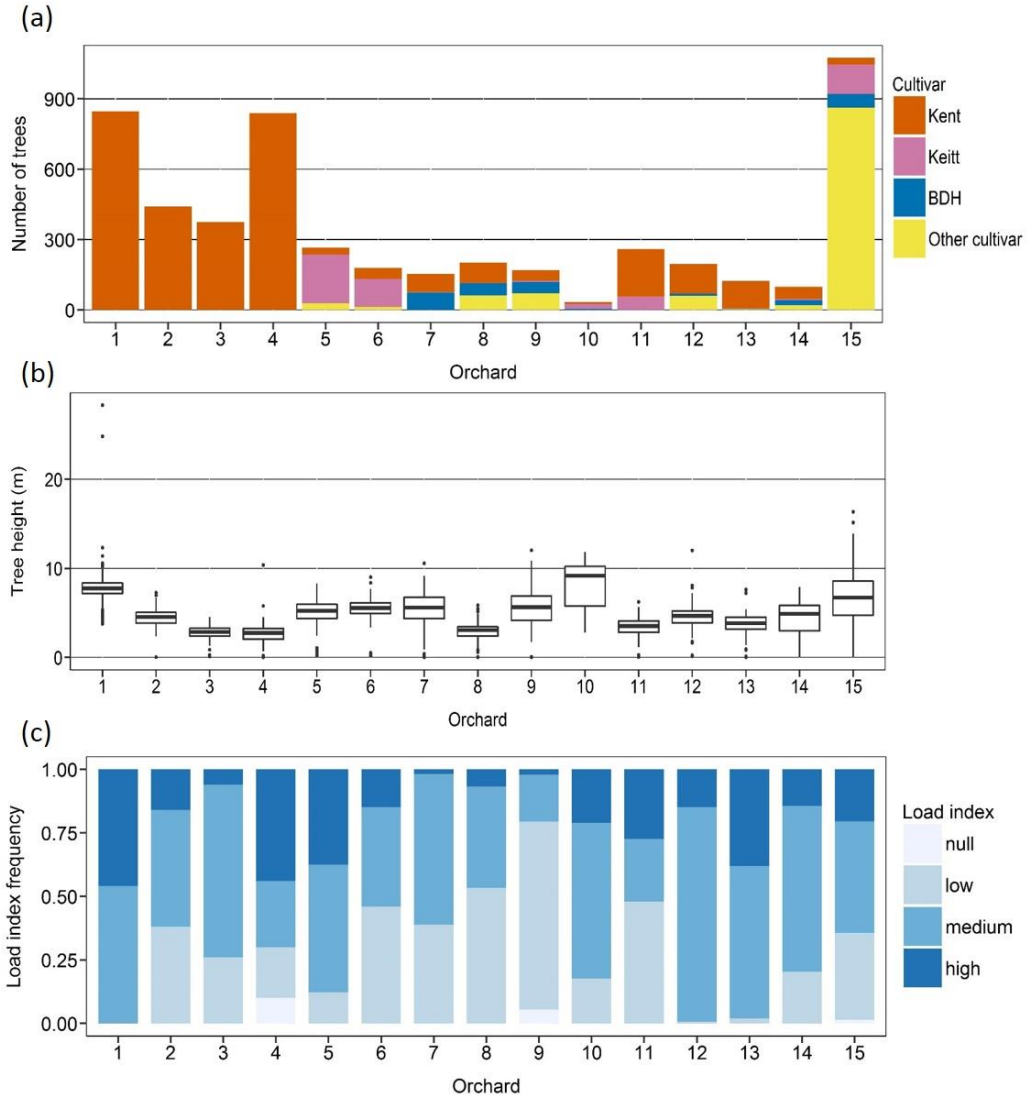


Fig. 64. Orchards description: (**a**) number of trees per cultivar; (**b**) tree height (m); (**c**) and frequency of load index.

Yield Mapping at the Orchard Scale

Orchard yield was calculated by summing all mango tree estimated productions per productive area (orchard cultivated area in hectare) for each orchard. Those results have been compared to measured yields by the growers in six orchards (data were available). Estimated yields varied from 2.5 to 39.6 $t.ha^{-1}$ depending on orchards. Estimated yields were consistent with growers' information for orchards 1, 10, and, 11 (Table 9). However, estimations were not aligned with surveyed yields for orchards 2, 3, and 4 (where data reliability

was medium or low). Thanks to the GIS software, we mapped individual tree productions in order to visualize orchard yield heterogeneity (see Fig. 61d for an example). Land cover maps also allowed us to compute some orchard spatial features related to mango trees: planting density (number of mango tree per hectare), mango specific area (% of area covered by mango), mango canopy volume, and Simpson index (object class diversity). These results are presented in Table 9.

We found that orchard yield was significantly and positively correlated to planting density, mango specific area, and mango canopy volume with Pearson's correlation coefficients ranging between 0.63 and 0.76 (*p*-values < 0.05). Cultivar proportion was also found to be an important driver of orchard yields since the four cultivars have significant differences on yield (Table 8). No significant relationship was found between orchard landscape diversity (Simpson index) and yields. Generally, we found that orchard yields were positively linked with planting density and mango canopy volume (correlated to mango specific area), though it was not the case for some orchards, such as orchards 9 and 15. Despite high mango canopy volume in these orchards (more than 20,000 $m^3.ha^{-1}$), their yields were below 5 t/ha^{-1}.

Table 9. Orchard spatial features and yields.

#	Area [ha]	Planting Density [trees.ha^{-1}]	Mango Specific Area [%]	Mango Volume [m^3.ha^{-1}]	Simpson Index	Orchard Yield [t.ha^{-1}]	
						Estimated	Grower
1	2.2	386	49.7	62,792	0.51	39,619	41,079 ***
2	2.1	214	37.9	25,878	0.47	14,627	6923 *
3	2.8	138	6.0	3043	0.13	2012	3667 **
4	2.2	380	13.6	5012	0.27	6737	1060 *
5	3.7	72	14.9	19,513	0.25	8518	NA
6	1.7	107	19.4	12,131	0.35	8661	NA
7	1.1	169	30.6	15,511	0.70	10,400	NA
8	1.1	205	15.1	6019	0.56	5908	NA
9	2.2	86	23.1	25,624	0.73	3908	NA
10	1.3	25	11.2	8383	0.53	7469	7634 ***
11	1.5	172	20.0	7907	0.56	11,193	10,526 ***
12	1.2	170	30.2	13,073	0.67	11,694	NA
13	0.9	148	16.2	4539	0.66	7867	NA
14	0.9	113	21.8	8924	0.47	6447	NA
15	13.8	80	29.0	227,040	0.56	2346	NA

* level of reliability of the data (* low, ** medium, *** high).

I.4. Discussion

a. Computational Time

As in (Torres-Sánchez et al. 2015), computational time is a recognized limit of our work. In the following paragraph, we present the time needed for each step of the presented method when applied on one average-

sized orchard (around 2 ha). The computer used for calculation was a Dell Precision M4700 (Dell Inc., Round Rock, TX, USA) with a 16 GB of RAM, an Intel Core i7 processor and a NVIDIA Quadro K1000M graphic card of 2 GB. The first steps of orthomosaic and DSM building took 2 h 50. Duration for these steps mainly depended on the covered area and number of images to proceed. Second, most of the processing time was spent on GEOBIA classification, which took from 4 h 30 to 8 h 10, depending on orchard complexity and whether or not the level 2 classification was applied. Time processing for GEOBIA classification could be summarized as follow: from 1 h to 2 h for both segmentation; from 45 min to 1 h 30 for manual sampling; 20 min to 1 h 00 for RF training and from 1 h 20 to 4 h 40 for RF application for both classifications. When used, mango cultivar classification added 25 min to 1 h 40 to the processing time. Finally, RF application took some 20% of 50% of the entire processing time, mostly because of the large number of objects to classify (between 38,000 and 580,000). To summarize, an average processing time of 15 h per orchard was needed to apply the presented methodology. Some steps of the procedure required manual operations, and improvement of these steps is discussed in the next paragraph.

b. Challenges of GEOBIA in Heterogeneous Agroecological Landscapes

To map land cover in each orchard, we used a standard GEOBIA classification process on four-band images (R, G, B, and CHM) acquired by the UAV. Due to their high flexibility, low cost, and the VHR images, UAVs are widely used for the purpose of land cover mapping (Lisein et al. 2015; Michez et al. 2016; Nevalainen et al. 2017) and tree crown delineation (Díaz-Varela et al. 2015; Torres-Sánchez et al. 2015; Jiménez-Brenes et al. 2017; Birdal et al. 2017). Each of these two tasks requires its own method and algorithm. GEOBIA procedures are recognized to be more accurate than per-pixel classification on complex landscapes (Blaschke 2010). In this work, segmentations allowed us to obtain thousands of objects which were then classified using successively two Random Forest (RF) classifications: a first (level 1) to discriminate mango trees from other tree species or objects (building, grass, etc.); and a second (level 2) applied on mango tree objects to determine their cultivar. These two classifications might be time consuming, as they increase the number of steps. However, as shown by our results, the accuracy of the land cover maps is enhanced by them. Moreover, the spectral and textural similarities between mango and citrus trees and among mango cultivars constrained us to first isolate mango trees and then classify mango cultivars. Such a hierarchical approach significantly improved RF classification accuracy, as shown in previous works (Lebourgeois et al. 2017).

In our study, segment classifications into 10 classes including four tree species (“mango”, “citrus”, “cashew” and “mimosa”) reached an average overall accuracy of 89% (Table 7). These results confirmed the high efficiency of RF for tree species classification, and were comparable to recent studies on tree species classification in forests. For instance, Nevalainen et al. (2017) obtained an accuracy of 95% for the classification of four tree species in a boreal forest with RF on hyperspectral imagery. Lisein et al. (2015) achieved a classification error of 16% to classify five tree species by using RGB combined with a near infrared (NIR) band. Our results have shown that accuracy decreased when orchard landscapes got more complex (high number of similar classes) and when diversity increased (Simpson index), especially in cases where there were several mango cultivars to discriminate among (Table 7). Visible discrepancies between mango cultivars made level 2 classification possible. However, for some orthomosaics, mango cultivars were not clearly distinct because of flight conditions (high level of shadow or wind) and tree physiological status. To offset this issue, other spectral bands could be added by switching the RGB camera for NIR acquisition or using hyperspectral sensors, as recommended for tree species classification (Michez et al.

2016; Fassnacht et al. 2016; Maschler et al. 2018). In future works, such solutions should be examined, considering that acquisition cost and processing time will increase (Torresan et al. 2017).

Even if feature selection is not imperative to improving RF classification accuracy (Belgiu and Drăguţ 2016; Ma et al. 2017), we assume that such a method would decrease computation time by keeping only relevant features. In this work, a decision tree (implemented under eCogition) was applied to rank and select the most important features (Table 6). However, other studies have used more advanced methods, and Ma et al. (2017) have compared several methods to determine the most adapted features for RF classifier.

The GEOBIA procedure used in this study was also limited by its semi-automation, mainly during the building of (1) the training dataset and (2) post-processing for tree crown delineation steps. Firstly, because UAV images were acquired under different environmental conditions (daytime, luminosity, wind speed, etc.), tree aspect was highly variable between orchards, justifying the use of at least one training dataset per orchard (i.e., 25 training datasets built in total: 15 orchards x 1 or 2 classifications). Each training dataset, built by manual sampling of objects in each class, was time consuming and might have led to class imbalance impacting RF classification accuracy (Ma et al. 2017). An option could be to transfer a trained RF classifier to another study area, even if the efficiency of such technique remains unclear (Juel et al. 2015). Second, due to the high variability of tree structure and plantation patterns within and between orchards, crown delineation was adapted for each situation. Consequently, tree crown delineation ended up being a limiting factor, because for some complex orchards, a step-by-step merging method was used to fuse objects belonging to the same tree crown, and a few crown trees embedded in the same object were manually divided and reshaped. One could think that increasing the scale parameters in multiresolution segmentation or using another segmentation method such as mean-shift segmentation could better delineate trees (Maschler et al. 2018). However, we deliberately chose to over-segment trees, as classification showed better results, (Ma et al. 2015) and in order to avoid the non-segmentation of small trees. The delineation of individual trees remains a complex process, especially under heterogeneous, agroecological landscapes, and there are plenty of automatic or semi-automatic algorithms (Ke and Quackenbush 2011) for which efficiency has to be tested in this context. In addition, testing and evaluating new tree crown delineation methods should be completed.

c. Tree Structure Assessment

The crown volume and area of each tree were computed from Canopy Height Model layer obtained from UAV structure-from-motion process (Lisein et al. 2015). In our study, the estimated tree height was stable and robust, as RMSE was equal to 0.47 m (RMSE% = 11%), which was comparable to similar works using RGB cameras. For example, Torres-Sánchez et al. (2015) reported errors in tree height assessment ranging from 0.22 to 0.53 m, while Díaz-Varela et al. (2015) obtained RMSE% from 6% to 20%. Stability in tree height estimations confirmed that UAV-based tree structure parameters can be used as inputs for modelling. However our estimations of tree heights from CHM slightly underestimated measured tree heights (regression slope = 0.77). If exact tree heights were needed, one could calibrate CHM values using linear regressions fitted in Fig. 62. Nevertheless, after the tree structure extraction, we noticed that tree height values were not consistent for only 2.1% of our trees (low or zero value). These inconsistencies resulted from errors during point cloud densification at DSM construction step. After taking a closer look, we noticed that most of these errors were coming from one orchard (1), which has the highest canopy cover (mango volume > 60 000 $m^3.ha^{-1}$) and tree density (386 $trees.ha^{-1}$). Interestingly, previous studies showed that the point cloud generated by structure-from-motion photogrammetry failed to measure the DSM in areas with dense canopy cover (Wallace et al. 2016). Under such conditions, one should use another source of terrain

information such as LiDAR (Wallace et al. 2016). Another solution could be to adapt the UAV's flight plan in order to enhance 3D triangulation (flight height, grid flight, or viewing angle (Zhang and Kovacs 2012). Finally, a manual corrective procedure of the point cloud for each misestimated tree might also enhance the accuracy of tree structure assessment.

d. Yield Estimation at the Orchard Scale

The tols described in this study allowed us to rapidly and precisely map mango yields at the orchard scale across the diversity of mango cropping systems. Quantification of tree species, mango cultivars, estimates of their structure, and load index were implemented as independent variables in models for tree production estimations. These models were fitted to 150 calibration trees depicting the tree structure and fruit load of the three main cultivars ('Kent', 'Keitt' and 'BDH') found in Niayes area. When evaluated with measured production on 60 trees, models reached satisfying accuracies, with R^2 greater than 0.77 and RMSE% of validation ranging from 19.67% to 28.83%. Few studies used above-ground remote sensing to estimate tree production. For instance, Rahman et al. (2018) used multispectral satellite imagery on three mango orchards to compute vegetation indices and tree crown area, and linked them to tree production across two years. While their models developed at the individual orchard level for both years reached high efficiencies ($R^2 > 0.79$), weaker correlations were obtained by combining data in all orchards ($R^2 = 0.70$), or by using a single year ($R^2 < 0.57$). Consequently, the correlation between tree production and remotely-assessed features was not generic, and has to be calibrated for each orchard and each year to include climate and site effects. In some previous works, mango yield estimation and, for some, mapping, were achieved by a direct assessment of production of each tree of the orchard by ground machine vision systems composed of a mobile platform (a car or a robot) equipped with GPS and multiple sensors. These studies reach high efficiency for individual tree production estimation, providing accurate yield map of the orchard when performed on all the trees. For example, Payne et al. (2014) detected mango fruits by RGB images automatically taken from a car driving down the plantation. The relationship between detected mangoes and the actual tree load was 0.92. The most advanced system was recently proposed by Stein et al. (2016), who used a mobile robot equipped with RGB and LiDAR sensors. This system provided a 3D map of the tree structure and estimated production with an error rate of only 1.36% compared to validation trees. However, such expensive and time-consuming tools can hardly be employed in complex, diversified orchards and on large areas, which currently represent the majority of the mango cropping systems worldwide, and especially those in developing countries (Rey et al. 2007; Grechi et al. 2013).

Here, the proposed method achieved the breakthrough of rapidly and precisely mapping mango production without above ground imagery, but rather by linking productions with tree structural parameters. However, the relationship between production and tree crown structure, even measured with the highest precision of a LiDAR, is known to be poor for mango trees, as noticed by Stein et al. (2016). This poor relationship also results from endogenous physiological factors of the tree, as exemplified by Dambreville et al. (2013). By introducing the load index as a proxy of fruit load that is easily measurable at the orchard scale, we solved this issue (Rahman et al. 2018). We acknowledged that this variable encompassed various effects such as climate, sites, and management practices that could not be estimated in our procedure. Further studies should test the robustness of the tree production models under various conditions (different years and other study areas). As load index is summarized for each cultivar from a transect of 50 trees at the orchard scale, we assumed that tree production was estimated by weighting this variable over the orchard (Eq. 13). We assumed that 50 sampled trees were enough to capture the variability within small orchards (less than 2 ha,

the majority in the study area). Moreover, in the study area, large orchards are conventional monospecific systems in which tree structure and production are homogeneous, resulting in a lower sampling intensity needed to capture the orchard variability. Under these conditions, we believe that our sampling methods provide balanced sampling intensity among orchards. Despite this simplification, it was possible to accurately and rapidly map and estimate yield at the orchard scale (Fig. 61d). Future works should investigate the effect of the transect method on the accuracy of orchard yield estimation. Additionally, easier sampling and geolocalization of the load index should be explored: this could be done by a visual, ground-based assessment of a given proportion of the trees, or by UAV-based imagery.

Finally, over the 15 studied orchards, production was estimated on 5400 mango trees using the developed models and empirically assessing the load index at the orchard scale. Yields at the orchard scale resulted from the sum of the production of each mango tree of the orchard. Comparisons with growers' information was possible only for 6 orchards, and our estimation was highly accurate for three of them (which had the highest level of reliability). For the other three, differences might be explained by the fact that growers measured their production on plots (>10 ha) larger than our studied area. In these cases, yield measurements by growers are coarse, and do not take into account in-field variations (planting density, tree structure, microclimate, watering heterogeneity, etc.). For instance, the grower of orchard 4 measured production on a plot of 30 ha, but admits that yield is closer to 8 $t.ha^{-1}$ in the 2.21 ha area we studied. Nine out of fifteen orchard growers had no estimation of their actual yields that represent the main case in the studied area, but also, in most of the orchards of developing countries (Carletto et al. 2015). At best, a manual count of a low number of trees is performed to obtain an average number of fruits per tree that is multiplied by the number of trees in the orchard; an estimate that is inaccurate and labor-intensive, and more importantly, that does not inform the grower about the yield distribution within the orchard. To our knowledge, our methodology is the first that allows estimating the orchard yield at tree scale precision by using consumer-grade tools of remote sensing (RGB camera and UAV). We consequently provided a tool for easily quantifying yield heterogeneity and distribution within the orchard; a tool that will inform growers on practices to address the needs of each tree (pruning, watering, etc.) and to plan orchard management (harvest, packing, etc.).

I.5. Conclusions

In this paper, we introduced a new and innovative methodology for yield estimation and mapping at the orchard scale. A set of tools for characterizing and quantifying the structure of mango trees was set up in order to map yields at the orchard scale. The described procedure permitted us to obtain two precise outputs from UAV imagery analysis: (i) a land cover map, and (ii) a Canopy Height Model (CHM). From the latter, three tree structure parameters were computed: tree height, crown area, and crown volume. These variables, combined with a load index, were implemented as independent variables in models for accurate production estimations of each individual tree in the orchard. The procedure is used for yield mapping at the orchard scale. Further works on the GEOBIA classification will make it possible to improve procedure automaticity. By using land cover maps and CHM, other variables describing orchard composition and configuration are computable to explain yield variations. To our knowledge, ours is the first methodology using low-cost tools (RGB camera and UAV) that achieves mapping of tree structures, species, and yield in orchards. This tool addresses the issue of accurate estimates of mango orchard yields, which remains a challenging task, especially in developing countries.

Supplementary Materials: The following are available online at www.mdpi.com/link, Figure S1: Image abacus used by expert in the field to estimate load index for (a) 'Kent', (b) 'Keitt', (c) and 'BDH' cultivar. Load index categories (low, medium and high) are displayed in column and different tree heights (small, medium and tall) are represented in line. Table S1: Mean fruit weight and standard deviation (SD) for the three variety in Niayes region. Table S2: Description of the 150 calibration trees: cultivar; number of fruit detected by the KNN-based machine vision and yield measured; load index; and tree structure parameters (tree height, crown area and volume).

Acknowledgments: This work was supported by the French National Research Agency under the Investments for the Future Program (ANR-16-CONV-0004) and by the PixYield Creativity and Scientific Innovation action funded by CIRAD, the French agricultural research and international cooperation organization. We received funds for covering the costs to publish in open access. Additionally, we are grateful to Pierre-Éric Lauri and Frédéric Normand for insightful comments on an earlier version of this work. The authors thank three anonymous reviewers for their valuable comments and constructive suggestions on the manuscript.

Author Contributions: J.S., É.M., and É.F. conceived and designed the experiments; J.S., C.A.B.S., and É.F. performed the experiments; J.S. and É.F. analyzed the data; All authors contributed to the data interpretations and participated in the edition of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

Baatz M, Schäpe A Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl J, Blaschke T, Griesbner G (eds) Angewandte Geographische Informations-Verarbeitung, XII. Wichmann Verlag, Karlsruhe, Germany, pp 12–23

Bargoti S, Underwood J (2017) Deep fruit detection in orchards. In: Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, pp 3626–3633

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Birdal AC, Avdan U, Türk T (2017) Estimating tree heights with images from an unmanned aerial vehicle. Geomat Nat Hazards Risk 8:1144–1156. https://doi.org/10.1080/19475705.2017.1300608

Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65:2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Carletto C, Jolliffe D, Banerjee R (2015) From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. J Dev Stud 51:133–148. https://doi.org/10.1080/00220388.2014.968140

Carletto C, Savastano S, Zezza A (2013) Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. J Dev Econ 103:254–261. https://doi.org/10.1016/j.jdeveco.2013.03.004

Crane JH, Salazar-Garcia S, Lin T-S, et al (2009) Crop Production: Management. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 432–483

Dambreville A, Lauri P-E, Trottier C, et al (2013) Deciphering structural and temporal interplays during the architectural development of mango trees. J Exp Bot 64:2467–2480. https://doi.org/10.1093/jxb/ert105

Díaz-Varela R, de la Rosa R, León L, Zarco-Tejada P (2015) High-Resolution Airborne UAV Imagery to Assess Olive Tree Crown Parameters Using 3D Photo Reconstruction: Application in Breeding Trials. Remote Sens 7:4213–4232. https://doi.org/10.3390/rs70404213

Fare Y, Dufumier M, Loloum M, et al (2017) Analysis and Diagnosis of the Agrarian System in the Niayes Region, Northwest Senegal (West Africa). Agriculture 7:59. https://doi.org/10.3390/agriculture7070059

Fassnacht FE, Latifi H, Stereńczak K, et al (2016) Review of studies on tree species classification from remotely sensed data. Remote Sens Environ 186:64–87. https://doi.org/10.1016/j.rse.2016.08.013

Faye E, Rebaudo F, Yánez-Cajo D, et al (2016) A toolbox for studying thermal heterogeneity across spatial scales: from unmanned aerial vehicle imagery to landscape metrics. Methods Ecol Evol 7:437–446. https://doi.org/10.1111/2041-210X.12488

Giannetti F, Chirici G, Gobakken T, et al (2018) A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. Remote Sens Environ 213:195–205. https://doi.org/10.1016/j.rse.2018.05.016

Gongal A, Amatya S, Karkee M, et al (2015) Sensors and systems for fruit detection and localization: A review. Comput Electron Agric 116:8–19. https://doi.org/10.1016/j.compag.2015.05.021

Grechi I, Sane CAB, Diame L, et al (2013) Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68:447–466. https://doi.org/10.1051/fruits/2013094

Haralick RM, Shanmugam K, Dinstein I (1973) Textural Features for Image Classification. IEEE Trans Syst Man Cybern SMC-3:610–621. https://doi.org/10.1109/TSMC.1973.4309314

Jiménez-Brenes FM, López-Granados F, de Castro AI, et al (2017) Quantifying pruning impacts on olive tree architecture and annual canopy growth by using UAV-based 3D modelling. Plant Methods 13:55. https://doi.org/10.1186/s13007-017-0205-3

Juel A, Groom GB, Svenning J-C, Ejrnæs R (2015) Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. Int J Appl Earth Obs Geoinformation 42:106–114. https://doi.org/10.1016/j.jag.2015.05.008

Ke Y, Quackenbush LJ (2011) A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. Int J Remote Sens 32:4725–4747. https://doi.org/10.1080/01431161.2010.494184

Lande R (1996) Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. Oikos 76:5. https://doi.org/10.2307/3545743

Lebourgeois V, Dupuy S, Vintrou É, et al (2017) A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). Remote Sens 9:259. https://doi.org/10.3390/rs9030259

Lisein J, Michez A, Claessens H, Lejeune P (2015) Discrimination of Deciduous Tree Species from Time Series of Unmanned Aerial System Imagery. PLOS ONE 10:e0141006. https://doi.org/10.1371/journal.pone.0141006

Ma L, Cheng L, Li M, et al (2015) Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. ISPRS J Photogramm Remote Sens 102:14–27. https://doi.org/10.1016/j.isprsjprs.2014.12.026

Ma L, Fu T, Blaschke T, et al (2017) Evaluation of Feature Selection Methods for Object-Based Land Cover Mapping of Unmanned Aerial Vehicle Imagery Using Random Forest and Support Vector Machine Classifiers. ISPRS Int J Geo-Inf 6:51. https://doi.org/10.3390/ijgi6020051

Maschler J, Atzberger C, Immitzer M (2018) Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data. Remote Sens 10:1218. https://doi.org/10.3390/rs10081218

Meyer GE, Neto JC (2008) Verification of color vegetation indices for automated crop imaging applications. Comput Electron Agric 63:282–293. https://doi.org/10.1016/j.compag.2008.03.009

Michez A, Piégay H, Lisein J, et al (2016) Classification of riparian forest species and health condition using multi-temporal and hyperspatial imagery from unmanned aerial system. Environ Monit Assess 188:146. https://doi.org/10.1007/s10661-015-4996-2

Mukherjee SK, Litz RE (2009) Botany and Importance. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 1–18, ISBN 978-1-84593-489-7

Nevalainen O, Honkavaara E, Tuominen S, et al (2017) Individual Tree Detection and Classification with UAV-Based Photogrammetric Point Clouds and Hyperspectral Imaging. Remote Sens 9:185. https://doi.org/10.3390/rs9030185

Normand F, Lauri P-E, Legave J-M (2015) Climate change and its probable effects on mango production and cultivation. Acta Hortic 21–31. https://doi.org/10.17660/ActaHortic.2015.1075.1

Okamoto H, Lee WS (2009) Green citrus detection using hyperspectral imaging. Comput Electron Agric 66:201–208. https://doi.org/10.1016/j.compag.2009.02.004

Payne A, Walsh K, Subedi P, Jarvis D (2014) Estimating mango crop yield using image analysis using fruit at 'stone hardening' stage and night time imaging. Comput Electron Agric 100:160–167. https://doi.org/10.1016/j.compag.2013.11.011

Payne AB, Walsh KB, Subedi PP, Jarvis D (2013) Estimation of mango crop yield using image analysis – Segmentation method. Comput Electron Agric 91:57–64. https://doi.org/10.1016/j.compag.2012.11.009

Qureshi WS, Satoh S, Dailey MN, Ekpanyapong M (2014) Dense segmentation of textured fruits in video sequences. In: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on. IEEE, pp 441–447

Rahman M, Robson A, Bristow M (2018) Exploring the Potential of High Resolution WorldView-3 Imagery for Estimating Yield of Mango. Remote Sens 10:1866. https://doi.org/10.3390/rs10121866

Rey J-Y, Diallo TM, Vannière H, et al (2007) The mango in French-speaking West Africa: varieties and varietal composition of the orchards. Fruits 62:57–73. https://doi.org/10.1051/fruits:2006051

Sarron J, Sané CAB, Borianne P, et al (2020) Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions? Acta Hortic 8

Stein M, Bargoti S, Underwood J (2016) Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors 16:1915. https://doi.org/10.3390/s16111915

Torresan C, Berton A, Carotenuto F, et al (2017) Forestry applications of UAVs in Europe: a review. Int J Remote Sens 38:2427–2447. https://doi.org/10.1080/01431161.2016.1252477

Torres-Sánchez J, López-Granados F, Serrano N, et al (2015) High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. PLOS ONE 10:e0130479. https://doi.org/10.1371/journal.pone.0130479

Underwood JP, Hung C, Whelan B, Sukkarieh S (2016) Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. Comput Electron Agric 130:83–96. https://doi.org/10.1016/j.compag.2016.09.014

van Ittersum MK, van Bussel LGJ, Wolf J, et al (2016) Can sub-Saharan Africa feed itself? Proc Natl Acad Sci 201610359. https://doi.org/10.1073/pnas.1610359113

van Wart J, Kersebaum KC, Peng S, et al (2013) Estimating crop yield potential at regional to national scales. Field Crops Res 143:34–43. https://doi.org/10.1016/j.fcr.2012.11.018

Wallace L, Lucieer A, Malenovský Z, et al (2016) Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds. 16

Zhang C, Kovacs JM (2012) The application of small unmanned aerial systems for precision agriculture: a review. Precis Agric 13:693–712. https://doi.org/10.1007/s11119-012-9274-5

II. Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa

La méthodologie décrite précédemment et permettant la modélisation de la production des arbres par l'outil drone est une manière efficace d'obtenir et de spatialiser le rendement du verger (Chapitre 2 – Partie I). Cependant cette méthode fait face à plusieurs limites liées au temps de traitement et à la difficulté pour les producteurs d'accéder à des drones. Cette Partie II est un article décrivant la spatialisation du rendement dans trois vergers représentatifs des systèmes de culture (extensif, diversifié et intensif) en utilisant l'outil d'estimation de la production à l'échelle de l'arbre (voir Chapitre 1 – Partie II). La spatialisation de la production permet ensuite de décrire les hétérogénéités spatiales du rendement dans chaque verger. Ces hétérogénéités spatiales sont ensuite prises en compte pour mettre en place des stratégies d'échantillonnage adaptées. Ces stratégies d'échantillonnage constituent la seconde méthode pour l'estimation du rendement et qui n'utilise pas l'imagerie drone.

Cet article scientifique est en cours de finalisation et doit être soumis à une revue à comité de lecture (revue visée *European Journal of Agronomy*).

- Sarron J.; Faye É.; Malézieux É. **Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa.**

Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa

Julien Sarron[1,2,3,*], Emile Faye[1,2,3], Eric Malézieux[1,2]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]HortSys, University Montpellier, CIRAD, F-34090 Montpellier, France
[3]Centre pour le Développement de l'Horticulture, ISRA, Dakar 14000, Senegal
*Corresponding author: julien.sarron@cirad.fr

Abstract

Production of fruit tree within an orchard depends on both internal (structure, age, cultivar) and external (soil, topography, climate, and interactions with other trees) factors. At the upper scale, orchard yield relies on both individual tree production (and its temporal variation) and orchard structure (i.e., planting density, species diversity, and spatial arrangement). Estimating the orchard yield prior to harvest is a critical step that requires to consider the spatial heterogeneity of tree production, especially in complex and smallholder orchards (e.g., agroforestry systems). However, most studies focus on monospecific orchards. The aim of this study is to characterize the within-field variability of different types of orchards using specific methodologies in order to accurately estimate their yields. Our study focused on three different mango-based orchards located in Senegal, West-Africa, representative of extensive, diversified, and intensive cropping systems. Using unmanned aerial vehicle and ground imageries, the structure and production of all the trees were accurately estimated and mapped. Hence, the spatial variability of planting pattern and tree features (i.e., crown volume, production, and fruit load) were quantified for each orchard. The planting pattern of the diversified orchard was locally regular, then clustered at the field scale reflecting the temporal evolution of the orchard. The estimates and coefficient of variations (CVs) of the tree production were used to test different sampling strategies (random or clustered) for orchard yield estimation. The minimum number of trees to sample to reach 10% of error in the yield estimation was assessed for each strategy and cropping system type. We found that, for the extensive and intensive orchards, sampling trees by taking into account clusters of volume would not decrease the sampling effort compared to a random sampling. Nevertheless, an improvement of the accuracy of the yield estimation was obtained with increasing the sampling effort up to 10 and 20% of total trees for the intensive and extensive orchards, respectively. In the diversified orchard, the clustering of trees according to their crown volume decreased both the CV of yield estimation, from 21% to 6.6%, and the sampling effort, from 37 to 20% of the tree. This study showed that the combination of adapted sampling strategies and efficient tools for the estimation of tree production and orchard structure characterisation allow for accurate yield estimations of simple to complex orchards.

Keywords: perennial crops, tree production, spatial analysis, sampling method, unmanned aerial vehicles, machine vision

II.1. Introduction

The structure of orchards, defined by tree species composition and their spatial arrangement, first depends on the past and present cultural practices and the long-term strategy of the growers. But the development of perennial crops, as any plant formation, also rely on the spatiotemporal environment they are growing in, including soil properties and conditions (Grotti et al., 2019), topography (Aggelopoulou et al., 2010),

climate (Pan and Dong, 2018), and intra- and interspecific interactions (Aerts, 1999; Ngo Bieng et al., 2013b). Hence, variations of both practices and environmental factors in space and in time may induce heterogeneity in the structure (e.g., crown volume) of the trees. For fruit crops, the heterogeneity in tree conditions (structure and vigour) additionally leads to a spatiotemporal variation of the tree production, i.e., the number or weight of fruits per tree (Noble et al., 2018; Perry et al., 2010). The orchard yield, defined as the sum of individual tree production per hectare, is determined by both the orchard structure and tree production heterogeneities. At the regional or global scale, the diversity of orchard structure can vary from simple (e.g., monospecific commercial orchard) to highly complex (e.g., agroforestry systems in the humid tropics) with numerous associated species (Jagoret et al., 2017; Malézieux, 2012). In any cases, the orchard structure impacts the orchard yield, by directly defining the planting density (i.e., number of trees per hectare) and indirectly influencing the individual tree production by driving competition between trees (Menzel and Le Lagadec, 2017). It is commonly admitted that smallholder orchards are more diversified and more complex, displaying irregular planting patterns for instance, than large commercial orchards which are mostly monospecific (Malézieux, 2012). Management practices aim to reduce the tree spatial and temporal heterogeneities as growers seek to homogenize fruit quality while maximizing and maintaining the level of production over the years (Whelan and McBratney, 2000). However, in many cases, management practices are not uniformly applied. Smallholders may manage or modify their orchard locally: for instance, trees can be pruned in some parts of the orchard while, in others, trees are replaced by other species, etc. (Jagoret et al., 2011). Conversely, growers of large monospecific orchards may apply management practices homogeneously at the field or row scale (Whelan and McBratney, 2000). Thus, the effect of spatial deviations in management practices superimpose to the environment and orchard structure local disparities. In such context, we expect a higher variability of within-field tree production in smallholder systems than in the intensive ones.

Methodologies used for spatial analysis at the field scale reflect different approaches and objectives. Some studies on within-field heterogeneity take part of the precision agriculture that seeks to match resource inputs and management practices in order to optimize plant needs and maximize the yields (Whelan and McBratney, 2000). These studies mainly focus on monospecific orchards in temperate and, to a lesser extent, in tropical areas. Spatial deviation of tree structure, tree production, and fruit quality have therefore been described and related to various factors such as soil, topography, or other tree features (e.g., crown volume or vegetation index) in apple (Aggelopoulou et al., 2013, 2010), citrus (Mann et al., 2011; Zaman and Schumann, 2006), peach (Perry et al., 2010), or mango (Anderson et al., 2018) orchards. Due to a better accessibility of remote sensing tools, tree structure and production can be directly measured, using vegetation indices, and mapped at resolution finer than the tree (Usha and Singh, 2013). However airborne and satellite imagery remain limited by the high cost per hectare for data acquisition, by a low flexibility, and by contrasting results on remotely-sensed variables and tree production estimations (Anderson et al., 2018). On ground, computer vision systems allow to characterize tree structure and production using sensors carried by hands for simple tools (such as cameras or smartphones), or by a mobile platform for automated and data-intensive monitoring tools (Borianne et al., 2019; Stein et al., 2016). While these tools reach high accuracies for tree features assessment, an accurate mapping of within-field heterogeneity still requires expensive and high-tech solutions that are poorly applicable in complex orchards (i.e. orchards with irregular tree arrangement and numerous species). Other studies, taking part of the landscape ecology, assessed the plant diversity and spatial interactions in complex systems (e.g., Ngo Bieng et al., 2013a). Nevertheless, studies considering the within-field spatial variation of yields remain scarce in complex perennial-based orchards. This is mainly due to the lack of adapted tools for accurate yield measurement at

the tree and the orchard scales in that kind of systems (Carletto et al., 2015; Malézieux et al., 2018). The main reason is that the complexity of such systems makes the use of remote sensing technologies complicated, costly, and intensive in fieldworks (high resolution and expensive sensors needed). However, in the last decade, unmanned aerial vehicles (UAVs) have progressively proven to be an interesting alternative for precise mapping of tree features (Pádua et al., 2017). The acquisition of very high resolution (VHR) images with low operation and equipment costs, and with short revisiting period made UAVs appropriate for monitoring simple to complex wooded areas like forest or agroforestry systems. For instance, Sarron et al. (2018) developed an UAV-based workflow for tree structure and yield mapping in monospecific and diversified mango-based cropping systems in West Africa. However, while the authors succeed to estimate and map the orchard yield using tree structure as a proxy, the individual tree production was not measured.

Beyond the spatial within-field heterogeneity assessment, orchard yield estimation prior to harvest is critical for growers but remains inaccurately measured in most cases, especially in developing countries (Carletto et al., 2015). Indeed, when growers have the ability, yields are estimated before harvest by an overall visual evaluation of the orchard or by measuring the production of some sampled trees. In the case of tree sampling, Anderson et al. (2018) reported that 5% of the trees are usually randomly sampled. Then, orchard yield is extrapolated by multiplying the average number of fruit per sampled tree by the total number of trees in the orchard. Such method conducts to an inaccurate estimation, especially in complex orchards, as it does not take into account the within-field variability in the production per tree. Hence, the optimal sampling method for an accurate yield estimation has to be a trade-off between a low number of trees to sample (to reduce the sampling effort) and a relevant selection of trees to integrate the within-field production variability and decrease estimation errors. Precision agriculture address this issue as within-field heterogeneities can be used for spatial clustering of trees in function of their features (structure or production) and to build sampling strategies (Peeters et al., 2015; Anderson et al., 2018). In monospecific mango orchards, Anderson et al. (2018) estimated the minimum number of trees required to sample in order to reach an acceptable level of error of orchard yield estimation using different sampling strategies and tree production mapping methods. The authors succeeded in accurately estimating the orchard yields using different sampling strategies, which relied on an *a priori* estimation of tree production using remote sensing tools (satellite or machine vision system with RGB/LiDAR sensors). As a result, it is necessary to focus on the adaptability of these sampling strategies in more complex orchards where the use of remote sensing tools for prior estimation of tree production is restricted.

The present study explores the within-field heterogeneity among orchards with different levels of complexity and evaluates different sampling strategies to accurately estimate the orchard yield. We focused our study on mango (*Mangifera indica L.*) production in West Africa. In this region, mango production is of key importance for food security and its production has doubled in the last decade (FAOSTAT, 2017). Agriculture is mostly practiced in small farms - around 80% of farms in this region are of less 2 ha (Lowder et al., 2016) – and under a diversity of perennial-based cropping systems. Thus, different mango-based cropping systems co-exist in West Africa, from smallholder diversified orchards to large commercial-based monospecific orchards (Grechi et al., 2013). As mentioned above, these systems may display more or less within-field variability in tree production. Production variabilities are even more pronounced for mango trees than for others fruit crops because the species shows irregular reproductive behaviour (i.e., asynchronism and alternate bearing) (Dambreville et al., 2013). Yields of mango-based systems in West-Africa have been previously studied at the tree and orchard scales (Sarron et al., 2018). However, to our

knowledge, the within-orchard tree heterogeneities in structure and production have never been described. As seen previously, using the orchard spatial variability has shown interesting results for yield estimation prior to harvest. In this study, our first objective was to characterize and analyse the within-field heterogeneities in orchard structure and tree production for three mango-based orchards depicting the three main mango-based cropping systems found in West-Africa. Innovative tools based on UAV and ground image analysis were mobilized in that purpose. In a second time, heterogeneities were integrated in order to determine the best suited strategy of sampling to estimate the yields considering the structure of each orchard.

II.2. Materials and methods

a. Study sites

The study was carried out in three mango-based orchards of the Niayes region, a major agricultural production basin in western Senegal that covers more than 2,500 km², and located between 14°4' to 15°0' N and 16°6' to 17°2' W. The region is part of the Sudano-Sahelian zone and is characterized by unimodal rainfall from July to September (mean rainfall of 430 mm per year) and a relatively fresh and humid climate generated by the proximity to the Atlantic Ocean (with monthly mean temperatures ranging from 16 °C to 35 °C). The landscape is characterized by a succession of dunes, with sandy and non-lessive soils, and depressions, with clay and peaty soils (Fare et al., 2017). In these topographic depressions the groundwater comes up to the surface, ensuring accessibility to the resource for irrigation. Combined with the favourable climate, these conditions make this region one of the major fruit production areas in Senegal.

The three orchards have been selected to represent the three main types of mango-based cropping systems identified in the Niayes region (Grechi et al., 2013). Types are defined according to the crop composition (tree species and/or annual crops) and mango management practices. The first selected orchard, called "extensive", represents the smallholder extensive orchard type. It is mainly composed of mango trees planted on bare soil with no or low level of management practices (no pruning nor irrigation and chemicals). The second orchard, called "diversified", represents the smallholder diversified orchards. It is composed of mango trees along with other fruit tree species (citrus, papaya, etc.) and occasionally with market gardening underneath the trees (Fig. 65a). The selected orchard was highly diversified, composed of about half mango and half citrus trees, and displayed a large area dedicated to annual crops cultivated below the tree canopy. Thus, this orchard was similar to an agroforestry system and was chosen to be clearly distinct from the two others. The third orchard, called "intensive", represents the monospecific commercial-based orchards composed of one or two export cultivars ('Kent' and 'Keitt', Rey et al., 2007) with high level of external inputs. Orchards productive area, i.e., the orchard area when buildings are removed, varied from 1.19 to 1.52 ha and the mango planting density was 217, 174 and 383 trees.ha^{-1} for the extensive, diversified and intensive orchards respectively (Table 10).

b. Orchard land cover map and tree structure parameters

In order to characterize each orchard, an unmanned aerial vehicle (UAV) photogrammetry and object-based image analysis procedures were used following the method described in Sarron et al. (2018). Briefly, each selected orchard has been overflown using a UAV (DJI Mavic Pro, DJI Inc, China) in order to compute a high resolution orthomosaic (1.30 cm.pixel^{-1}; Fig. 66a) and a Canopy Height Model (CHM) using a photogrammetry procedure. The two layers were then analysed through an object-based classification

workflow to obtain a land cover map, containing 11 object classes, on which individual tree crown were delineated and tree species identified. Land cover map combined with the CHM permitted to accurately compute three tree structure parameters: tree height, crown area, and crown volume. Tree species identification permitted to classify mango, citrus (*Citrus sinensis L.*, *Citrus reticulate*, *Citrus paradisi*), cashew (*Anacardium occidentale*), papaya (*Carica papaya*), palm tree (*Eleis guineensis*), and other anecdotic, or bushy species (gum acacia: *Senegalia Senegal*, sapodilla: *Manilkara zapota*, or soursop: *Annona muricata*). In addition, areas of annual crops were delineated when present. Land cover maps were expressed as geospatial vectors with each object, including the delineated trees, corresponding to a geolocalized polygon and described in an adjunct database (.shapefile format) to be imported in GIS or spatial analysis software. Shapefile databases described object type (tree species class, annual crop, building, bare soil, etc.) and the tree structure parameters. They were then enriched with information on mango production and cultivars (Section II.2.c).

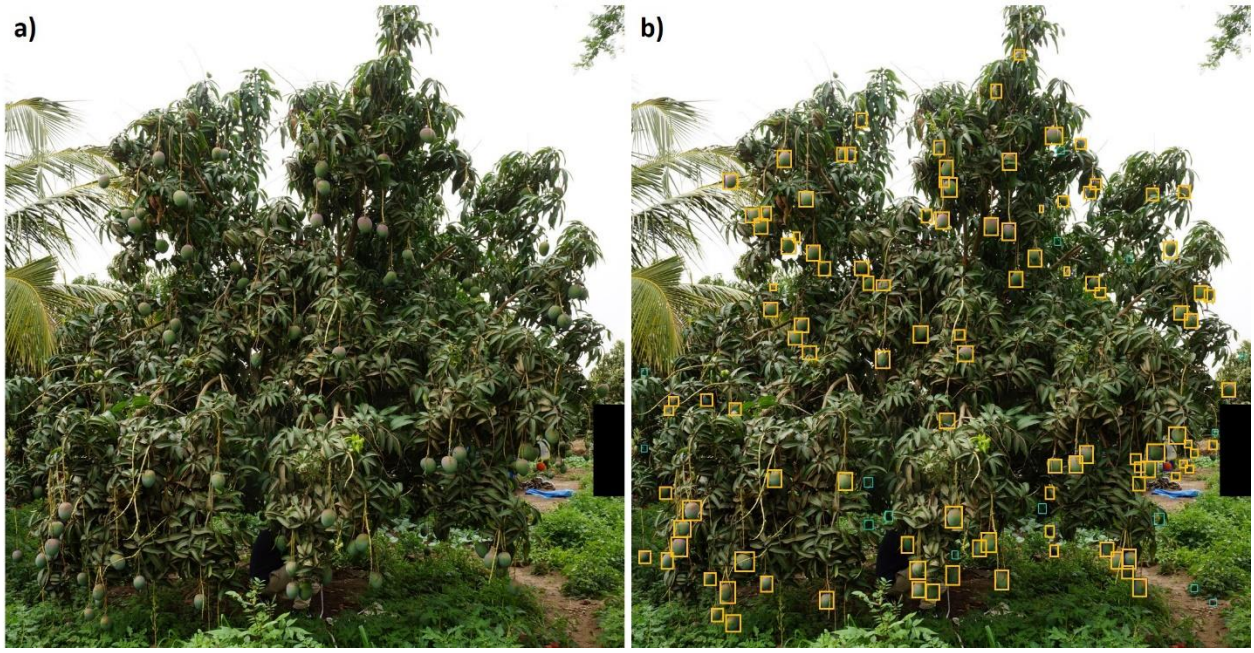


Fig. 65. (a) A mango tree of 'Kent' cultivar in the diversified orchard and (b) production estimation from machine vision system.

c. Estimation of mango tree production

During June 2018, one month to two weeks before harvest, the production of all mango trees was individually estimated in each orchard using a machine vision system. Each mango tree has been imaged on two opposite sides using a hand-held Sony Nex-7 RGB camera (Sony Corporation, New York, NY, USA) at a fixed focal length of 18 mm and 6000x4000 pixel size. During this image acquisition, the cultivar was identified by visual inspection of the tree. Then, a deep-learning image analysis algorithm was used to detect and count mangoes in the image (Fig. 65b; Borianne et al., 2019). The algorithm was a region-based convolutional neural network (CNN), called Faster R-CNN, and trained to accurately detect mango fruits in the crown. The algorithm reached a detection accuracy of 90% to detect mangoes from 'Kent', 'Keitt' and 'Boucodiékhal' cultivars, the three main cultivars found in the Niayes region. Details on Faster R-CNN functioning, parametrization, and evaluation are described in Borianne et al. (2019). Then, to estimate the

actual number of fruits in the tree, regression models were fitted between the number of fruit detected by the algorithm (automatic count) and the actual number of fruits counted in the field (field count). The three models calibrated for the main cultivars reached satisfying coefficient of determination (R^2) ranging from 0.92 and 0.95 following the cultivar and a normalized RMSE (NRMSE) of around 7%.

Thus, the actual number of fruits was obtained and geolocalized in the shapefiles of land cover maps for each mango tree. These estimates were then converted into individual tree production (in kg) using the average fruit weight measured for each cultivar in 2018. The fruit load (in $kg.m^{-3}$) was computed by dividing the tree production by the tree crown volume computed using UAV (Section II.2.b).

d. Orchard spatial metrics

Orchard structure was characterized by plant species diversity (or composition), tree density, and tree spatial configuration (or planting pattern). First, the orchard plant species diversity was described by the following metrics: i) the proportion of mango, citrus, and other trees (by grouping all trees that were neither mango nor citrus) in the total number of trees; ii) the number of mango cultivars; iii) the annual crop ratio, i.e. the ratio between area of annual crops and the total productive area; and iv) the Simpson's diversity index SI (Lande, 1996) computed using the area of each vegetation class. Second, plant population density variables were computed from the land cover map: tree planting density (i.e., number of trees per ha), tree planting density of mango (i.e., number of mango trees per ha), and the canopy volume of mango trees in the orchard (i.e., the total mango tree crown volume per ha). Last, the tree planting pattern was described using the Ripley's *L(r)* function for point pattern analysis (Ripley, 1977). The function calculated the expected number of neighbours, within a radius distance below or equal to r, of any point (i.e., tree centroid) in the orchard. The spatial organization of trees was compared to a complete spatial randomness (CSR) of the tree planting pattern as described in previous studies (Ngo Bieng et al., 2013a). Statistical significance of the difference between the actual and the CSR spatial structures was determined by comparing the observed *L(r)* value with a 95% confidence simulation envelope based on Monte Carlo simulations (n = 1000) of *L(r)* at CSR. The *L(r)* function allows to statistically determine the type of tree spatial point pattern at a fixed radius r: a *L(r)* value below the confidence envelope corresponds to a regular pattern; a *L(r)* value close to 0 or within the confidence envelope corresponds to a random pattern; and a L(r) value above the confidence envelope corresponds to a clustered pattern. Using the Ripley's *L(r)* function, planting pattern of mango trees was described for the three orchards at every distance between 0 to 15 meters radius. In the diversified orchard, L(r) values for all trees were also computed due to the high presence of other tree species.

e. Spatial heterogeneity in mango tree features

Land cover map and geolocalized estimations of tree structure and production allowed us to display three key features of mango trees: crown volume, production and fruit load. In addition, average (av.) and coefficient of variation (CV), computed as (average/standard-deviation)*100, were calculated for each features. Statistical comparison of tree features between orchards were tested using non-parametric Kruskal-Wallis test at a 95% confidence level because the three features presented a logarithmic distribution. In case of significant differences between tree populations, a pairwise comparison was conducted using non-parametric Dunn test. Using the same statistical analysis, tree features of the two dominant cultivars 'Kent' and 'Keitt' were compared in the diversified orchard.

To analyse the spatial heterogeneity of tree features, the spatial autocorrelation of each variable was computed. Spatial autocorrelation corresponds to the dependence of an observed value at a specific location with the values at neighbouring locations (Bivand et al., 2013). Similarly to the point pattern analysis (Section 2.4), spatial autocorrelation focuses on continuous variables and allows to determine if the identical values of the variable are rather randomly distributed, clustered, or dispersed. Several studies reported analysis of spatial autocorrelation within fields (e.g., Perry et al., 2010). In our case, the autocorrelation of each mango tree feature was calculated at the orchard scale using global Moran's *I* index (Bivand et al., 2013):

$$I = N \frac{\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j W_{ij}(x_i - \bar{x})^2} \quad (14)$$

where N is the total number of mango trees; x_i, x_j, and $\bar{x}$ are the variable values for the centred tree *i,* the neighbour trees *j,* and the variable mean respectively; W_{ij} is the spatial weight between centred-neighbour trees *i* and *j*. For each mango tree, the spatial weight of neighbour mango trees, within a fixed Euclidean distance, was affected using standardized values (i.e., the sum of weight values of all neighbour equal 1). We arbitrary fixed the Euclidian distance to 7 meters as it roughly corresponds to the maximum distance to encompass close mango trees at the orchard scale. Moran's *I* vary between -1 (perfect dispersion) to 1 (perfect clustering). Analysis of spatial autocorrelation and Moran's *I* significant test were proceeded using the R package '*spdep'* (Bivand et al., 2013).

f. Tree clustering

In the perspective of orchard yield estimation, clustering of trees of similar features may be useful to reduce the sampling effort especially in case of spatial dependence. Hence, a clustering of trees in function of their crown volume was carried out for each orchard. As described by Peeters et al. (2015), two types of clustering methods exist depending on whether spatial dependency information is included or not. In basic statistical clustering, methods do not take into account the spatial context of the data and the spatial autocorrelation is not considered. However, in case of an evidenced spatial autocorrelation, the value of the variable observed on a location depends on the values displayed by the neighbouring points. This spatially dependencies values might be useful for point clustering (Anderson et al., 2018). Thus, we used the local Getis-Ord G_i index, a local autocorrelation statistic, to describe the local spatial dependency of crown volume at the tree scale (Ord and Getis, 1995). The index evaluates the degree to which each feature is surrounded by similar values in its neighbourhood:

$$G_i(d) = \frac{\sum_j W_{ij}(d) x_j}{\sum_j x_j} \quad (15)$$

where G_i is the Getis-Ord index for the *i*-th centred tree within a distance *d*; x_j is the variable values for the neighbor *j* of tree *i*; W_{ij} is the spatial weight between centred-neighbour trees *i* and *j*. The degree of local autocorrelation for each point with its neighbours was given by the z-scores of the G_i test. Then, for each orchard two k-means clustering were proceeded based on tree crown volume: one using the observed values and one using the G_i z-scores. The k parameter of the k-means was set to three in both cases to build three clusters of trees of low, medium and high value of the variable.

g. Testing tree sampling strategies for orchard yield estimation

The tree production maps permitted to evaluate the interest of taking into account the within-field spatial heterogeneity to accurately estimate the orchard mango yield. We built three different strategies: i) *random sampling* (‘*random’*) of a fixed percentage of trees; ii) *cluster sampling* (‘*clust*’) based on tree clustering using observed value of crown volume; iii) *spatial cluster sampling* (‘*clust_sp*’) using crown volume spatial dependencies (G_i z-scores). For each strategy, it was possible to estimate the optimal number of trees to sample. As described by Anderson et al. (2018), the minimum number of trees to sample to obtain a desired percentage error (PE) can be computed by taking into account the CV of tree production:

$$n = \frac{S \times N}{S + N} \tag{16}$$

with:

$$S = \frac{(t \times CV)^2}{PE^2} \tag{17}$$

where, n is the minimum number of trees; N is the total number of trees in the orchard; t is the Student t statistic; CV is the coefficient of variation of tree production; and PE is the percentage of error. For a PE at 0.95 probability the t value is fixed to 1.66 using Student table (degree of freedom > 100). In this study, PE was fixed to 10% as it is the maximum percentage of error acceptable for yield estimation (Anderson et al. 2018, and personal observations).

The cluster sampling strategies took into account the within-field tree variability in crown volume using spatial dependency (“*clust_sp*”) or not (“*clust*”). From the two clusters described in Section II.2.f, the minimum number of trees to reach a PE of 10% in each cluster was computed using Eq. 16. This first assessment showed a minimum number of trees to sample systematically higher in each clustering than for the *random* strategy (see Results). In addition, these requirements exceeded 100 trees in the extensive and diversified orchard whatever *clust* or *clust_sp* was used. Therefore, only the sampling effort of the *random* strategy has been optimized (‘*random_optim’)* using the number of trees required to obtain 10 PE and was tested to evaluate the error of yield estimation. However, we also tested each type of strategies (*random*, *clust*, or *clust_sp*) using the same number of sampling trees fixed at a given percentage of total number of trees. This percentage was 20% of the trees for the extensive and diversified orchards, and 10% of the trees for the intensive one and were chosen as they correspond to a reasonable number of trees to sample (~50 trees) in each orchard. Then, the two clustered strategies at 20% or 10% (‘*clust’* and ‘*clust_sp’*) were compared with the *random* sampling at the same percentage (‘*random_20%’* or ‘*random_10%’*) for each orchard. Finally, we added the random sampling strategy at 5% (‘*random_5%’*) of the trees which corresponded to the sampling intensity commonly used by growers (Anderson et al. 2018 and personal observations).

Thus, in total, five sampling strategies were designed and tested for each orchard. For each sampling strategy, orchard production was calculated using the following procedure: 1) the average tree production of the sampled trees was calculated based on the estimation of individual tree production (Section II.2.b); 2) the orchard production was extrapolated by multiplying the average tree production by the number of trees in each cluster or in the entire orchard (for strategies with no clustering). Since all strategies included

a spatial random tree selection, misestimates of the total orchard yields could happen depending on tree selection. To avoid this effect, each random sampling was simulated 1,000 times. Simulations allowed to compute the CV of orchard production estimations gave by each sampling methods to compare them. For this purpose, the Feltz and Miller's statistical test for the equality of CV from different populations was used at a 99% confidence level, due to high differences in CVs (Feltz and Miller, 1996). In order to evaluate the effect of taking into account or not the cultivar in the random sampling strategy, the same five strategies were developed for yield estimation of 'Kent' and 'Keitt' trees in the diversified orchard (the only orchard with enough 'Keitt' trees). For each strategy, the random tree selections were weighted by the proportion of both cultivars in the entire orchard or clusters. The CV of cultivar-weighted selection was compared with the CV of a random selection with no consideration of the cultivars proportion using the Feltz and Miller's test.

Table 10. Orchard structure for the extensive, diversified and intensive systems.

Variable	Extensive	Diversified	Intensive
Productive area (ha)	1.24	1.52	1.20
Number of trees	287	531	456
% of mango	0.94	0.50	1
% of citrus	0.03	0.38	0
% of others trees	0.03	0.12	0
Number of mango trees	269	264	456
Number of mango cultivars	>4	>4	1
% of 'Kent' cv.	0.89	0.60	1
% of 'Keitt' cv.	0.03	0.31	0
% of 'Boucodiékhal' cv.	0.02	0.02	0
% of other cultivars	0.06	0.07	0
Annual crop ratio	0	0.24	0
Simpsons's index (SI)	0.36	0.70	0.07
Tree planting density (trees.ha^{-1})	231	349	383
Mango tree planting density (trees.ha^{-1})	218	174	383
Mango canopy volume (m^3.ha^{-1})	6,415	9,905	19,735

II.3. Results

a. Orchard structure

Land cover maps allowed to accurately describe the orchard structure i.e., the plant species diversity (composition) and the spatial arrangement of trees (Fig. 66b). As expected, the lowest Simpson's diversity index was obtained for the intensive orchard corresponding to a pure grove of mango trees (SI = 0.07, Table 10). The extensive orchard showed an intermediate level of species diversity (SI = 0.36) while the diversified orchard reached the highest species diversity (SI = 0.70). In the latter, whereas mango remained the main tree crop (50% of the total trees), the area was shared with citrus (38%) and other trees (12%), such as palm trees and sapodilla. In addition, annual crops were important in the diversified orchard as they represented around a quarter of the total productive area. The three orchards gathered a total of 989 mango trees with identified cultivar: 87% of 'Kent', 9% of 'Keitt', 1% of 'Boucodiékhal' and 3% of other local cultivars (Table 10). The group of other local cultivars contained more than 5 cultivars including 'Sewe' and 'Dieg

bou gatt' (Grechi et al., 2013). Both extensive and intensive orchards looked quite similar regarding their mango cultivar diversity: mango trees of the extensive orchard are mainly 'Kent' cultivar (90% of the 269 mango trees) and the intensive orchard was a pure grove of 'Kent' cultivar (Fig. 66c). Differently, the diversified orchard contained more than three mango cultivars with a large representation of 'Kent' (60%) and 'Keitt' (31%), and, to a lesser extent, 'Boucodiékhal' (2%) (Fig. 66c and Table 10).

Mango planting density and orchard canopy volume are known to be positively correlated variables (Menzel and Le Lagadec, 2017). In our study, both mango planting density and tree crown volume, which were significantly different among orchards (Kruskall-Wallis p-value < 0.001, Table 11), impacted the mango canopy volume. The highest volume of mango canopy (20,000 $m^{-3}.ha^{-1}$) had been observed for the intensive system due to a high planting density of 383 $trees.ha^{-1}$ and a high average tree crown volume of 43.28 m^3. On the opposite, the lowest mango canopy volume was observed for the extensive orchard: canopy volume was 6,500 $m^{-3}.ha^{-1}$ with an intermediate mango planting density of 218 $trees.ha^{-1}$ (Table 11) and a low average crown volume of 24.84 m^3 (Table 11). The diversified orchard had an intermediate mango canopy volume (10,000 $m^{-3}.ha^{-1}$) with an average tree crown volume of around 37.52 m^3 and a low mango planting density of 174 mango trees per ha due to the presence of trees from different species (total tree planting density was 349 $trees.ha^{-1}$, the same as in the intensive orchard). Extensive and diversified orchards displayed identical planting patterns with mango trees planted regularly around 3.5 m as shown by the plots of *L(r)* values (Fig. 67a). These results are confirmed by land cover maps as mango trees looked regularly planted locally for both orchards (Fig. 66). The diversified orchard showed an additional pattern, with mango trees clustered at large radius distance of 14 m (*L(r)* values significantly above the envelope) and, as shown in Fig. 2, trees appeared more concentrated on the upper right corner of the orchard. When considering all tree species, planting pattern in this orchard remained similar to the mango planting pattern with a regular planting at 2.5 m (Fig. 67b) and clustered planting above 10 m. In the intensive orchard, two regular patterns were observed at lower distances, 2 m and 5.5 m respectively, but at lower intensity (lower absolute *L(r)* values). In fact, the intensive orchard presented two distances for which planting pattern was regular: the first pattern detected at around 2 m corresponded to the in-row tree spacing (3 m, according to the grower and measured in the field), while the second regular pattern at around 5.5 m corresponded to the between-rows spacing (5m). A clustered pattern seemed to appear at 14 m (Fig. 67), which probably corresponded to an effect of the bottom edge as no trees were planted there.

b. Spatial heterogeneity in mango tree features

Descriptive statistics and plots of frequency distribution of mango tree features (i.e., crown volume, tree production, and fruit load) are presented in Table 11 and Fig. 68 respectively. Differences between orchards were significant for all features with p-value of Kruskal-Wallis tests below 0.001. The diversified orchard displayed an average tree production (av. = 75.61 kg) and average fruit load (av. = 2.44 $kg.m^{-3}$) significantly higher than the intensive one (61.56 kg and 1.75 $kg.m^{-3}$). Nevertheless, all features of the diversified orchard were more spread out (Fig. 68) and showed higher CV values (between 69 and 87%) than for the intensive orchard, which displayed the lowest CV (between 33 and 67%) among the three orchards. On the opposite, the extensive orchard had the lowest tree production (av. = 41.60 kg) and crown volume (av. = 24.84 m^3) but a high fruit load of 2.31 $kg.m^{-3}$ on average (Table 11). Thus, trees in the extensive orchard were in average smaller but with a higher fruit load. But tree production and fruit load, likewise tree crown volume, had CVs close or above 100% confirming the high observed spread out (Fig. 68). Tree features appeared more homogeneous in the intensive orchard than in the two others.

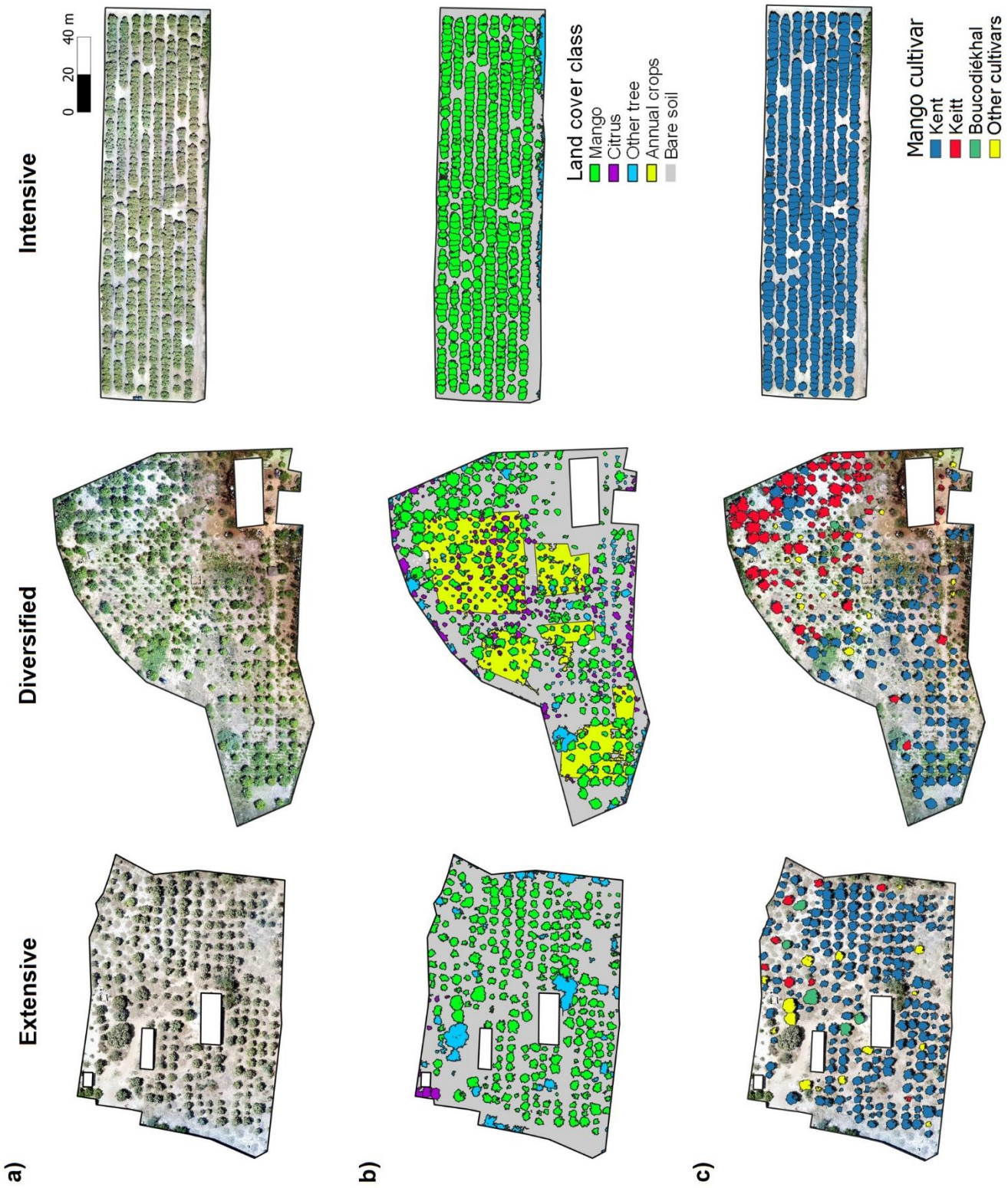


Fig. 66. (a) UAV-based orthomosaic, (b) land cover and planting arrangement maps and (c) mango tree cultivars in the three extensive, diversified, and intensive orchards. Tree crowns were delineated through an object-based procedure. Cultivars were directly identified in the field.

Cultivars comparison in the diversified orchard showed that both 'Kent' and 'Keitt' had significantly different crown volumes, with a lower volume for 'Kent' trees ('Kent' = 38.03 m^3; 'Keitt' = 48.59 m^3), but similar productions (both av. = 83.5 kg) due to a significantly higher average fruit load of 2.75 $kg.m^{-3}$ for 'Kent' compared to an average of 1.95 $kg.m^{-3}$ for 'Keitt' trees (data not shown).

The spatial heterogeneity of the tree features was estimated by Moran's *I* (Eq. 14) to identify spatial trends and is mapped in Fig. 69. All tree features presented a significant (p-value < 0.01) and positive Moran's *I* values in all orchard except for the fruit load in diversified and intensive orchards. However, despite the significance of the tests, the Moran's *I* index was higher than 0.2 only for the crown volume and for tree production in the diversified orchard and for tree production and fruit load in the extensive orchard (Fig. 69). Despite these four features were not defining a perfect clustering (i.e., Moran's *I* = 1), some clustering tendencies could be observed in addition to the high variability (CV) in the frequency distribution of values (Fig. 68). For instance, extensive orchard exhibited a higher fruit load in the up-right quarter than in the rest of the area (Fig. 69c). In the intensive orchard, only the tree production showed a spatial dependence but the Moran's *I* value was low and no spatial autocorrelation appeared in the map (Fig. 69b).

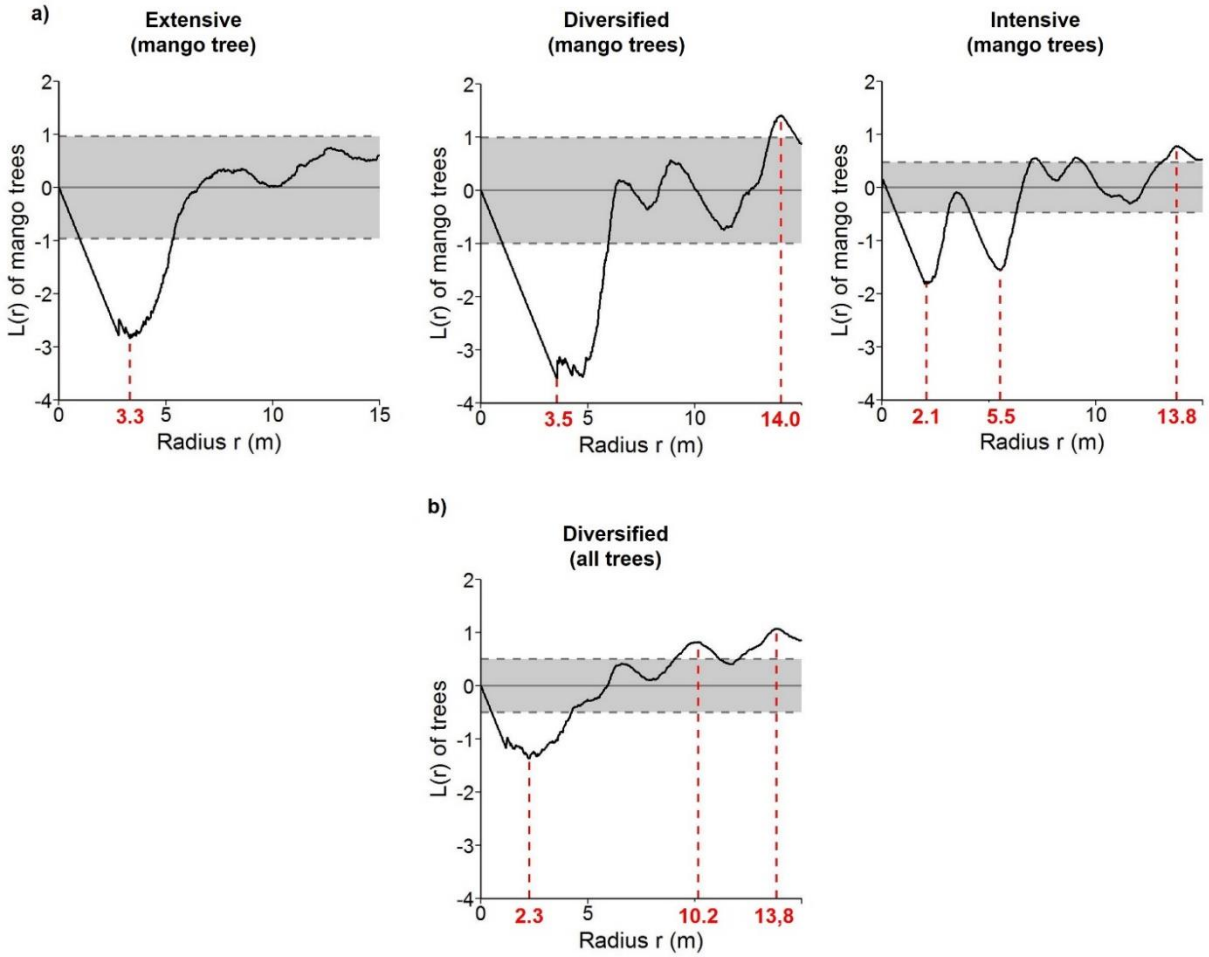


Fig. 67. Variation of Ripley's *L*(*r*) values with size of area analyzed for a) mango trees in the extensive, intensive and diversified orchards; and b) for all trees in the diversified orchard. All trees in the diversified orchard where represented by 50% of mango and 50% of other species (citrus mainly). The grey zone refers to the 95% confidence envelope into which *L(r)* corresponds to a random pattern. Vertical red dashed lines refer to the radius where regular (*L(r)* below the envelope) or cluster (*L(r)* above the envelope) are observed.

Table 11. Summary statistics of tree features for extensive, diversified, and extensive orchards: tree crown volume (assessed by UAV), tree production (assessed by ground image analysis), and fruit load. Letters indicate groups according to Dunn's tests.

Variable	Extensive		Diversified		Intensive	
	Average	CV (%)	Average	CV (%)	Average	CV (%)
Crown volume (m^3)	24.84c	95.22	37.52b	86.81	43.28a	43.69
Production (kg)	41.60c	116.18	75.61a	77.36	61.56b	32.67
Fruit load ($kg.m^{-3}$)	2.31a	141.13	2.44a	68.85	1.75b	66.86

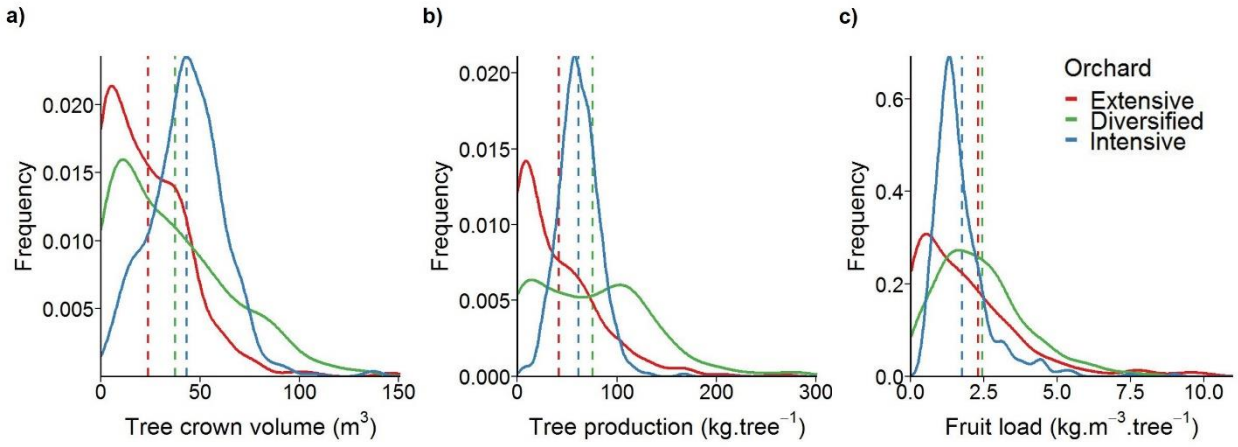


Fig. 68. Frequency distribution of the (a) tree crown volume; (b) tree production; and (c) tree fruit load in the extensive, diversified, and intensive orchards. Vertical dashed lines refer to the average value of the feature in each orchard.

c. Sampling strategies for orchard yield estimation

The minimum number of trees to sample in order to reach 10 PE in *random*, *clust*, and *clust_sp* strategies was computed depending on the tree production CV (Table 11) and using Eq. 16. Results show that the minimum number of trees to reach 10 PE was higher in the extensive orchard (between 56 and 77% of the total mango trees), than in the diversified orchard (between 37 and 63% of the total mango trees), and in the intensive orchard (between 6 and 16% of the total mango trees) depending on the sampling strategy (Fig. 70). Results for the *random* sampling gave a number of trees to sample of 150, 99, and 28 for the extensive, diversified, and intensive orchards, respectively (number of trees that were used to build the *random_optim* strategy).

Observed orchard yields were of 11.2, 20.0, and 28.1 $t.ha^{-1}$ for extensive, diversified, and intensive orchards, respectively (Table 12). Statistical comparison of sampling strategies gave contrasting results (Fig. 70). In all situations, the *random* sampling strategy of 5% of trees (*random_5%*), commonly used by growers, had the highest CV of 29.39, 20.58, and 6.57% in the extensive, diversified, and intensive orchard, respectively. In addition, the yield estimation with *random_5%* often exceeded or fell below 40% of the actual yield in the extensive and diversified orchards (Fig. 70). For the extensive orchard, the optimized sampling (*random_optim*) of 56% of mango trees reached the best CV of 6.23% while *random_20%*, *clust*, and *clust_sp* are equivalent and have a CV of around 13.5% (Table 12a).

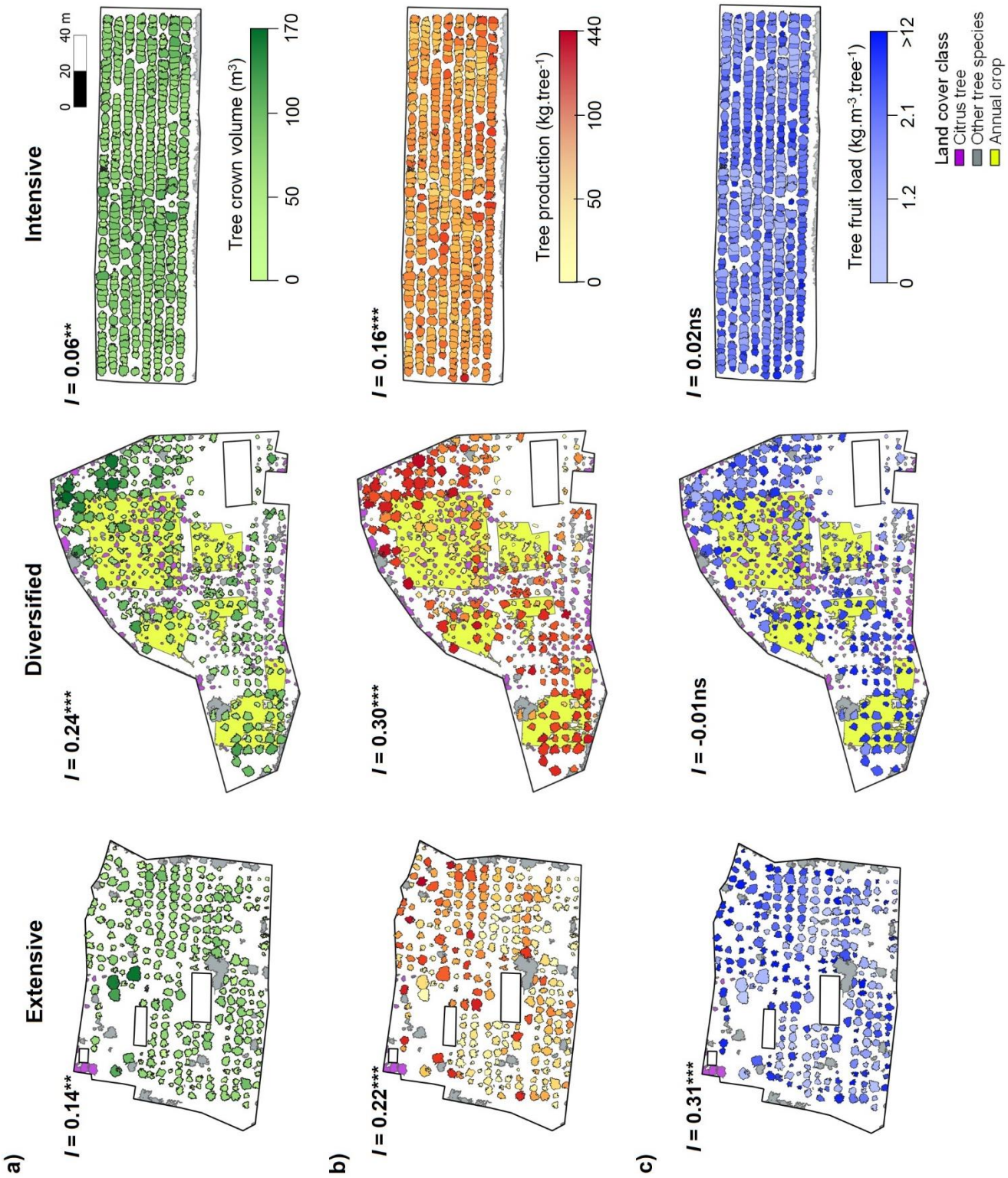


Fig. 69. Maps of spatial heterogeneity of the studied tree features: (a) tree crown volume (in m^3); (b) tree production (in kg); (c) and tree fruit load (in $kg.m^{-3}$) in the extensive, diversified, and intensive orchards. Moran's *I* statistics for global spatial autocorrelation and its significance at 95% confidence level (ns: non-significant; ** p-value < 0.01; *** p-value < 0.001) are displayed for appreciation of the global spatial autocorrelation.

Table 12. Results of the yield estimations and the number of trees sampled in the (a) extensive, (b) diversified, and (c) intensive orchards. Observed yield was compared to the estimations from five different sampling strategies.

(a) Extensive

	Orchard production		Tree sampled	
	Average (kg)	CV (%)	#	% of total
Observed	11189		269	100
Optimal sampling: *'random_optim'*	11199	6.23a	150	56
5 percent sampling: *'random_5%'*	11277	29.39c	14	5
20 percent sampling: *'random_20%'*	11117	14.08b	54	20
Cluster sampling: *'clust'*	11184	13.05b	54	20
Spatial cluster sampling: *'cluster_sp'*	11199	14.12b	54	20

(b) Diversified

	Orchard production		Tree sampled	
	Average (kg)	CV (%)	#	% of total
Observed	19962		264	100
Optimal sampling: *'random_optim'*	19993	6.11a	99	37
5 percent sampling: *'random_5%'*	20059	20.58c	14	5
20 percent sampling: *'random_20%'*	19935	9.49b	53	20
Cluster sampling: *'clust'*	20017	6.58a	53	20
Spatial cluster sampling: *'cluster_sp'*	19899	9.25b	53	20

(c) Intensive

	Orchard production		Tree sampled	
	Average (kg)	CV (%)	#	% of total
Observed	28072		456	100
Optimal sampling: *'random_optim'*	28116	5.92b	28	6
5 percent sampling: *'random_5%'*	28029	6.57c	23	5
10 percent sampling: *'random_10%'*	28085	4.49a	46	10
Cluster sampling: *'clust'*	28076	4.29a	46	10
Spatial cluster sampling: *'cluster_sp'*	28126	4.5a	46	10

Table 13. Difference in yield estimation CV (%) between the strategy when the cultivar proportion is taken into account (i.e., random sampling is weighted by proportion of cultivars) and the same strategy with no consideration of cultivar proportion, applied on 'Kent' and 'Keitt' trees in the diversified orchard. Significance of the differences were tested using the Feltz and Miller's test at a 99% confidence level (ns: non-significant; ** p-value < 0.01; *** p-value < 0.001).

	Difference in CV (%)
Optimal sampling: *'random_optim'*	-1,35***
5 percent sampling: *'random_5%'*	-0,92ns
20 percent sampling: *'random_20%'*	0,03ns
Cluster sampling: *'clust'*	-0,56**
Spatial cluster sampling: *'cluster_sp'*	-0,21ns

For the diversified orchard, *random_optim*, and *clust* strategies revealed as the best strategies with a CV of 6.11 and 6.58 respectively, while *random_20%* and *clust_sp* came afterward with a CV of around 9.4% (Table 12b). Finally, the *random_10%, clust* and *clust_sp* strategies in the intensive orchard attained the lowest CV which were below 5% while the *random_optim* CV reached a CV of 5.92% (Table 12c).

The diversified orchard was composed of 60% of 'Kent' and 31% of 'Keitt' cultivars that had the same production on average but different volumes (see II.3.b). Taking into account the cultivar proportion increased the minimum number of trees to sample, which reached 52% of the total trees for the *random_optim* strategy compared to 37% obtained without consideration of the cultivar (Fig. 70c and d). When random sampling was weighted by proportion of 'Kent' and 'Keitt', a significant reduction of 1.35 and 0.56 points in CV could be expected for *random_optim* and *clust* strategy, respectively (Table 13). Contrary to the cases with no cultivar taken into account, the best strategy for the diversified orchard, when weighted by cultivar proportion, was the *random_optim* which reached a CV of 4.5%.

II.4. Discussion

a. Characterisation of orchard heterogeneities

Few studies address within-field variability of tree features and tree production in mango orchards even monospecific. However, our results on the variation of tree production in the intensive orchard are consistent with those of Anderson et al. (2018) who highlighted a CV of the fruit number of mango trees varying from 27 to 93% in ten commercial mango orchards. Because drivers of spatial heterogeneity are scale dependent and variability of tree features are low, the studied area (1.20 ha) might be too small to observe a spatial tendency in the distribution of tree features. For instance, Mann et al. (2011) highlighted a yield and crown volume CVs of 50.43 and 53.94% respectively and a moderate to strong spatial dependence of tree features and soil characteristics over a 10ha citrus grove in USA. Nonetheless, a spatial heterogeneity of tree production was detected by Aggelopoulou et al. (2010) in a small 0.8 ha apple orchard with a low CV of 21%, and was correlated with the orchard topography. The Niayes region shows poor soil fertility (low C and N contents) and local water shortages (Fare et al., 2017). Thus, consequent heterogeneities may have been corrected, in the intensive orchard, by high levels of inputs used, especially through irrigation and fertilisation, hence contributing to explain the relative homogeneity and random spatial distribution of tree features.

Regarding the extensive orchard, the Simpson's diversity index (SI) remained low despite the presence of other tree species, because of the dominance of 'Kent' mango trees. However, extensive and intensive orchards had significantly different planting arrangements: density of mango tree planting was twice higher in the intensive orchard than in the extensive one. Despite the lower tree density reducing the competition between trees, the tree crown volume in the extensive orchard was the lowest (Table 11) and was even lower than the mean value of 66 m^3 for extensive orchards in the region (Sarron et al., 2018). Tree age, vigour, and/or management practices are possible factors explaining the low tree crown volume in the extensive orchard. Indeed, tree population in this orchard is mainly composed of small trees corresponding to either young individuals or trimmed trees that may explain the log-normal form of the crown volume values distribution compared to the normal distribution in the intensive orchard (Fig. 68). Moreover, the extensive orchard was not irrigated, which could conduct in a reduction of tree growth and development, particularly in dry areas (Schaffer et al., 2009). In addition, the tree features in the extensive orchard reached the highest CV among the study sites, which were associated with a significant spatial heterogeneity (Table 11 and Fig. 69). However, this orchard displayed a quite regular mango planting pattern (Fig. 67) and a low spatial

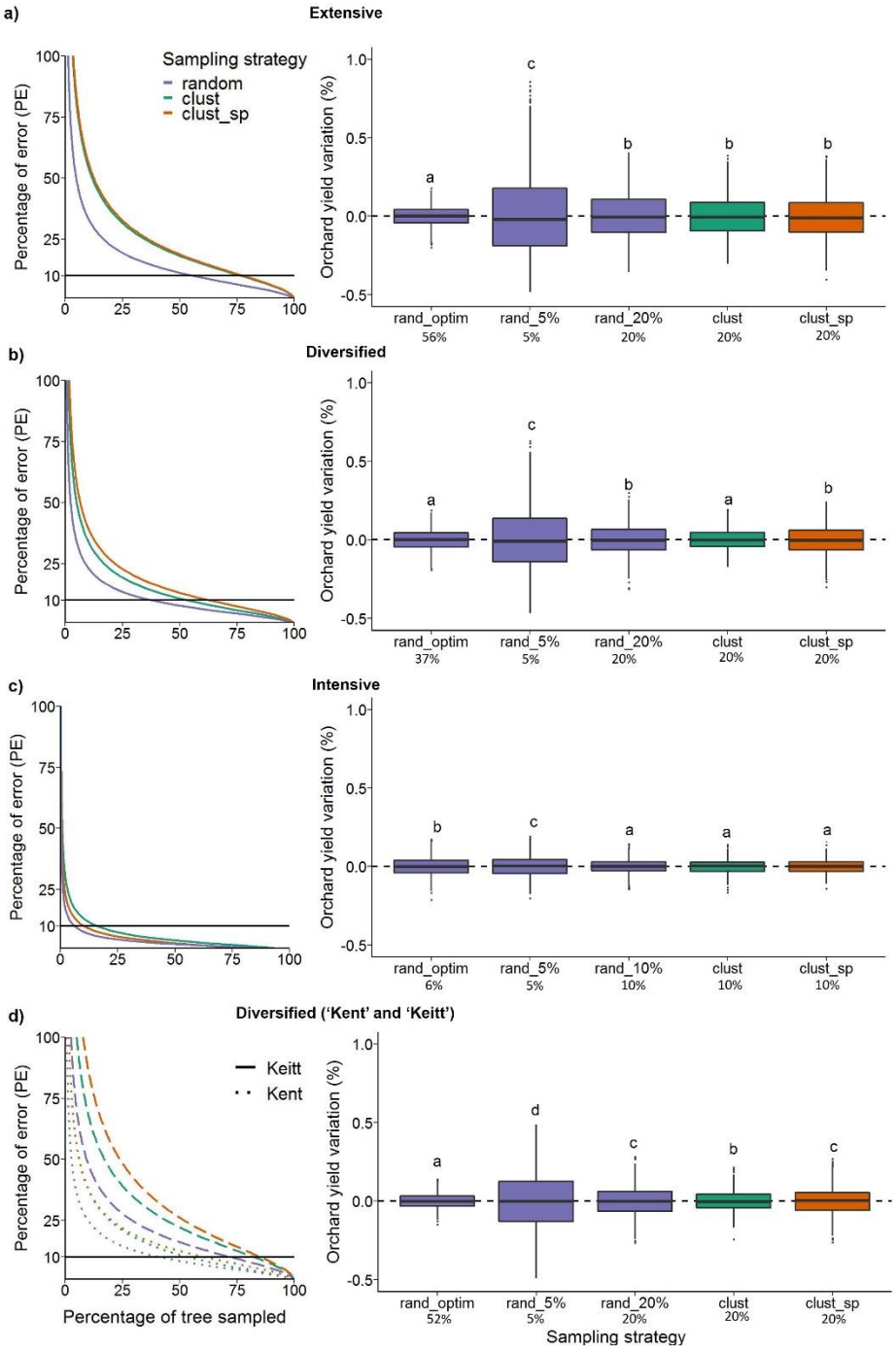


Fig. 70. Percentage of error (PE) in function of the percentage of tree sampled (left panel) and variation of orchard yield estimates for the different sampling strategies (right panel) for the (a) extensive, (b) intensive, and (c) diversified orchards. (d) corresponds to percentage of tree sampled and evaluation of sampling strategies for 'Kent' and 'Keitt' cultivars in the diversified orchard. Five sampling strategies are displayed: optimized random sampling ('rand_optim'), random sampling of 5 and 10% or 20% of the total trees ('rand_5%', 'rand_10%' and 'rand_20%'), cluster sampling using volume ('clust') or spatial statistics on volume ('clust_sp'). In the left panel, PE is expressed in function of the tree production coefficient of variation (CV) and horizontal line correspond to 10 PE. In the right panel, orchard yield variation refers to the difference between the estimation and the actual yield divided by the actual yield. Letters indicate groups according to Feltz and Miller's tests at 99% confidence.

heterogeneity in tree crown volume, confirming that smallholders pay attention to build regular patterns, as well as industrial growers. Consequently, the orchard structure is not the main driver explaining the spatial heterogeneity of tree production. Possible drivers of heterogeneity are i) the spatial variation of soil characteristics (both fertility and water content) that might be exacerbated by the low input characterizing this cropping system; ii) and the morphologic and/or genetic heterogeneity of planting material, especially the rootstocks, at the time of plantation.

Contrastingly, the diversified orchard appeared as the most diverse as half of the trees were other species than mango, and annual crops occupied a large part of its area. This balanced occupation of the area by different species resulted in a high SI of 0.70 (Table 10). Due to its planted diversity and structure, the diversified orchard may be considered as multi-strata agroforestry system, such as cocoa systems (Jagoret et al., 2017). Indeed, this orchard is composed of three vertical layers: the upper layer with mango trees, the intermediate layer with citrus and other small trees and, on the ground, the annual crops layer. Like the extensive orchard, the diversified orchard showed a regular mango tree pattern (at 3.5 m) even when all trees were considered (Fig. 67). Nevertheless, the tree density of this orchard appears also clustered at a larger scale (around 10 m). This may indicate that the grower tended to plant mango trees regularly at the beginning and then have progressively replaced some old or died trees, added tree crops from other species or trimmed branches to graft new cultivars. These changes may have progressively disorganized the orchard structure and conduct to a clustering of tree density and volume (Jagoret et al., 2011). The spatial heterogeneities of this orchard were demonstrated by a high CV and a significant Moran's *I* index for tree crown volume and tree yield. Considering the feature map, tree density and volume were higher in the upper right part of the field. The higher canopy volume observed in that part of the orchard may be explained by the high proportion of 'Keitt' trees (Fig. 66) as 'Keitt' trees reach higher crown volume compared to 'Kent' (Section III.3.b and Sarron et al., 2018). Another driver of the spatial heterogeneity may be the tree age as the grower explained that the upper right part was the oldest part of the orchard. Soil variabilities and management practices are also possible drivers of orchard heterogeneity, however additional data depicting those factors are actually necessary. Finally, whereas the tree crown volume and tree production both showed spatial tendencies, the fruit load was randomly distributed in the diversified orchard. Despite the heterogeneity evidenced in this orchard, it seems to have no effect on the tree performances.

b. Choosing the "best" sampling strategy

The orchard spatial characterisation could be used for a better estimation of the orchard yield using adapted sampling methods, which corresponded to the second objective of this work. Five sampling strategies were built in function of a random sampling (*random*) or a sampling based on a clustering of trees considering the crown volume values or spatial dependencies (i.e., *cluster*, *cluster_sp*). The minimum number of trees required to reach a percentage of error (PE) of 10 was estimated and used to build the *random_optim* strategy. This strategy reached an acceptable CV of around 6% in all orchards (Table 12). However, the number of trees to sample to obtain 10 PE varied greatly among orchards. In smallholder extensive and diversified orchards, the sampling numbers for the *random_optim* strategy represented 56% and 37% of total trees respectively (representing more than 100 trees). For the three orchards, the cluster samplings *clust* and *clust_sp* always required a higher number of trees to reach 10 PE compared to the *random* strategy. For these reasons, the two cluster strategies were tested with a number of trees to sample of about 50 (i.e, 20% in extensive and diversified orchards and 10% in the intensive, Table 12). Such percentages greatly decrease the sampling effort to an acceptable level that would be viable for growers.

In all orchards, the strategy to sample 5% of the trees (*random_5%*) reached the highest CV which was even superior to 20% in extensive and diversified orchards. Hence, when using *random_5%* strategy the probability that the yield estimation is 20% higher or lower than the actual yield (i.e., orchard yield variation above 0.2 or below -0.2, Fig. 70) was of 47 and 32% in the extensive and diversified orchards respectively. This meant that this sampling strategy, commonly used by growers (Anderson et al. 2018, and personal observations), is no longer relevant especially for smallholder orchards which display high CV of tree production (Table 11).

In the intensive orchard, the *random_optim* strategy requires 28 trees to be sampled or 6% of the total number of trees. This value is within the range of 2.5 to 41% of sampled trees to attain 10 PE found by Anderson et al. (2018) for commercial mango orchards. In the intensive orchard, the tree clustering did not improve the accuracy of orchard yield estimation. However, in that case, the non-clustered sampling *random_10%* gave better results than the *random_optim* strategy with a decrease of CV of 1.5 points (from 6% to 4.5%, Table 12c). In this cropping system, growers generally sample only 5% of the trees. Therefore, an increase in random sampling intensity, up to 10% (or around 50 trees), for yield estimation should be recommended and could be supported by fast tools for yield estimation. In the extensive orchard, the strategies *random_20%*, (i.e. 54 trees sampled) is more suitable due to lower sampling intensity. Despite a within-field spatial heterogeneity evidenced in the extensive orchard, both strategies based on tree clustering did not improve the CV of the sampling yield estimation compared to the *random_20%* strategy (Fig. 70). However, a sampling of 20% of trees decreased by half the CV of yield estimation when compared to the sampling of 5% if trees (Table 12a). Thus, for this orchard, the random sampling of 20% of trees would be recommended as a compromise between yield estimation accuracy and sampling intensity. In the diversified orchard, best yield estimations were obtained by *random_optim* and *clust* strategies (Fig. 70). This meant that clustering trees depending on of their crown volume halved the number of trees to sample and obtained a CV value of 6.6% (Table 12b). In addition, a small improvement is obtained by weighted the sampling number using the proportion of 'Kent' and 'Keitt' cultivars. However, this implies that the proportion of tree cultivars is accurately known.

Conclusions on whether or not tree clustering helped increasing the accuracy of orchard yield estimations were contrasted. On one hand, the improvement of yield estimation by tree clustering were unsuccessful for intensive and extensive orchards. These results confirm observations made by Anderson et al. (2018) who observed no improvement in yield estimation even by clustering trees in function of their spatial variation of fruit production. On the other hand, the cluster of trees using their crown volume values permitted to improve the accuracy of yield estimation in the diversified orchard. Several studies underline the interest of tree clustering to delineate orchard zones for precision management (Mann et al., 2011; Zaman and Schumann, 2006), however a few have used this method for improving yield estimations. Peeters et al. (2015) described a better tree partitioning using the spatial clustering with Getis-Ord statistics, over the non-spatial clustering. But this was not the case here. Indeed, the use of G_i statistic to group trees gave higher CV than simply using the crown volume values. The contrasting results obtain in the present study on mango orchards highlight that non-spatial tree clustering may improve yield estimation only in complex orchards containing several cultivars. Unlike other studies (Aggelopoulou et al., 2013; Anderson et al., 2018), we choose not to integrate tree production or fruit quality information in the clustering process. This gave the advantage to build cluster of trees with no *a priori* information on their productivity. In addition, tree structure varies less over time than tree production especially for mango trees which have an alternate bearing behaviour. Nonetheless, more research is needed on tree clustering when using other tree features

more correlated to the tree production, like vegetation index (Zaman and Schumann, 2006) and trunk circumference (Peeters et al., 2015), or other clustering methods (Aggelopoulou et al., 2013). Finally, the existence of tools for real-time yield estimation in the field (Stein et al., 2016) should stimulate interest and development of innovative tools for guided real-time sampling.

II.5. Conclusion

In this study, we could characterize the structure and within-field heterogeneities of three different orchards representative of the three main cropping systems found in the Niayes region, a major mango producing and exporting region in Senegal. We used a validated procedure combining photogrammetry and object-based classification to accurately delineate tree crown, extract tree features, and identify tree species. The analysis of the orchard structure and within-field spatial heterogeneity of planting pattern and tree features allowed us to highlight the different patterns of planting density and tree features in the three mango orchards: the intensive orchard tended to be homogeneous in tree feature and planting pattern while the extensive orchard showed higher variability in tree performance and higher spatial heterogeneity. In the diversified orchard, the planting pattern and the spatial heterogeneity of tree structure reflected the temporal evolution of the orchard configuration and composition. In addition, the complexity of the diversified orchard, which can be considered as an agroforestry system, did not prevent a random distribution of the fruit load.

The spatial estimation of tree production was used to test different sampling strategies that included random sampling on the entire orchard and on clusters build from trees crown volume. We concluded that the random sampling of 5% of the trees, commonly applied by growers, gave high yield estimation errors in the extensive and diversified orchards due to high variability in tree production. In addition, clustering of trees depending on the crown volume do not improve the yield estimation in extensive and intensive orchard. Thus, an increase of sampling effort up to 20% and 10% of the trees is recommended for the extensive and intensive orchards, respectively. In the diversified orchard, a sampling of 20% of trees based on non-spatial clustering of trees by taking into account tree crown volume, weighted or not by cultivars proportion, is recommended as it reduced the sampling effort.

To the author's knowledge, this is the first analysis of within-field variability carried out on mango-based orchards in West-Africa. As such analysis of smallholder extensive and diversified orchards is quite novel, no comparison with other studies was possible. Whereas these orchards depicted the three distinct cropping systems found in the study area, additional research on an extended orchard network will permit to better compare the cropping systems. Moreover, this study opens the research field for a within-field analysis of perennial crop yields, a necessary step for improving management practices of small growers of perennial crops in the tropics.

References

Aerts, R., 1999. Interspecific competition in natural plant communities: mechanisms, trade-offs and plant–soil feedbacks. J. Exp. Bot. 50, 29–37. https://doi.org/10.1093/jxb/50.330.29

Aggelopoulou, K., Castrignanò, A., Gemtos, T., Benedetto, D.D., 2013. Delineation of management zones in an apple orchard in Greece using a multivariate approach. Comput. Electron. Agric. 90, 119–130. https://doi.org/10.1016/j.compag.2012.09.009

Aggelopoulou, K.D., Wulfsohn, D., Fountas, S., Gemtos, T.A., Nanos, G.D., Blackmore, S., 2010. Spatial variation in yield and quality in a small apple orchard. Precis. Agric. 11, 538–556. https://doi.org/10.1007/s11119-009-9146-9

Anderson, N.T., Underwood, J.P., Rahman, M.M., Robson, A., Walsh, K.B., 2018. Estimation of fruit load in mango orchards: tree sampling considerations and use of machine vision and satellite imagery. Precis. Agric. https://doi.org/10.1007/s11119-018-9614-1

Bivand, R., Pebesma, E.J., Gómez-Rubio, V., 2013. Applied spatial data analysis with R, second edition. ed, Use R! Springer, NY.

Borianne, P., Sarron, J., Borne, F., Faye, É., 2019. Deep Mangoes: from fruit detection to cultivar identification in color images of mango trees., in: DISP'19, International Conference on Digital Image and Signal Processing. Oxford, UK, p. 7., arXiv:1909.10939.

Carletto, C., Jolliffe, D., Banerjee, R., 2015. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. J. Dev. Stud. 51, 133–148. https://doi.org/10.1080/00220388.2014.968140

Dambreville, A., Lauri, P.-E., Trottier, C., Guedon, Y., Normand, F., 2013. Deciphering structural and temporal interplays during the architectural development of mango trees. J. Exp. Bot. 64, 2467–2480. https://doi.org/10.1093/jxb/ert105

FAOSTAT, 2017. FAO, Food and Agriculture Organization of the United Nations, Rome, Italy [WWW Document]. URL http://www.fao.org/faostat/en/#home

Fare, Y., Dufumier, M., Loloum, M., Miss, F., Pouye, A., Khastalani, A., Fall, A., 2017. Analysis and Diagnosis of the Agrarian System in the Niayes Region, Northwest Senegal (West Africa). Agriculture 7, 59. https://doi.org/10.3390/agriculture7070059

Feltz, C.J., Miller, G.E., 1996. An asymptotic test for the equality of coefficients of variation from k populations. Stat. Med. 15, 647–658. https://doi.org/10.1002/(sici)1097-0258(19960330)15:6<647::aid-sim184>3.0.co;2-p

Grechi, I., Sane, C.A.B., Diame, L., De Bon, H., Benneveau, A., Michels, T., Huguenin, V., Malezieux, E., Diarra, K., Rey, J.-Y., 2013. Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68, 447–466. https://doi.org/10.1051/fruits/2013094

Grotti, M., Chianucci, F., Puletti, N., Fardusi, M.J., Castaldi, C., Corona, P., 2019. Spatio-temporal variability in structure and diversity in a semi-natural mixed oak-hornbeam floodplain forest. Ecol. Indic. 104, 576–587. https://doi.org/10.1016/j.ecolind.2019.04.014

Jagoret, P., Michel, I., Ngnogué, H.T., Lachenaud, P., Snoeck, D., Malézieux, E., 2017. Structural characteristics determine productivity in complex cocoa agroforestry systems. Agron. Sustain. Dev. 37. https://doi.org/10.1007/s13593-017-0468-0

Jagoret, P., Michel-Dounias, I., Malézieux, E., 2011. Long-term dynamics of cocoa agroforests: a case study in central Cameroon. Agrofor. Syst. 81, 267–278. https://doi.org/10.1007/s10457-010-9368-x

Lande, R., 1996. Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. Oikos 76, 5. https://doi.org/10.2307/3545743

Lowder, S.K., Skoet, J., Raney, T., 2016. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. World Dev. 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041

Malézieux, E., 2012. Designing cropping systems from nature. Agron. Sustain. Dev. 32, 15–29. https://doi.org/10.1007/s13593-011-0027-z

Malézieux, É., Rabodomanantsoa, A., Sarron, J., Faye, É., 2018. Forecasting crop yields for perennials in the tropics: looking for an innovative multiscale methodology. Presented at the Workshop CLAND : forecasting crop yields from data, models, and expert knowledge, Paris, France, p. 1.

Mann, K.K., Schumann, A.W., Obreza, T.A., 2011. Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data. Precis. Agric. 12, 457–472. https://doi.org/10.1007/s11119-010-9189-y

Menzel, C.M., Le Lagadec, M.D., 2017. Can the productivity of mango orchards be increased by using high-density plantings? Sci. Hortic. 219, 222–263. https://doi.org/10.1016/j.scienta.2016.11.041

Ngo Bieng, M.A., Gidoin, C., Avelino, J., Cilas, C., Deheuvels, O., Wery, J., 2013a. Diversity and spatial clustering of shade trees affect cacao yield and pathogen pressure in Costa Rican agroforests. Basic Appl. Ecol. 14, 329–336. https://doi.org/10.1016/j.baae.2013.03.003

Ngo Bieng, M.A., Perot, T., de Coligny, F., Goreaud, F., 2013b. Spatial pattern of trees influences species productivity in a mature oak–pine mixed forest. Eur. J. For. Res. 132, 841–850. https://doi.org/10.1007/s10342-013-0716-z

Noble, A.E., Rosenstock, T.S., Brown, P.H., Machta, J., Hastings, A., 2018. Spatial patterns of tree yield explained by endogenous forces through a correspondence between the Ising model and ecology. Proc. Natl. Acad. Sci. 115, 1825–1830. https://doi.org/10.1073/pnas.1618887115

Ord, J.K., Getis, A., 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. Geogr. Anal. 27, 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Pádua, L., Vanko, J., Hruška, J., Adão, T., Sousa, J.J., Peres, E., Morais, R., 2017. UAS, sensors, and data processing in agroforestry: a review towards practical applications. Int. J. Remote Sens. 38, 2349–2391. https://doi.org/10.1080/01431161.2017.1297548

Pan, J., Dong, L., 2018. Spatio-temporal variation in vegetation net primary productivity and its relationship with climatic factors in the Shule River basin from 2001 to 2010. Hum. Ecol. Risk Assess. Int. J. 24, 797–818. https://doi.org/10.1080/10807039.2017.1400373

Peeters, A., Zude, M., Käthner, J., Ünlü, M., Kanber, R., Hetzroni, A., Gebbers, R., Ben-Gal, A., 2015. Getis–Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. Comput. Electron. Agric. 111, 140–150. https://doi.org/10.1016/j.compag.2014.12.011

Perry, E.M., Dezzani, R.J., Seavert, C.F., Pierce, F.J., 2010. Spatial variation in tree characteristics and yield in a pear orchard. Precis. Agric. 11, 42–60. https://doi.org/10.1007/s11119-009-9113-5

Rey, J.-Y., Diallo, T.M., Vannière, H., Didier, C., Kéita, S., Sangaré, M., 2007. The mango in French-speaking West Africa: varieties and varietal composition of the orchards. Fruits 62, 57–73. https://doi.org/10.1051/fruits:2006051

Ripley, B.D., 1977. Modelling spatial patterns. J. R. Stat. Soc. B 39, 172–212.

Sarron, J., Malézieux, É., Sané, C., Faye, É., 2018. Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. Remote Sens. 10, 1900. https://doi.org/10.3390/rs10121900

Schaffer, B., Urban, L., Lu, P., Whiley, A.W., 2009. Ecophysiology, in: Litz, R.E. (Ed.), The Mango: Botany, Production and Uses. CABI, Cambridge, USA, pp. 170–209.

Stein, M., Bargoti, S., Underwood, J., 2016. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors 16, 1915. https://doi.org/10.3390/s16111915

Usha, K., Singh, B., 2013. Potential applications of remote sensing in horticulture—A review. Sci. Hortic. 153, 71–83. https://doi.org/10.1016/j.scienta.2013.01.008

Whelan, B.M., McBratney, A.B., 2000. The "null hypothesis" of precision agriculture management. Precis. Agric. 2, 265–279. https://doi.org/10.1023/A:1011838806489

Zaman, Q.-, Schumann, A.W., 2006. Nutrient Management Zones for Citrus Based on Variation in Soil Properties and Tree Performance. Precis. Agric. 7, 45–63. https://doi.org/10.1007/s11119-005-6789-z

III. Understanding drivers of mango yields in Sub-Saharan Africa: insights from regional agronomic diagnosis

Les travaux sur l'estimation du rendement du verger ont permis d'obtenir deux méthodes (par modélisation et échantillonnage) intéressantes et adaptés aux vergers de la zone des Niayes. Le rendement de chaque verger de l'échantillon a été estimé en utilisant la méthode de modélisation par drone (combinant l'extraction de la structure des arbres et la modélisation), mise en place plus tôt dans la thèse et décrite dans le Chapitre 2 – Partie I. Cette partie est un article décrivant l'analyse de variabilité inter-verger dans la zone des Niayes (Fig. 32, p. 66). La méthode du diagnostic agronomique régional a été appliquée afin de comparer les performances des 29 vergers de l'échantillon. Cette méthodologie permet d'analyser les vergers à l'échelle régionale et d'identifier les facteurs impactant le rendement.

Cet article scientifique en cours de finalisation doit être soumis à une revue à comité de lecture (revue visée *Agronomy for Sustainable Development*).

- Sarron J.; Faye É.; Malézieux É. **Understanding drivers of mango yields in Sub-Saharan Africa: insights from regional agronomic diagnosis.**

Understanding drivers of mango yields in Sub-Saharan Africa: insights from a regional agronomic diagnosis

Julien Sarron[1,2,3,*], Emile Faye[1,2,3], Eric Malézieux[1,2]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]HortSys, University Montpellier, CIRAD, F-34090 Montpellier, France
[3]Centre pour le Développement de l'Horticulture, ISRA, Dakar 14000, Senegal
*Corresponding author: julien.sarron@cirad.fr

Abstract

In West Africa, fruit trees like mango (*Mangifera indica L.*) trees provided several ecosystem services, especially in agroforestry systems, which are useful for food security. However, due to lack of information at the tree and orchard scales, the productivity performances of the mango-based cropping systems are still not well understood in this region. The aim of this study was to carry out an analysis of factors that are responsible for orchard yield variations among three main cropping systems (extensive, diversified, and intensive) in the Niayes region, a major mango production basin in Senegal. We adapted the regional agronomic diagnosis to decipher the effects of orchard structure (i.e., species diversity, planting density, tree configuration, etc.), management practices and environment on three orchard productivity variables: the orchard yield, the tree production and the fruit load (i.e., yield by unit of biomass). Comparison of cropping systems showed that the intensive orchards reached the highest orchard yield, however the diversified systems observed the highest average tree production. The extensive system had the lowest orchard yield, tree production and fruit load despite this system had similar mango biomasses than the diversified. The low performances of the extensive system was mainly due to the poor maintenance of the trees (e.g., pruning, inputs, etc.). Analysis of drivers showed that mango planting density was the main driver of the higher yields obtained in the intensive system. On the opposite, the lower density of the diversified system seems compensated by a little increase of biomass at the tree scale conducting in higher tree production *P*, negatively correlated to the planting density. This study gave new insights for the orchard yields analysis and highlighted the usefulness of the fruit load variable to compare cropping system in West Africa.

Keywords: tree crop, cropping system, agroforestry

III.1. Introduction

In the world, fruit tree crops provide important ecosystem services, among them of course the production of fruits, an important component for healthy human diets (WHO 2004; Demestihas et al. 2017). Mango (*Mangifera indica L.*) is 5th fruit produced in the world mainly in semi-arid, tropical and subtropical areas (Mukherjee and Litz 2009; FAOSTAT 2017). Mango production is of main importance for food security in developing countries such as in West Africa where its production has doubled in the last decade (FAOSTAT 2017). In this region, various mango-based cropping systems co-exist, from smallholder extensive orchards to large commercial-based monospecific orchards (Grechi et al. 2013) but access to information on orchard performance and management practices remains scarce (Carletto et al. 2015). Mango systems highly vary in terms of structure (tree pattern, species diversity), management, and productivity. Hence, West African mango-based systems may display more or less complexity in plant composition and variability in tree

production and orchard yield. But key drivers of orchard productivity in the different mango-based systems remain unknown in these systems.

The structure of an orchard is defined by the tree species composition and the tree spatial arrangement (planting density and pattern) which both depend on the past and present cultural practices and the long-term strategy of the growers (Malézieux et al. 2009; Jagoret et al. 2011). The orchard structure combined with environmental factors (soil and climate) can impact the orchard productivity (Fig. 71) by defining the development of productive biomass and light interception of the orchard canopy, determining the production potential at the orchard scale (Menzel and Le Lagadec 2017). However, at the tree scale, the tree production is impacted by several factors related to orchard structure (i.e., competition between trees), management practices, and the climatic conditions during the reproductive cycle (Fig. 71, Schaffer et al. 2009). Depending on the system, grower priorities may vary. In monoculture, growers seek to maximize the yield (i.e., the output per hectare) of a sole crop. They tend to improve orchard yield by maximizing the crop canopy biomass and in the same time reducing between-tree competition for light (Menzel and Le Lagadec 2017). On the other hand, in agroforestry and diversified systems, growers seek for both diversification of income and services and resilience (Malézieux 2009).

At the regional scale, a production basin of perennial crops may display a high variability in cropping systems. This is true for mango (Grechi et al. 2013) but also cocoa or others perennials (Jagoret et al. 2011). A comparison of the structure and performances of cropping systems at the regional scale is a preliminary step before understanding key drivers of yield and quality, improving management practices, and defining innovative cropping systems. Regional agronomic diagnosis is a methodological framework that seek to relate cropping system performance with farmers' practices (Doré et al. 1997, 2008). This methodology has been mainly used to identify and rank drivers of variability in annual crop yields at the regional scale (Doré et al. 2008). The regional agronomic diagnosis is based on an on-farm assessment and comparison of variables including management practices and environmental conditions (soil and climate). The methodology combines the use of data which can be modelled (e.g., to estimate potential yield) or assessed (field measurements or surveys), and analysis methods like regression combined with principal component analysis (PCA) (Chopin and Blazy 2013; Jagoret et al. 2017). Whereas this methodology has mainly be focused on annual crops (e.g., Valantin-Morison et al. 2007) and, to a lesser extent, on short-life perennials like banana (Chopin and Blazy 2013), few studies used regional agronomic diagnosis methodology to compare perennial-based cropping systems. In a study on cocoa yields in agroforests, Jagoret et al. (2017) highlighted the effects of tree basal area and rate of unproductive adults on the cocoa yield.

One main issue that hampers the use of regional agronomic diagnosis for perennial-based cropping systems is the difficulty to define and assess orchard performances especially in complex systems (Jagoret et al. 2017). Indeed, different ways exist to evaluate the orchard productivity performances. The orchard yield can be measured by summing up the production of individual trees (in number or weight of fruits) in a defined area. However, the sole yield becomes inappropriate when comparing the outputs of an agroforestry systems. Indeed, due to the high cultivated diversity in agroforestry systems, the weight of fruits of one tree species constitutes only a part of the total output per hectare the grower can expect (Malézieux et al. 2009). The fruit load is also an intermediate variable to characterize the orchard productivity. The fruit load of a tree is commonly measured in physiology as a ratio between the number of fruits (or biomass of fruits) and leaves area at the branch level (Léchaudel et al. 2002). Hence the fruit load can be defined at the tree or orchard scale when dividing the fruit production by biomass variables (e.g., tree crown volume or orchard canopy volume).

In this study we decipher the main factors of mango-based orchards productivity in a given production region: the Niayes region in Senegal. For this purpose, the productivity performance, structure, management, and environmental conditions were described for each orchard. Three variables of the orchard productivity performances have been measured: the orchard yield (in $t.ha^{-1}$), the tree production (in kg), and the fruit load (in $kg.m^{-2}$). We adapted the regional agronomic diagnosis method to analyse the performances of three main cropping systems in the area.

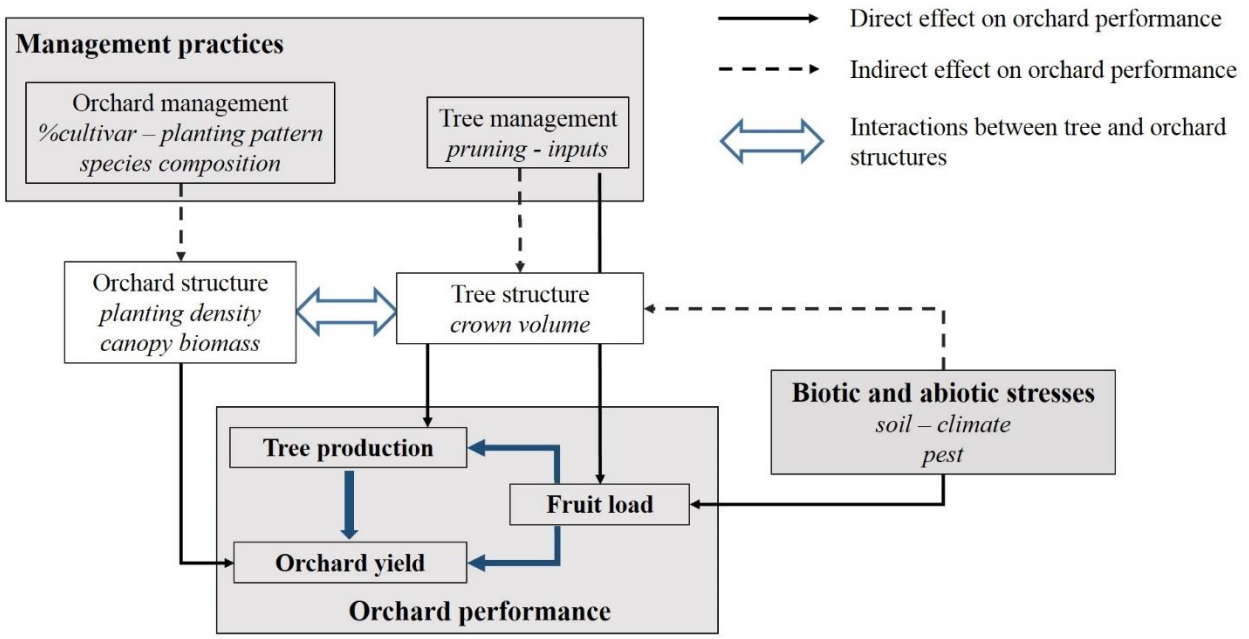


Fig. 71. Summary of the main interactions between management practices (orchard and tree scales), biotic and abiotic stresses (environment), and orchard productivity performances (fruit load, tree production and orchard yield).

III.2. Materials and methods

a. Study sites : mango-based orchards in the Niayes area, Senegal

The Niayes region is part of the Sudano-Sahelian zone and is characterized by an unimodal rainfall from July to September (mean rainfall of 430 mm per year) and a relatively fresh and humid climate generated by the proximity to the Atlantic Ocean (with monthly mean temperatures ranging from 16 °C to 35 °C). The landscape is characterized by a succession of dunes, with sandy and non-lessive soils, and depressions, with clay and peaty soils (Fare et al. 2017). In these topographic depressions the groundwater comes up to the surface, ensuring accessibility to the resource for irrigation. Combined with the favourable climate, these conditions make this region one of the major fruit production areas in Senegal.

Twenty-nine mango-based orchards have been sampled in a 500 km² zone of the Niayes region and located between 14°4’ to 15°0’ N and 16°6’ to 17°2’ W. These orchards were selected to be spatially distributed in the studied area and to depict the three main mango-based cropping systems identified in the region (Grechi et al. 2013). The typology classified orchards according to their crop composition (tree species and/or annual crops) and mango management practices. The first type (n = 9), called “extensive”, corresponds to smallholder orchards mainly composed of mango trees planted on bare ground with no or a low level of management practices (no pruning nor irrigation and chemicals). In this system, the mango trees were mainly of ‘Kent’ cultivar (Rey et al. 2007). This type displayed a high variability in mango tree structure

(i.e., crown volume) and production (Sarron et al., sub.). The second type (n = 10), called "diversified", corresponds to smallholder diversified orchards composed of mango trees of different cultivars along with other fruit tree species (citrus, cashew, papaya, etc.) and occasionally with market gardening underneath the trees. Some highly diversified orchards of this type corresponded to a multi-strata agroforestry system with mango trees in the upper layer, other trees in the intermediate layer, and annual crops on the ground. In the diversified system, the levels of management practices on mango trees were intermediate, and most growers pruned and fertilized trees occasionally, while trees were irrigated in few orchards. The third system (n = 10), called "intensive", corresponds to the monospecific commercial-based orchards composed of one or two export cultivars ('Kent' and 'Keitt', Rey et al. 2007) with a high level of external inputs and homogeneous planting patterns (Sarron et al., sub.). The productive areas of orchards (i.e., the orchard area without building elements and where crops can be grown) was of 1.3, 2.9, and 2.7 ha in average for the extensive, diversified, and intensive systems respectively. The average mango planting densities were 147.2, 122.4, and 226.2 trees.ha^{-1} for the extensive, diversified, and intensive systems, respectively.

b. Equations of the orchard productivity performances

The yield of an orchard Y is commonly defined as the weight of fruit harvested per unit of area (in t.ha^{-1}) and is a key output of orchard performance (Menzel and Le Lagadec 2017). Considering trees as major components of the orchard, the yield can be expressed as the sum of the production of individual trees divided by the orchard area:

$$Y = \frac{\sum_{j=1}^{N} P_j}{A} \times 10^3 \quad (18)$$

where: Y is the yield of the orchard in t.ha^{-1}; N is the number of trees in the orchard; P_j is the fruit production of the j-th tree in kg; and A is the productive area in ha. However, if the individual tree production P_j of all trees is not accessible, then the Eq. 18 of orchard yield can be simplified as follows:

$$\hat{Y} = \bar{P} \times d \times 10^3 \quad (19)$$

with:

$$d = \frac{N}{A} \quad (20)$$

where: Y is the orchard yield approximated by the average tree production in kg.ha^{-1}; $\bar{P}$ is the average tree production of the orchard in kg; d is the tree planting density in trees.ha^{-1}. As the actual production of all trees may be too fastidious and costly to measure, the average tree production $\bar{P}$ can be estimated using different sampling strategies (Anderson et al. 2018; Sarron et al., sub.).

Another way to express the yield is to assess the average fruit load *fl* of the orchard. While fruit load is commonly studied at the tree scale (Léchaudel et al. 2002), in this study we computed the orchard fruit load as the weight of fruit produced per unit of productive biomass. In the case of mango tree, the fruits are mostly produced on the periphery of the canopy due to terminal flowering and concentration of the vegetative growth, flowering, and fruiting at the distal part of the growth unit (Normand et al. 2009). Hence,

we acknowledged the mango crown silhouette area (in m²) as best definer of the productive biomass available for fruit production. Then, we set the average orchard fruit load as the orchard yield divided by the estimated total crown silhouette area of the orchard (i.e., the sum of all mango crown silhouette area per ha) expressed in weight of fruit per unit of canopy area (kg.m^{-2}). Thus:

$$\hat{Y} = fl \times A_o \quad (21)$$

where: fl is the orchard fruit load in kg.m^{-2} of canopy; A_o is the orchard crown area, i.e., the total crown area of mango per ha, in m^2.ha^{-1}. Thus, the average tree production P can be expressed using the fruit load:

$$\bar{P} = \frac{\hat{Y}}{d} = fl \times A_m \quad (22)$$

where: A_m is the average mango tree crown area in the orchard in m^2.tree^{-1}. The orchard fruit load is a variable that reflects the productivity of the orchard per unit of biomass. The equations (19) and (21) highlighted three variables related to orchard productivity performance: i) the orchard yield Y_p; ii) the average tree production $\overline{P}$; and iii) the orchard fruit load fl. In addition, these equations show the importance of the variables related to mango biomass at the tree and orchard scales: the mango tree crown area A_m, the planting density d, and the orchard crown area A_o.

c. Data collection

Data on drivers of yields were collected in the 29 orchards of the sample. In total, 55 variables (Table 14) were used to describe the production performances (n = 3), mango biomass (n = 3), orchard structure (n = 32), management practices (n = 5), and environment (n = 12) of each orchard. Yields were assessed during 2017 and 2018 using a methodology described in Sarron et al. (2018) and summarized below.

Orchard land cover and yield estimation

Orchard structure and yields were assessed using an unmanned aerial vehicle (UAV), and a procedure that combines photogrammetry and geographic object-based image analysis (GEOBIA) described in Sarron et al. (2018). Briefly, each orchard was overflown using an UAV (DJI Mavic Pro, DJI Inc, China) to obtain a very high resolution orthomosaic (1.30 cm.pixel^{-1}). The photogrammetry procedure computed a 3D point cloud from which a Canopy Height Model (CHM) was extracted to measure the height of each trees. Using the orthomosaic and the CHM, the GEOBIA step allowed to delineate and classify each object in 11 classes including annual crops and 7 classes for tree species: mango, citrus (*Citrus sinensis L.*, *Citrus reticulate*, *Citrus paradisi*), cashew (*Anacardium occidentale*), papaya (*Carica papaya*), palm tree (*Eleis guineensis*), and other tree species (*Senegalia Senegal*, *Manilkara zapota*, or *Annona muricata*). Thus, by combining this land cover map, with trees accurately delineated, and the CHM, three structure parameters of each tree were calculated: the tree height, crown area, and crown volume. These variables, combined with a load index, which takes into account year and management effects, were implemented as independent variables in models for accurate estimation of the orchard mango yield. Models reached satisfying accuracies with R^2 greater than 0.77 and RMSE% ranging from 20% to 29% when evaluated with the measured production of 60 validation trees (Sarron et al. 2018).

Table 14. List of variables for performance, orchard structure (mango biomass, composition, and configuration), management practices, and environment (yearly and site specific).

Category of variable	Type of variables (55 variables)	Code	Unit
Orchard performance	Orchard yield (n = 1)	*Y*	$t.ha^{-1}$
	Fruit load (n = 1)	*fl*	$kg.m^{-3}$
	Tree production (n = 1)	*P*	kg
Mango biomass	Mango tree crown area (n= 1)	A_m	m^2
	Orchard crown area (n = 1)	A_o	$m^2.ha^{-1}$
	Planting density (n = 1)	*d*	
Orchard composition	Planting density (other tree, n =3)	*dens*	$trees.ha^{-1}$
	Volume density (n = 4)	*vol*	$m^3.ha^{-1}$
	Tree proportion (n = 4)	*prop*	%
	Volume proportion (n = 3)	*prop_vol*	%
	Orchard age (n = 1)	*age*	
	Annual crops proportion (n = 1)	*prop_annual*	%
	Diversity index (n = 7)	*tree_sp* (# of tree species) *mango_cv* (# of mango cultivars) *SHEI, SHDI* (Shannon's Index) *SIEI, SIDI* (Simpson's Index) *PR* (patch richness)	
Orchard configuration	Tree aggregation (n = 7)	*AI* (aggregation index) *PROX** (proximity) *CLUMPY* (mango clumpiness)	
	Mango tree shape (n = 2)	*CIRCLE**(circumscribing circle)	
Management practices	Levels of input (n = 5)	*prun*; *phyto*; *fert* *irr* (mango and citrus)	categorical
Environment (yearly temperature)	Hours of temperature (n = 3)	*flw_degree*; *Tstress*; *GDH*	hour
	Temperature (n = 3)	*Tmean*; *Tmin*; *Tmax*	°C
Environment (site-specific)	Climate (WordClim 2) (n =3)	*Tmean_WC* (temperature)	°C
		srad_WC (solar radiation)	$kJ\ m^{-2}\ day^{-1}$
		prec_WC (precipitation)	mm
	Soil content (n = 3)	*soil_S* (sand)	$kg.kg^{-1}$
		soil_O (organic)	$g.kg^{-1}$
		soil_N (nitrogen)	$g.kg^{-1}$

Using this methodology, the total fruit produced per orchard (in kg), computed as the sum of the production of each mango tree, was assessed for 2017 and 2018. The orchard productive area (i.e., the orchard area without building elements and where crops can be grown) was deduced from the land cover map and allowed to compute the orchard yield *Y*. In addition, the land cover map allowed to identify each mango tree individually in order to measure the mango planting density *d*. Using Eq. 19, the average tree production *P* was computed. Both land cover and CHM gave information on the tree height (*h*) and projected crown (i.e., horizontal crown seen from above). Then, the area of the crown silhouette of each mango tree was computed by fitting a cylinder model ($A = \pi * h * D$, with *D* the diameter of the projected crown). We deduced the average tree crown area A_m and the total crown area A_o of each orchard. The latter was used to compute the fruit load *fl* using Eq. 21.

Metrics of orchard structure

Several variables describing the orchard structure have been computed using the land cover map and the CHM. These variables depicted: i) the mango biomass (d, Ao, and Am previously described); ii) the orchard other tree density and plant species diversity (or composition, n = 23); and iii) the orchard planting arrangement (or configuration, n = 9). As some species were less abundant or anecdotic in all the trees studied, tree species were grouped into 'mango', 'citrus', and 'other tree'. First, 7 variables of orchard tree density (i.e., planting density and volume of biomass) were directly issued from the UAV outputs and expressed the density (*dens*) of citrus, and other tree which in number of individuals (in trees.ha^{-1}) or by volume of biomass per ha (in m^3.ha^{-1}). Second, the 16 variables of orchard composition included the proportion in volume and in number of individuals of each tree group in the total trees. The proportion of 'Kent' mango trees over the other cultivars was added (% of 'Kent' tree). In addition, the proportion of annual crop was computed as the ratio between area of annual crops and the productive area. This group of orchard composition variables contained also orchard diversity indices: numbers of tree species, number of mango cultivars, Shannon's and Simpson's diversity (*SHDI*, *SIDI*) and evenness (*SHEI*, *SIEI*) indices, and patch richness (*PR*). Finally, the 9 variables of orchard configuration described the tree spatial arrangement (aggregation, patch density, proximity, and clumpiness) mango tree shape indices were computed at the orchard or mango scales, respectively. This later group of variables, and diversity indices were all computed thanks to FRAGSTATS software (McGarigal et al. 2012).

Management practices

During the year 2017, management practices on mango trees were surveyed in each orchard as some practices might vary among the same cropping system. Information on tree pruning, chemical applications (phytosanitary product, fertilizer), irrigation of mango and citrus trees were investigated with the highest precision possible (date and frequency, quantity of input, tool used, etc.) as probable drivers of orchard performances (*Y*, *P*, and *fl*), tree structure and orchard volume density (Fig. 71). Because these data were obtained with low details in most cases -as most producers do not record their practices- they were summarized into categorical variables describing the levels of application frequency (i.e., none, low, medium, or high), for pruning, phytosanitary products and fertilizers, or type of irrigation system (Table 14). As unknown in most cases, the rootstock cultivars were not investigated. However, the planting year of the orchard was obtained and corresponded to the tree age in intensive orchards (all trees of the same age) and to the age of the oldest trees in smallholder orchards due to several tree replacement that occurred.

Weather and soil data

The orchard environmental conditions were described using climatic and soil data. In order to distinguish the effect of climate on the yearly production from the effect of long-term climatic conditions (i.e, site-specific effect) on the tree growth and development, two types of climatic data were used. To characterize the effect of 2017 and 2018 weather on the yearly yields, temperature loggers (HOBO U23 Pro v2, Onset Computer Corporation, USA) placed in each orchard at 2 m high under shelter recorded air temperature every 10 minutes during 2017 and 2018. Then, temperature (T) was summarized at hour interval (mean, minimum and maximum) from February to October of each year to only integrate the period of the mango reproductive stage in the Niayes region (i.e., from flowering in February to late harvest in October). Six variables of yearly temperature were used: *Tmean*, *Tmin*, and *Tmax* as the mean, minimum, and maximum temperature during the reproductive cycle; *flw_degree* as the cumulative degree hours stimulating flowering (i.e., number of hours with T <18°C day and <10°C night, Schaffer et al. 2009); *Tstress* as the cumulative degree hours of temperature stress (T > 40°C); *GDH* as the cumulative degree hours stimulating growth (Tbase = 18°C, Schaffer et al. 2009).

To depict the long-term climate of each orchard (i.e., the climatic site effect), we used climatic variables (temperature, rain, and solar radiation) derived from 1-km resolution WorldClim 2 database (Fick and Hijmans 2017) for 1970 – 2000. These climatic variables were site-specific while the yearly variables on temperature varied from sites and years. Soil variables have been computed from the 'AfSoilGrids250m' map of soil properties for Africa at 250 m resolution (Hengl et al. 2015) available at ISRIC soil data hub (http://www.isric.org/data/AfSoilGrids250m). From this soil dataset, three soil variables (clay fraction, carbon organic, and N contents) were used to describe orchard soil properties.

d. Statistical analysis

A comparison of the three types of cropping systems was performed on the criteria of orchard productivity performances, structure and composition. As most variables presented a logarithmic distribution, we used non-parametric Kruskal-Wallis tests at a 95% confidence level. The first step in the analysis of drivers of orchard yields consisted of characterizing the structure of relationships between orchard performances variables (*Y*, *P*, and *fl*) and mango biomass variables (*d*, A_m, and A_o). For this purpose, we performed a principal component analysis (PCA) to display the correlation of orchard yield *Y* with other output variables. In addition, correlation between each pair of variables were calculated using the Spearman's *r* coefficient. Effect of the year on *Y* was studied using the non-parametric Wilcoxon signed-rank test to compare paired difference of yields of the same orchard with the yields variation of the sample. The second step, involved a factor analysis for mixed data (FAMD) dedicated to explore the entire dataset of orchard descriptors (n = 55) which included quantitative and qualitative variables (Pagès 2004). FAMD uses the same principles as the PCA while taking into account both continuous and categorical variables in the analysis. The FAMD explicated the correlations between all variables and the ones that described best the orchard sample. Finally, from this set of best variables, one or two variables of each groups of descriptors depicting orchard structure, management practices, and environment (soil and climate) were selected. Relationships between orchard performance (*Y*, *fl*, and *P*) and the determinants were determined using multiple linear regression with backward stepwise variable selection (Jagoret et al. 2017). All statistical tests and component analysis (PCA, FAMD) were performed using R and *FactoMineR* R package (Lê et al. 2008).

Table 15. Features (mean and SD) of the three main mango-based cropping systems in the Niayes region, Senegal

	Extensive		Diversified		Intensive	
	mean	SD	mean	SD	mean	SD
Productive area	1.32	0.57	2.92	3.78	2.7	0.97
Y	4.29b	2.69	7.05a	3.27	13.87a	10.50
fl	0.95b	0.45	1.56a	0.60	1.97a	1.09
P	31.40b	21.17	77.19a	64.32	65.25ab	30.09
A_m	32.35a	11.70	50.99a	31.98	37.79a	18.93
A_o	4,423b	1,618	4,607b	1254	7,486a	3,945
d	147.19b	54.4	122.39b	58.06	226.20a	115.85
Total tree planting density	185.88a	68.86	223.50a	91.91	227.32a	114.8
% of mango	79.67b	7.81	58.19b	22.04	99.13a	2.66
% of citrus	11.28b	9.72	33.14a	22.92	0.71c	2.21
% of other tree	9.05a	4.31	8.66a	5.56	0.15b	0.46
% of annual crop (in the total area)	2.50b	4.61	12.10a	9.25	0.00c	0.00
% Kent tree (in the mango area)	69.90a	18.91	43.26b	26.83	79.17a	34.04
Shannon diversity index (*SHDI*)	1.01b	0.31	1.19a	0.22	0.03c	0.10
Mango clumpiness (*CLUMPY*)	0.95ab	0.02	0.96a	0.01	0.93b	0.02

Table 16. Spearman's rho correlation between orchard productivity performances and mango biomass drivers.

	fl	*P*	*d*	A_o	A_m
Y	0.74***	0.67***	0.54***	0.58***	0.06ns
fl		0.46***	0.39**	-0.03ns	-0.29*
P			-0.14ns	0.48***	0.63***
d				0.34**	-0.58***
A_o					0.46***

III.3. Results

a. Cropping system characterisation

Results on cropping systems are summarized in Table 15. Considering the orchard structure (i.e., species composition and tree configuration), the highest mean mango planting density (226 trees.ha^{-1}) was displayed by the intensive system which is mainly of 'Kent' cultivar (80% of all mango trees). However, when considering all the tree species, the average planting density was equivalent in all systems (about 200 trees.ha^{-1}). The diversified system displayed the highest diversity with a third of citrus trees, 9% of other tree species, and 12% of annual crops (Kruskal-Wallis test p-value < 0.05). Thus, the orchard spatial composition, assessed by the Shannon's diversity index (*SHDI*) which measures the proportion of land occupied by each tree species or annual crops, depended on the cropping system. The diversified system

reached a *SHDI* of 1.19 on average then followed by the extensive system (*SHDI* = 1.01) while the intensive system came afterward with a very low diversity (*SHDI* = 0.03). Mango tree clumpiness was measured with the *CLUMPY* index and was close to 1 for all system meaning that the trees tended to be spatially aggregated (*CLUMPY* = 1) than disaggregated (*CLUMPY* = -1) or randomly distributed (*CLUMPY* = 0). Aggregation seemed to be higher in the diversified orchards (*CLUMPY* = 0.96) than in the other systems.

The extensive orchards showed the lowest *Y* (av. = 4.3 $t.ha^{-1}$), *fl* (av. = 0.95 $kg.m^{-2}$), and *P* (av. = 31.40 kg). The intensive system showed higher performance with *Y* and *fl* at 13.9 $t.ha^{-1}$ and 2 $kg.m^{-2}$ respectively, but appeared not significantly different from the diversified orchards (*Y* = 7.0 $t.ha^{-1}$; *fl* = 1.6 $kg.m^{-2}$) due to the high variability of ±10 $t.ha^{-1}$ (SD). However, the tree production *P* was the highest in the diversified orchard and reached 77.20 kg equivalent to the average of intensive system (av. = 65.25 kg) and more than twice higher than in the extensive system.

b. Orchard productivity

Interactions among variables of orchard productivity (*Y*, *P*, and *fl*), mango density (*d*), and crown areas (A_m, A_o) are highlighted on Table 16 (correlations) and Fig. 72 (PCA). *Y* was related to the two other variables of performance, the tree production *P* (r = 0.67) and the fruit load *fl* (r = 0.74). In addition, *Y* was positively correlated with the planting density *d* (r = 0.54) and the orchard crown area A_o (r = 0.58), whereas the average mango tree crown area A_m had no effect. The tree production *P* was positively correlated with *fl*, and crown areas A_m and A_o, whereas *d* had no impact. In addition, an increase in planting density *d* increased the total mango crown area A_o but reduced the average tree crown area A_m (see Discussion).

The Wilcoxon's rank test for comparing both study years reached p-value < 0.01 for all performance variables meaning that orchard's *Y*, *P*, and *fl* were significantly different between 2018 and 2017. Individuals repartition on the PCA in function of years confirmed this result as both years looked separated on axis 2 (Fig. 72). Orchard productivities were lower in 2018 than in 2017 (Y dropped by 37% in 2018 compared to 2017). Only two intensive orchards displayed an increase of yield in 2018 compared to 2017. When testing the effect of the cropping system on the relative difference of *Y* between years (ANOVA test), no significant effect on the decrease was found (p-value = 0.19).

c. Drivers of orchard yields

Results of the FAMD showed that four groups of quantitative variables appeared on axes 1 to 3 (Fig. 73). On the first axis, a first group represented diversity variables opposing orchards with high values of diversity indices (e.g., *PR*, *SHDI*, *tree_sp*) to orchards with high presence of mango trees (high values of *prop_mango*, *prop_vol_mango*, *d* and *prop_kent*). A second set of variables regrouped tree aggregation variables on the second axis (*AI*, *CLUMPY*, *PROX*) which were correlated positively to *P* and negatively to *fl*. A third group represented the variables of tree biomass with A_o, A_m, *age*, *vol_mango* and *vol_tree*. This group was more spread out as A_o seemed less correlated to mango tree *age*. However A_o, A_m, and *age* interacted all with the volumes of mango orchard (*vol_mango)* and total trees (*vol_tree)*. In addition, orchards with high values of A_m and *age* seemed to be highly aggregated (correlation with aggregation indices *AI* and *CLUMPY*) and consequently to be correlated positively with *P* but negatively with *fl*. These results validated the correlation coefficients between A_m, *P* and *fl* (Table 16). On the opposite, the orchard yield *Y* and fruit load *fl* showed strong correlation with the diversity of the orchard: while they were opposed to the diversity indices, they were positively correlated to the proportion of mango. In addition, *Y* were related with A_o, as shown also in

Table 13, and with *vol_mango*. A fourth group depicting climatic variables clearly appeared on axis 3 of the FAMD (Fig. 73). Nonetheless, despite the strong correlation within these group of variables, no explicit correlation of this group with orchard productivity was evidenced. Finally, the long-term precipitation (*prec_WC*) appeared separately on the representation of axes 1 vs. 3 and was correlated to *fl*.

The representation of qualitative variables (i.e., management practices) showed clear distinction between the intensive orchards, related to high level of input and drip irrigation, and the two other systems. Indeed, the high level of management practices (fertilisation, pruning and phytosanitary product) was clearly distinct from the medium and no application levels. Thus, despite a significant distinction in plant species diversity (Table 15), extensive and diversified orchards of the sample couldn't be separated based on their management practices (Fig. 73c). The use of jet/can irrigation system appeared separately correlated to the second axis and was shared by orchards of different cropping system. Indeed, this practice was used by one diversified and two intensive orchards.

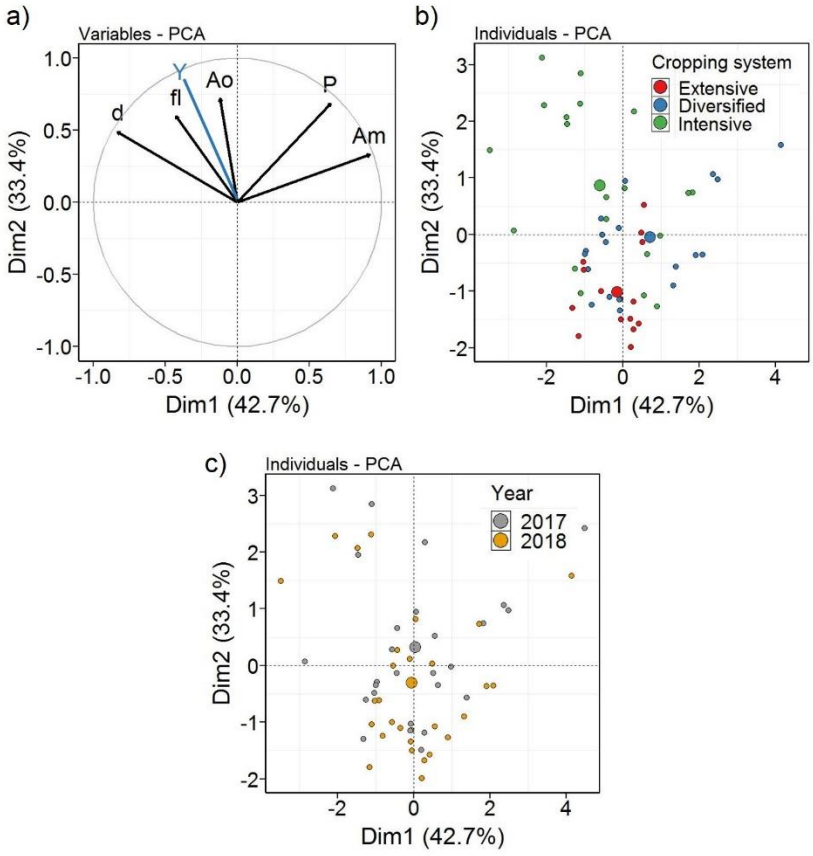


Fig. 72. Results of the Principal Component Analysis (PCA) with productivity performances and mango biomass variables: (a) quantitative variables with orchard yield (*Y*) projected as a complementary variable ; (b) studied orchards with colour reflecting the cropping systems ; and (c) studied orchards with the colour reflecting the year.

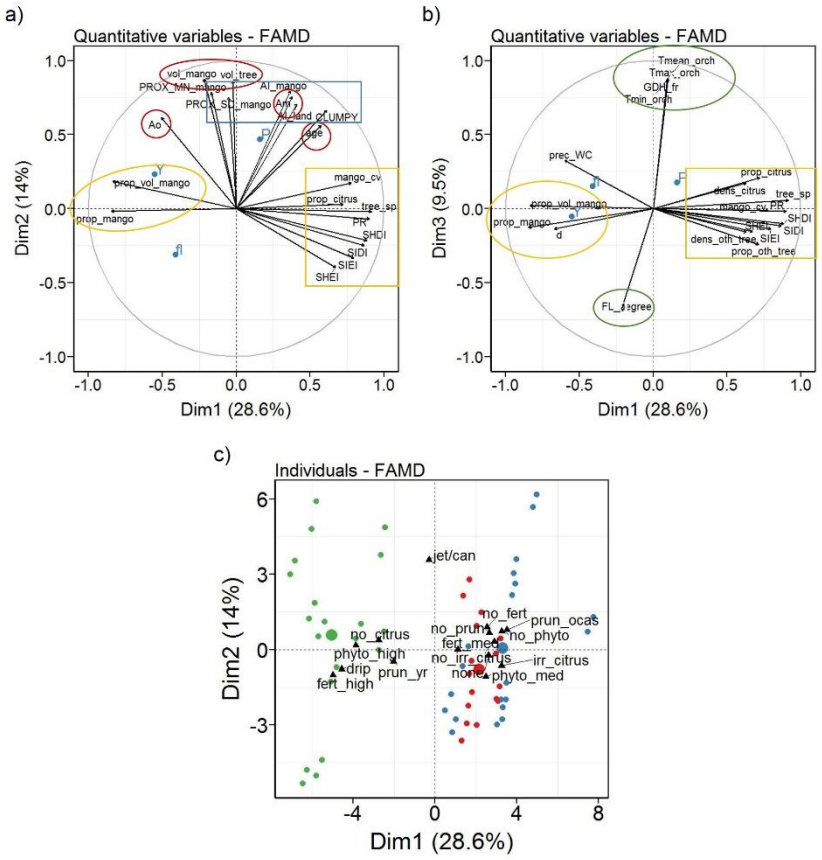


Fig. 73. Results of the factor analysis for mixed data (FAMD) with orchard structure, management practices and environment variable: (a) axis 1 vs 2 and (b) axis 1 vs. 3 of quantitative variables with orchard productivity variables (*Y*, *P*, and *fl*) projected as complementary variables ; (c) studied orchards with colour reflecting the cropping systems.

Using the PCA and FAMD results, we selected nine variables that were well correlated with each orchard productivity and represented each group of drivers (orchard structure, management practices and environment). Thus, this first set of variables were: *d*, A_m, A_o, *vol_mango*, *SHDI*, *CLUMPY*, *Tmean_orch*, *prec_WC*, *Year*, and *fert*. Fertilisation level (*fert*) were selected to represent management practices as they were all correlated (Fig. 73c). All models showed that *Y* (R^2 = 0.80), *P* (R^2 = 0.73), and *fl* (R^2 = 0.53) were greatly impacted by year effect confirming that orchard productivity performances were lower in 2018 than in 2017 (Table 17). In addition, effect of three important variables were evidenced: *fert*, *SHDI*, and *CLUMPY*. The 'high' fertilisation level had also a strong positive impact on orchard productivity performances. To a lesser extent, plant diversity (*SHDI)* positively impacted *Y* and *P*, whereas it had no impact on *fl* even if p-value (0.08) was close to 0.05 (Table 17). An increase in mango aggregation (*CLUMPY*) decreased both *P* and *fl* and might be explained by competition effect between trees. Finally, both *Y* and *P* were positively impacted by the mango biomass at the orchard (A_o, *d*) and tree (A_m) scales respectively.

Table 17. Regression statistics for orchard yield (*Y*), tree production (*P*), and fruit load (*fl*) with variable selected using a stepwise procedure.

Variable	Estimates	p-value	Model R^2
1) Orchard yield *Y*			0.80
d	0.27	0.004**	
A_o	0.61	<0.001***	
SHDI	0.19	0.016*	
Fert [none]	-0.29	0.057	
Fert [high]	0.62	0.003**	
Year [2018]	-0.40	0.002**	
2) Tree production *P*			0.73
A_m	0.98	<0.001***	
SHDI	0.19	0.04*	
CLUMPY	-0.20	ns	
Fert [none]	-0.46	0.008**	
Fert [high]	0.53	0.011*	
Year [2018]	-0.54	<0.001***	
3) Fruit load *fl*			0.53
SHDI	0.19	0.083	
CLUMPY	-0.29	0.014*	
Fert [none]	-0.50	0.030*	
Fert [high]	1.09	<0.001***	
Year [2018]	-0.60	0.002**	

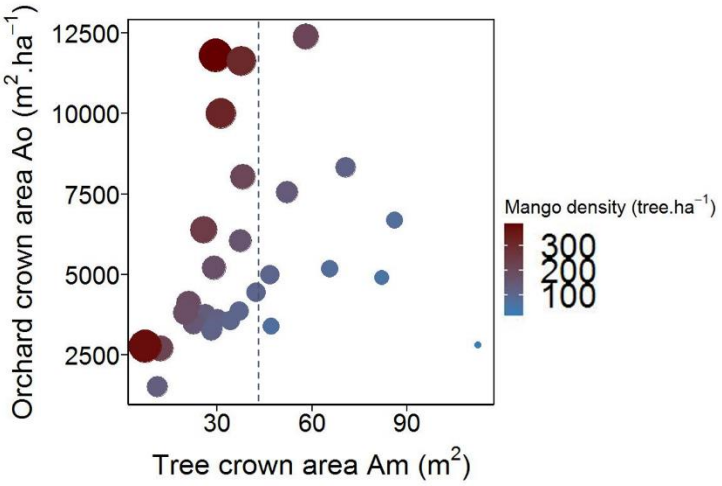


Fig. 74. Orchard crown area (A_o) in function of the tree crown area (A_m) with orchard density (*d*) displayed. The vertical dotted line indicates the maximum A_m attain by the orchard with the highest density (d > 200 trees.ha^{-1})

III.4. Discussion

a. Effect of mango biomass on tree production and orchard yield

In order to analyse drivers of orchard yields in mango-based orchards, interactions between variables of orchard productivity (*Y*, *fl*, and *P*) and the key drivers of mango biomass (*d*, A_m, and A_o) were evidenced. These drivers were chosen as they are first limiting factors of the light interception in the orchard (Connor et al. 2014; Menzel and Le Lagadec 2017). Despite no significant differences between diversified and intensive orchards, the highest yields of the study were observed in the intensive orchards that were well established (i.e., more than 6 years old). Yields of intensive orchards reached 15.5 $t.ha^{-1}$ on average when young intensive orchards are removed from the analysis (representing 2 of the 10 intensive orchards),

Intensive orchards displayed high *Y*, high crown area per hectare A_o and high mango tree densities d (Fig. 72). Interestingly, an increase in *d* or A_m both increased A_o values (positive correlations, Table 16) whereas the relationship between *d* and A_m was negative. At a certain level of planting density and tree development (age), A_m stopped to increase. This level of planting density could be estimated at around 200-250 $trees.ha^{-1}$ from which orchard with highest value of *d* had lower A_m, below 50 m^2 (Fig. 74). Thus, at high density, the orchard mango biomass is mainly driven by tree density than tree structure itself. Stagnation of A_m at high densities may be due to competitions between trees, and tree pruning practices (Connor et al. 2014). In the study, high density orchards were intensive orchards in which pruning practices are used every year to maintain the tree crown at a certain volume and optimize light interception (Sharma et al. 2006; Menzel and Le Lagadec 2017). Nonetheless, some low density orchards (< 80 $trees.ha^{-1}$) had A_o as high as in high density orchards due to A_m that might increase up to 50% (Fig. 74). Consequently, at low density, the drivers of orchard biomass are both tree density and the individual tree biomass. Finally, these results clearly evidenced that mango tree density is a key driver of individual tree crown and orchard canopy biomass in mango orchards.

Orchards with high *d* (between 300 and 400 $trees.ha^{-1}$) reached the highest A_o (between 10,000 and 12,500 $m^2.ha^{-1}$) and the highest Y values ranging from 25 to 35 $t.ha^{-1}$. On the opposite, orchards with lowest densities (d < 80 $trees.ha^{-1}$) showed A_m higher than 60 m² and consequently the highest *P* of the sample (110.8 kg on average). These results were concomitant with several studies, reviewed by Menzel et Le Lagadec (2017), which examined interactions between orchard biomass, planting density and productivity at the tree and orchard scale. Many studies showed positive relationships between individual tree biomass and tree production (Menzel et Le Lagadec 2017). Meanwhile an increase of planting density reduces tree biomass and productivity, the positive effect of this increase of planting density on the total orchard biomass and yield have been demonstrated. For instance, Sousa et al. (2012) showed negative relationship between several tree structure variable (such as crown height, crown diameter, and trunk diameter) and planting density. In addition, this study highlighted a continuous increase of mango yields with up to about 550 $trees.ha^{-1}$ that then decrease (Sousa et al. 2012). However, due to lack of consistency between studies - especially with the tree age, cultivar, and rootstock used- there is no consensus on the optimal planting density in mango orchard which vary from 200 to 4,000 $trees.ha^{-1}$ (Menzel and Le Lagadec 2017). In the present study, orchards with the highest *P* belonged to the diversified systems (Table 2) whereas most studies on mango productivity focused on monospecific orchards (Menzel and Le Lagadec 2017). However, this diversified system could be compare to other agroforestry systems in which the objectives of the grower were not to maximize the yield but rather to increase the production per tree (Jagoret et al. 2017). Hence, in

our study area, growers of diversified system seek to reduce mango planting density to occupy the area with other tree (mainly citrus) or annual crops while maintaining the mango tree production.

b. Effects of management practices and orchard structure on orchard productivity

Orchard structure and management practices showed distinct effects on *Y* and *P*. Despite a year effect was clearly evidenced, no effect of temperature was evidenced on *Y*, *P* or *fl*. Hence, the reduction of performances between 2018 and 2017 may be caused by other environmental factors such as rainfall and wind (Schaffer et al. 2009) that were not measured in this study. Nonetheless, models showed a positive effect of management practices (represented by fertilisation) and *SHDI* on *Y* and *P*. The effect of *SHDI* might be explained by the intermediate level of *Y* attained by the diversified orchards compared to the extensive orchards. Despite both systems displayed the same mango density (around 130 trees.ha^{-1}) and mango canopy biomass (no difference in A_o, Table 2), mango trees represented 80% of all trees in the extensive system (*SHDI* = 1.0) whereas diversified system showed higher diversity (*SHDI* = 1.20) (Table 2). In addition, most of extensive orchards received no input (e.g., 75% of growers applied no fertilizers), while diversified orchards generally received medium levels of inputs (e.g., 84% of growers applied fertilizers at least once a year). Thus, *SHDI* and management practices might be correlated in this study. However the fact that both factors had an effect on *Y*, *P* and *fl* supposed a little effect of plant diversity on these performances. When studying effects *SHDI*, *fert*, and their interaction of exclusively on the diversified orchards, we evidenced a positive and significant of *SHDI* on *P* (estimate = 3.40, p-value < 0.001) and on *fl* (estimate = 1.06, p-value = 0.04). Hence, within diversified system, increase of plant diversity might also increase the orchard productivity. One possible reason explaining this effect is that the presence of citrus and/or annual crops, which are irrigated, might increase the water availability to mango trees (Rathore et al. 2013). Nonetheless, further investigations on the effect of plant diversity on the orchard performances are required in the Niayes area. Such investigations should include a better quantitative assessment of management practices in addition with practices on other plant species with a priority on annual and citrus trees.

c. Using fruit load and regional agronomic diagnosis to compare mango-based cropping system

The fruit load (*fl*) was introduced in the study to explain the orchard productivity by unit of mango biomass. The correlation between *fl* and *d* corresponded to an indirect effect of the cropping system as all intensive orchards displayed higher mango densities (*d* = 226 trees.ha^{-1}, Table 15) but also higher level of inputs that might increase the number of fruits produced per unit of crown area. The negative correlation between *fl* and A_m was explained by the presence of an outlier orchard which corresponded to a young intensive grove of 5 years old with the lowest crown area (7 m^2 compared to 38 m^2 on average for the intensive system) of the sample. The small trees conducted to the highest *fl* values of the sample for this orchard (3.7 kg.m^{-2} on average on both years). By removing this outlier, the correlation between A_m and *fl* appeared non-significant (Spearman p-value = 0.13).

Thus, we considered that the fruit load *fl* was not correlated to mango biomass (Ao and A_m), whereas positive correlations with *P* and *Y* were evidenced. By reflecting the productivity for a fixed amount of biomass, *fl* permits to study the effects of exogenous factors on orchard performances. In this study, the use of accurate assessment of tree structure, based on UAV imagery (Sarron et al. 2018), strengthened the reliability of the

fl measurement. Then, we evidenced that such UAV-based workflow could be useful for orchard characterization especially agroforestry orchards (Pádua et al. 2017).

III.5. Conclusion

The methodology of regional diagnosis analysis allowed us to highlight the main drivers of orchard productivity in the Niayes area. Using PCA and FAMD methods followed by regression models, we highlighted the effect of mango biomass, plant diversity and management practices on orchard productivity. The fruit load *fl* was lower in extensive system compared to diversified and intensive system. More specifically, while extensive and diversified systems have similar mango biomasses (equality in A_o, A_m, and *d*), *fl* of diversified is 50% higher than for extensive systems which could be explained by the low level of management practices in this latter system. Hence, for an equivalent level of biomass, diversified and intensive systems both reach similar productivities. The higher yields obtained in the intensive system compared to diversified system (significant differences when young groves are removed, see 4.1) were mainly due to higher planting densities rather than higher fruit loads. Negative correlations between tree planting density and the individual tree biomass and production have been evidenced. On the opposite, the lower density of the diversified system seems compensated by a little increase of biomass at the tree scale conducting in higher tree production *P* as it was also demonstrated in other agroforestry systems. The lower yield in extensive system is explained by a decrease of *fl* that may be due to poor maintenance of trees (i.e., low level of management practices).

These results confirmed that the fruit load is an interesting variable to understand and compare orchard yields. The impossibility in this study to decipher the effects of the management practices originate from the fact that information is hard to obtain even at the farm scale in West Africa. Using categorical variables for management practices allowed to clearly distinguish intensive systems from the two others. However, more effort should be done on the assessment of management practices to better highlight management effects on orchard performances. This study highlighted the interest of using regional agronomic diagnosis to decipher drivers of yields of orchard from various cropping systems. Such methodology is useful to draw research hypothesis especially in regions were fruit trees are of key importance, such as in West Africa.

References

Anderson NT, Underwood JP, Rahman MM, et al (2018) Estimation of fruit load in mango orchards: tree sampling considerations and use of machine vision and satellite imagery. Precis Agric. https://doi.org/10.1007/s11119-018-9614-1

Carletto C, Jolliffe D, Banerjee R (2015) From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. J Dev Stud 51:133–148. https://doi.org/10.1080/00220388.2014.968140

Chopin P, Blazy J-M (2013) Assessment of regional variability in crop yields with spatial autocorrelation: Banana farms and policy implications in Martinique. Agric Ecosyst Environ 181:12–21. https://doi.org/10.1016/j.agee.2013.09.001

Connor DJ, Gómez-del-Campo M, Rousseaux MC, Searles PS (2014) Structure, management and productivity of hedgerow olive orchards: A review. Sci Hortic 169:71–93. https://doi.org/10.1016/j.scienta.2014.02.010

Demestihas C, Plénet D, Génard M, et al (2017) Ecosystem services in orchards. A review. Agron Sustain Dev 37:12. https://doi.org/10.1007/s13593-017-0422-1

Doré T, Clermont-Dauphin C, Crozat Y, et al (2008) Methodological progress in on-farm regional agronomic diagnosis. A review. Agron Sustain Dev 28:151–161. https://doi.org/10.1051/agro:2007031

Doré T, Sebillotte M, Meynard J-M (1997) A diagnostic method for assessing regional variations in crop yield. Agric Syst 54:169–188

FAOSTAT (2017) FAO, Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/faostat/en/#home

Fare Y, Dufumier M, Loloum M, et al (2017) Analysis and Diagnosis of the Agrarian System in the Niayes Region, Northwest Senegal (West Africa). Agriculture 7:59. https://doi.org/10.3390/agriculture7070059

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. Int J Climatol 37:4302–4315. https://doi.org/10.1002/joc.5086

Grechi I, Sane CAB, Diame L, et al (2013) Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68:447–466. https://doi.org/10.1051/fruits/2013094

Hengl T, Heuvelink GBM, Kempen B, et al (2015) Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLOS ONE 10:e0125814. https://doi.org/10.1371/journal.pone.0125814

Jagoret P, Michel I, Ngnogué HT, et al (2017) Structural characteristics determine productivity in complex cocoa agroforestry systems. Agron Sustain Dev 37:. https://doi.org/10.1007/s13593-017-0468-0

Jagoret P, Michel-Dounias I, Malézieux E (2011) Long-term dynamics of cocoa agroforests: a case study in central Cameroon. Agrofor Syst 81:267–278. https://doi.org/10.1007/s10457-010-9368-x

Lê S, Josse J, Husson F, others (2008) FactoMineR: an R package for multivariate analysis. J Stat Softw 25:1–18

Léchaudel M, Genard M, Lescourret F, et al (2002) Leaf-to-fruit ratio affects water and dry-matter content of mango fruit. J Hortic Sci Biotechnol 77:773–777. https://doi.org/10.1080/14620316.2002.11511571

Malézieux E, Crozat Y, Dupraz C, et al (2009) Mixing plant species in cropping systems: concepts, tools and models. A review. Agron Sustain Dev 29:43–62. https://doi.org/10.1051/agro:2007057

McGarigal K, Cushman SA, Ene E (2012) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. University of Massachusetts, Amherst

Menzel CM, Le Lagadec MD (2017) Can the productivity of mango orchards be increased by using high-density plantings? Sci Hortic 219:222–263. https://doi.org/10.1016/j.scienta.2016.11.041

Mukherjee SK, Litz RE (2009) Botany and Importance. In: Litz RE (ed) The mango: botany, production and uses. CABI, Cambridge, USA, pp 1–18

Normand F, Bello AKP, Trottier C, Lauri P-É (2009) Is axis position within tree architecture a determinant of axis morphology, branching, flowering and fruiting? An essay in mango. Ann Bot 103:1325–1336. https://doi.org/10.1093/aob/mcp079

Pádua L, Vanko J, Hruška J, et al (2017) UAS, sensors, and data processing in agroforestry: a review towards practical applications. Int J Remote Sens 38:2349–2391. https://doi.org/10.1080/01431161.2017.1297548

Pagès J (2004) Analyse factorielle de données mixtes. Rev Stat Appliquée 52:93–111

Rathore AC, Saroj PL, Lal H, et al (2013) Performance of mango based agri-horticultural models under rainfed situation of Western Himalaya, India. Agrofor Syst 87:1389–1404. https://doi.org/10.1007/s10457-013-9646-5

Rey J-Y, Diallo TM, Vannière H, et al (2007) The mango in French-speaking West Africa: varieties and varietal composition of the orchards. Fruits 62:57–73. https://doi.org/10.1051/fruits:2006051

Sarron J, Faye É, Malézieux É Integrating within-field spatial heterogeneity to estimate orchard yields: the case of mango-based systems in West Africa. [*submitted*]

Sarron J, Malézieux É, Sané C, Faye É (2018) Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. Remote Sens 10:1900. https://doi.org/10.3390/rs10121900

Schaffer B, Urban L, Lu P, Whiley AW (2009) Ecophysiology. In: Litz RE (ed) The mango: botany, production and uses. CABI, Cambridge, USA, pp 170–209

Sharma RR, Singh R, Singh DB (2006) Influence of pruning intensity on light penetration and leaf physiology in high-density orchards of mango trees. Fruits 61:117–123. https://doi.org/10.1051/fruits:2006010

Sousa CAF de, Cavalcanti MILG, Vasconcelos LFL, et al (2012) "Tommy Atkins" mango trees subjected to high density planting in subhumid tropical climate in northeastern Brazil. Pesqui Agropecuária Bras 47:36–43. https://doi.org/10.1590/S0100-204X2012000100006

Valantin-Morison M, Meynard J-M, Doré T (2007) Effects of crop management and surrounding field environment on insect incidence in organic winter oilseed rape (Brassica napus L.). Crop Prot 26:1108–1120. https://doi.org/10.1016/j.cropro.2006.10.005

WHO (2004) Fruit and Vegetables for Health. Report of a joint FAO/WHO workshop. Kobe, Japan

DISCUSSION ET CONCLUSION

Les objectifs de ma thèse étaient d'estimer et de spatialiser les rendements du manguier afin de mieux comprendre leurs hétérogénéités à différentes échelles : de l'arbre à la région. Pour ce faire, des méthodes nouvelles ont dû être élaborées pour caractériser l'arbre et sa production, ainsi que le verger et sa structure, en utilisant des outils innovants et frugaux (i.e., de faible coût). La première partie de cette discussion va porter sur les avantages et les limites de ces outils utilisés ou développés durant la thèse. En effet, ces outils, principalement basés sur l'analyse d'images, reposent sur des méthodes encore peu utilisées en recherche agronomique (réseaux de neurone, photogrammétrie, etc.). La fiabilité et les perspectives offertes par ces outils seront en particulier discutées dans le contexte de l'agriculture ouest-africaine.

Le principal questionnement méthodologique de la thèse porte sur l'imbrication des différentes échelles d'études. Plusieurs méthodes d'extrapolation du rendement, de l'échelle de l'arbre vers celle du verger, ont été mises au point. Ainsi, dans une deuxième partie, nous discuterons plus particulièrement des méthodes d'échantillonnages et de leurs limites. Bien que les travaux aient principalement porté sur les échelles de l'arbre et du verger, nous aborderons également les moyens qui pourraient être mis en œuvre pour l'estimation et la spatialisation des rendements à l'échelle régionale.

Enfin, dans une dernière partie, nous discuterons des méthodes d'analyse des facteurs d'hétérogénéité du rendement, en revenant sur deux éléments majeurs : i) le concept de 'production gap', qui constitue une première adaptation des 'yields gap' appliquée à l'échelle de l'arbre et ii) la variabilité du rendement intra- et inter-verger, appliquée à la diversité des systèmes de culture rencontrée en Afrique de l'Ouest.

I. Outils pour la caractérisation de l'arbre et du verger

I.1. Estimation de la production de l'arbre par analyse d'image

a. KNN vs. CNN

Les premiers travaux de la thèse ont consisté à mettre au point un outil précis d'estimation de la production de mangues à l'échelle de l'arbre. Nous nous sommes orientés vers la détection de fruits par analyse d'image. Cette technique n'est pas nouvelle en soi (Jiménez et al. 2000), mais l'accélération des puissances de calcul, la diminution du coût des capteurs et l'essor des réseaux de neurones en 2012 ont ouvert de nouvelles opportunités pour des outils performants et faciles d'utilisation (Kamilaris and Prenafeta-Boldú 2018). Deux algorithmes ont ainsi été testés et exploités pour la détection des mangues dans le houppier et reflètent les évolutions récentes apportées par la recherche en informatique. Le premier est basé sur une méthode de classification KNN (pour k-nearest neighbours, voir Chapitre 1 – Partie I et Fig. 75), un algorithme de *machine learning* assez simple à paramétrer et déjà utilisé pour la détection de plusieurs

espèces comme les pommes (Linker et al. 2012) ou les mangues (Qureshi et al. 2016). Nous avons pu démontrer que le KNN atteignait des performances similaires à celles de la littérature (F1-score de 0.71) pour détecter des mangues d'une seule variété ('Kent') issues d'un seul verger intensif et homogène dans la taille et la structure des arbres. Cependant, cet algorithme s'est très vite révélé limité pour détecter les mangues de différentes variétés présentes dans des systèmes de culture différents. Ainsi, lorsque l'algorithme est entrainé et utilisé pour détecter les mangues des trois variétés ('Kent', 'Keitt' et 'Boucodiékhal') et dans trois systèmes de culture (extensif, diversifié et intensif), ses performances diminuent de plus de 30% (F1-score de 0.48). S'ajoute à cette chute de la performance de détection, un temps de calcul très élevé lié à la classification centrée sur le pixel. Pour analyser des images RVB de 24 Mpixel, le KNN va utiliser les informations de chaque bande et également calculer des indices de couleurs et de textures. Le temps de calcul pour une seule image varie donc de 15 à 20 minutes incluant le calcul des indices, la classification, et le post-traitement de l'image (suppression du bruit, de l'arrière-plan, etc.).

Ce premier travail nous a permis de montrer les limites de la classification pixel-centrée pour des contextes aussi variables que le nôtre. Le choix des indices et le post-traitement sont des étapes dont la calibration manuelle conduit à l'augmentation des faux-positifs (mangue non détectée) et des faux-négatifs (autre objet détecté comme une mangue) d'une part et à la suppression de vrais-positifs d'autre part. Par exemple l'étape de post-traitement peut supprimer des mangues occultées (Fig. 76). L'hétérogénéité des mangues est également mise en cause dans l'origine de certaines erreurs. De par sa structure arborescente, le KNN est très sensible à la qualité des données d'entrainement et à la variabilité de l'aspect visuel de l'objet à étudier. Sur une image RVB, l'aspect des mangues varie en termes de couleur et de forme en fonction de la variété, de son stade de maturité, de sa position dans le houppier et de son éclairement (Fig. 15, p. 38). Plusieurs voies d'amélioration pourraient être proposées pour continuer à utiliser le KNN, comme la construction et le choix d'indices plus pertinents, l'augmentation du nombre d'indices, l'amélioration des données d'entrainement ou encore l'augmentation du nombre de classes (ex : une classe par variété) (Linker et al. 2012).

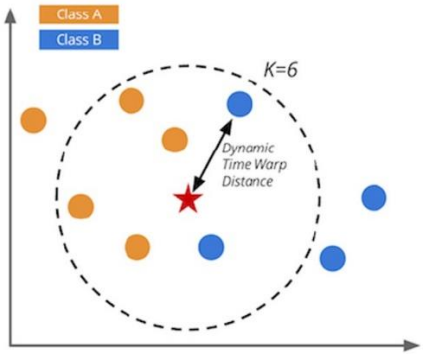


Fig. 75. Classification k-nearest neighbours (KNN). Pour classifier le pixel (étoile), l'algorithme s'intéresse à sa proximité avec les 6 plus proches voisins (k =6) projeté dans une espace à deux indices (source : github.com).

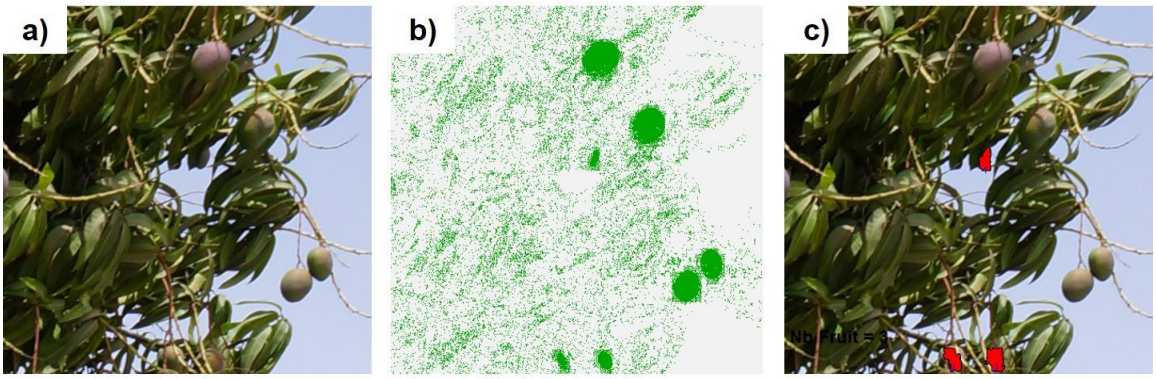
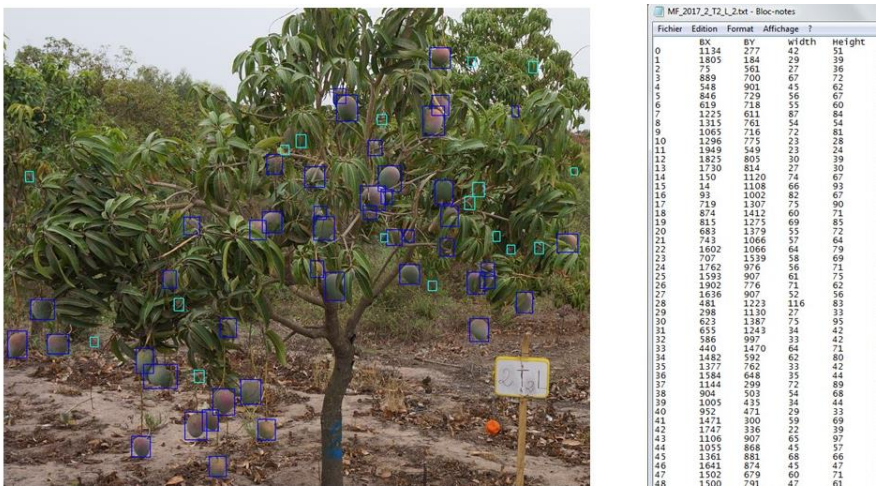


Fig. 76. Exemple montrant la suppression de mangues lors de la phase de post-traitement : (a) est l'image d'origine, (b) est la sortie de la classification et (c) est la sortie finale.



MF_2017_2_T2_L_2.txt - Bloc-notes

Fichier Edition Format Affichage ?

	BX	BY	Width	Height
0	1134	277	42	51
1	1805	184	29	39
2	75	561	27	36
3	889	700	67	72
4	548	901	45	62
5	846	729	56	67
6	619	718	55	60
7	1225	611	87	84
8	1315	761	54	54
9	1065	716	72	81
10	1296	775	23	28
11	1949	549	23	24
12	1825	805	30	39
13	1730	814	27	30
14	150	1120	74	67
15	14	1108	66	93
16	93	1002	82	67
17	719	1307	75	90
18	874	1412	60	71
19	815	1275	69	85
20	683	1379	55	72
21	743	1066	57	64
22	1602	1066	64	79
23	707	1539	58	69
24	1762	976	56	71
25	1593	907	61	75
26	1902	776	71	62
27	1636	907	52	56
28	481	1223	116	83
29	298	1130	27	33
30	623	1387	75	95
31	655	1243	34	42
32	586	997	33	42
33	440	1470	64	71
34	1482	592	62	80
35	1377	762	33	42
36	1584	648	35	44
37	1144	299	72	89
38	904	503	54	68
39	1005	435	34	44
40	952	471	29	33
41	1471	300	59	69
42	1747	336	22	39
43	1106	907	65	97
44	1055	868	45	57
45	1361	881	68	66
46	1641	874	45	47
47	1502	679	60	71
48	1500	791	47	61

Fig. 77. Annotations manuelles des mangues et sortie .txt avec coordonnées et dimension des boîtes englobantes.

Plutôt que de chercher à utiliser ces différents leviers d'amélioration pour le KNN, notre choix s'est porté sur l'utilisation d'un autre classifieur basé sur du *deep learning* (Fig. 28-29, p. 58-59). Ce choix s'est justifié par l'augmentation de l'utilisation des réseaux de neurones dans l'analyse d'image notamment en agriculture (Kamilaris and Prenafeta-Boldú 2018 ; Koirala et al. 2019b) et leur aspect open-source les rendant disponible à la communauté scientifique. Nous avons adopté un réseau de neurones de convolution (CNN, pour Convolutional Neural Network), le Faster-RCNN, dont l'efficacité pour détecter des mangues a été démontrée très récemment (Stein et al. 2016). Le réseau offre de nombreux avantages : il dépasse les autres algorithmes dans l'efficacité de détection des fruits (F1 score > 0.90) quel que soit le milieu et permet un gain considérable en temps de calcul. En effet, si l'entrainement du réseau de neurones sur des images annotées nécessite des ressources considérables (machine de calcul), sa mise en œuvre sur des images est très rapide (image analysée en quelques secondes) et ne nécessite que peu de ressources informatiques. Nous avons pu démontrer que l'efficacité du réseau Faster-RCNN était bien supérieure dans notre cas d'étude que le KNN avec un F1 score de 0.90 quels que soient la variété et le milieu considérés. Ces résultats sont

similaires à ceux de Bargoti et Underwood (2017) qui obtiennent un F1 score de 0.91 dans un verger intensif. Ainsi le réseau détecte plus de 90% des mangues présentent dans l'image.

Bien que les réseaux de neurones offrent de nombreux potentiels pour l'estimation de la production, nous avons pu, au cours de la thèse, explorer les limites de cet outil. En premier lieu, les réseaux de neurones requièrent un nombre important de données d'entrainement. Le CNN est un réseau qui détecte des régions d'intérêt sur l'image, il apprend donc à partir d'annotations sous forme de boîtes englobantes (Fig. 77). Les boîtes englobantes sont créées manuellement par des annotateurs experts qui repèrent chaque fruit sur l'image, l'entourent et exportent le fichier (.txt) contenant les coordonnées des boîtes de l'image et leur taille. Or, le nombre d'annotations nécessaires au réseau n'est pas prévisible à l'avance, il faut donc obtenir un maximum d'annotations dans l'espoir d'en avoir un nombre suffisant décrivant toutes les configurations et aspects possibles du fruit. Le réseau utilise ensuite une partie de ces annotations pour son paramétrage, l'autre partie étant gardée pour l'évaluation (calcul du F1-score). Le temps d'annotation et d'entrainement est donc fortement limitant. Pour entraîner le Faster-RCNN à détecter les mangues dans notre contexte, 10 000 annotations de fruits ont été réalisées par quatre annotateurs. L'entrainement a ensuite nécessité ~12 heures de calcul sur une machine possédant une carte graphique de 8 Go de mémoire vive et 1664 cœurs (NVIDIA Quatro M4000) installé sur un ordinateur à 64 Go de RAM et un disque dur flash. Le temps et la puissance de calcul nécessaire à l'entraînement ne sont donc pas négligeables. Cependant, une fois entraîné, le réseau obtient des résultats très prometteurs et analyse des images de 24 Mpixel en seulement 14 s. Une fois dégradé, il peut même être déployé sur les GPU des smartphones sans trop diminuer ses performances.

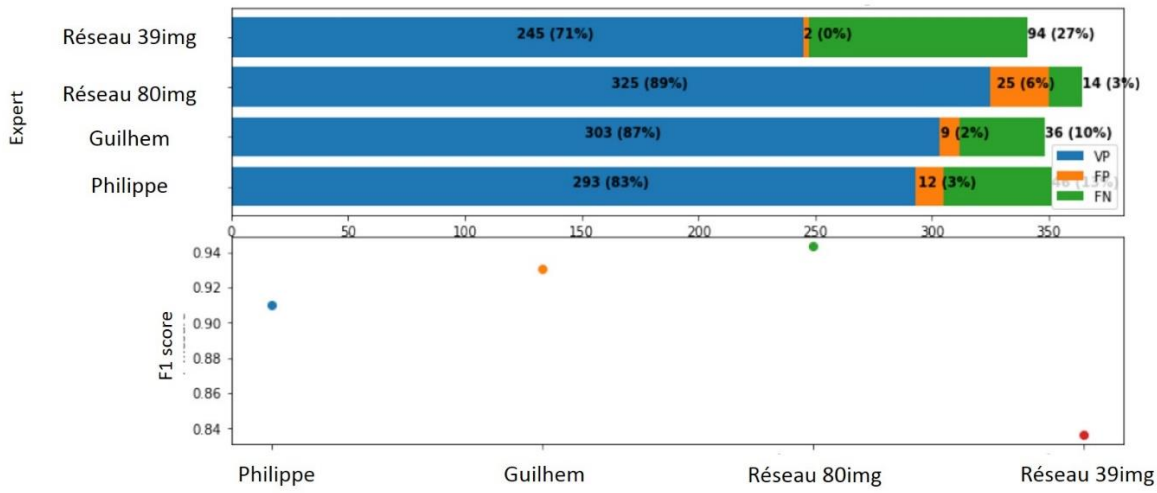


Fig. 78. Haut : Evaluation de variabilité de l'annotation de 2 annotateurs (Philippe et Guilhem) par rapport à un annotateur de référence. Réseau 39img correspond au réseau Faster-RCNN entraîné uniquement sur les images où les annotateurs étaient d'accords. Réseau 80img correspond au réseau entraîné avec toutes les images. VP : vrai positif; FP : faux positif; FN : faux négatif. Bas : F1 score obtenu pour chaque situation (source : G. Barthes, étudiant M1 CIRAD HortSys – AMAP).

Plus que le réseau en lui-même, les données annotées sont une ressource primordiale pour l'utilisation du *deep learning* en agriculture (Koirala et al. 2019b). Or, pour la mangue, et pour tous les fruits en général, la disponibilité et la qualité des données est variable. D'abord, nous avons pu mettre en évidence que l'évaluation du réseau dépend de la définition de l'objet à annoter. Il existe un effet 'annotateur' dans les performances du réseau : le réseau parvient à détecter des mangues que l'annotateur n'aura pas repérées (Fig. 78 et Fig. 53, p. 100). Il est cependant possible de pallier ce problème en retournant sur les images annotées. La définition de l'objet à annoter est également très subjective et varie d'un annotateur à l'autre. Pour les mangues, s'est en effet posé la question de quelles mangues inclure dans les annotations ? Certaines mangues ne sont en effet pas repérables par tous les annotateurs (ex : mangues occultées) ou certains objets similaires (feuilles vertes courbées) conduisent à des erreurs. D'autres, comme les mangues avortées, doivent être exclues de l'estimation de la production et donc de l'annotation. L'inclusion ou l'exclusion d'un objet dans l'annotation est donc subjective et dépendante de facteurs humains. Ces problématiques sont d'autant plus importantes que l'objet à détecter est complexe comme c'est le cas des inflorescences ou des pousses vertes (tests effectués en 2019 lors d'un stage de M1 du Cresi 'MangoLapse' intitulé « Identification d'organes spécifiques des stades phénologiques du manguier : évaluation des limites d'une approche par réseaux de neurones convolutifs »). La difficulté et le temps nécessaire à l'obtention des données d'entraînement en font un facteur limitant et une ressource pour l'utilisation des réseaux de neurones. Cette contrainte pousse les chercheurs à mutualiser leurs moyens en partageant leurs données d'annotation pour réduire les coûts d'entrainement (Koirala et al. 2019b). Cependant, ce partage n'est que très récent et les données ne sont parfois pas adaptées au type de réseau utilisé (convolutif vs. segmentation) ou aux questions de recherche.

b. La modélisation de la production

L'outil de détection de mangues, basé sur un réseau de neurones CNN, a donc été mis en place et utilisé sur les 300 arbres du réseau de vergers sur les années 2017 et 2018. Des premier tests, effectués en 2016 sur des manguiers sur le site du CIRAD de la Réunion, ont permis de définir le nombre d'images nécessaire pour obtenir facilement une estimation fiable du nombre de fruits présents dans l'arbre. Les arbres ont donc été photographiés sur deux faces opposées aux stades de maturité du fruit, c'est-à-dire de deux semaines à un mois avant la première récolte. Du fait de la structure spatiale du houppier, il est souvent impossible de photographier tous les fruits de l'arbre avec deux images. De plus, les deux images peuvent parfois contenir le même fruit (petits arbres) ou des fruits d'arbres voisins. Un modèle de calibration est donc nécessaire pour corriger le nombre de fruits détectés par le réseau de neurones ('automatic count') et estimer le nombre de fruits réellement contenu dans l'arbre ('field count') (Linker 2017). Une régression linéaire ou polynomiale a été construite pour chacune des trois variétés à partir du nombre de fruits réels mesuré

visuellement par 3 experts sur un total de 116 arbres échantillonnés dans les trois systèmes de culture et pendant les deux années d'étude. Ces régressions obtiennent des R^2 satisfaisants (entre 0.92 et 0.95) et ont été validées pour estimer le NRMSE (écart quadratique moyen normalisé par l'amplitude des valeurs) qui atteint environ 7% (Fig. 79). Ces résultats sont similaires à ceux trouvés par Koirala et al. (2019a) qui utilisent le même réseau et obtient des R^2 entre 0.94 et 0.98 sur des monocultures de manguiers. Le nombre de fruits estimé est ensuite multiplié par le poids moyen des fruits de la variété (voir Annexe IV) pour obtenir la production, en kg, de l'arbre.

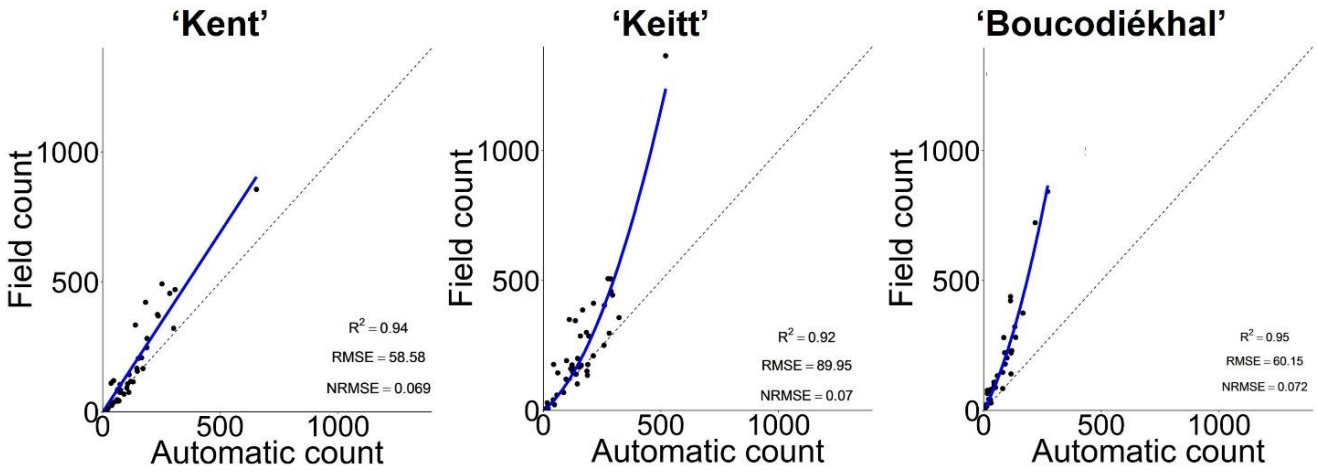


Fig. 79. Modèles d'estimation du nombre de mangues réelles à partir du nombre de mangues détectées pour les trois variétés étudiées

Alors que les études sur la détection des fruits sont nombreuses dans la littérature, seulement quelques travaux abordent l'estimation de la production à partir de ces détections. Dans la majorité des cas le nombre de fruits réel est estimée avec une régression simple (ou facteur de correction) calibrée sur la parcelle étudiée (Payne et al. 2013, Stein et al. 2016 ; Linker 2017). Dans le cadre de la thèse, nous avons opté pour un modèle calibré non pas par verger mais par variété. En effet, il était important d'avoir un modèle transposable à tous les vergers de la zone d'étude pour éviter une phase de calibration à chaque répétition ou dans chaque verger. Les faibles NRMSE montrent la stabilité des modèles quand la production est estimée sur un nouvel arbre de la même variété quel que soit son système de culture. Les différentes pentes obtenues par les régressions (Fig. 79) montrent par contre qu'une calibration par variété est nécessaire. Ceci s'explique par le fait que le rapport entre le nombre de fruits visibles (i.e., à l'extérieur du houppier) et le nombre de fruits invisibles (i.e., occulté ou à l'intérieur du houppier) n'est pas le même d'une variété à l'autre ; un résultat explicable par les différences architecturales entre les variétés, et/ou par certaines pratiques culturales comme la taille. Par exemple, la variété 'Boucodiékhal' n'est pas taillée, ce qui conduit à un plus grand nombre de fruits occultés par des feuilles ou à l'intérieur du houppier. Dans un objectif de

transférabilité sur d'autres variétés de mangues, d'autres systèmes de culture ou d'autres bassins de production, une phase préalable de validation de l'estimation sera nécessaire.

Plusieurs voies d'amélioration ont été proposées pour diminuer l'erreur d'estimation des modèles (RMSE). D'après Koirala et al. (2019a), le meilleur moyen de réduire l'erreur d'estimation serait d'affecter un facteur de correction par arbre. L'autre levier d'amélioration est de photographier un plus grand nombre de faces afin de capter tous les fruits du houppier. Payne et al. (2013) ont ainsi démontré qu'avec 4 images, l'estimation de la production du manguier était plus précise que pour 2 images. D'autres travaux abordent des approches stéréoscopiques ou multi-vues ('multi-view'). La stéréoscopie permet d'identifier le même fruit sur deux images prises du même point mais avec un angle différent et de supprimer les comptages doubles (Nuske et al. 2014). Les méthodes multi-vues consistent à utiliser un mécanisme de prise de vue automatique et de balayer l'arbre pour en obtenir plusieurs points de vue. Ces outils ont été développés sur mangue en Australie par Bargoti et Underwood (2017) et Stein et al. (2016). Ces équipes ont mis au point un robot équipé d'un GPS, d'un LiDAR et d'un capteur RVB pour balayer tous les arbres du verger et obtenir un plan continu des fruits (Fig. 30, p. 59). Leurs résultats montrent que le robot peut détecter jusqu'à 98.7 % des fruits de l'image (Stein et al. 2016). La stéréoscopie et la multi-vue sont des méthodes très précises mais qui nécessitent des matériels beaucoup plus couteux et complexes d'utilisation que ceux que nous avons utilisés (un capteur RVB classique, soit un simple appareil photographique ou smartphone). S'ajoutent à cela des temps de traitement augmentés par le nombre de données et une complexité à les traiter (photogrammétrie et fusion de données). Au contraire, nous avons montré que l'utilisation d'outils frugaux est suffisante pour obtenir une estimation fiables ($R^2 > 0.90$) en utilisant uniquement deux images de l'arbre. Au vu de la diversité des systèmes étudiés, les performances des modèles d'estimation de la production sont jugées très satisfaisantes.

I.2. Caractérisation de vergers complexes

La structure des arbres et des vergers (densité de plantation, motif de plantation) ont été estimées au moyen de l'analyse d'image. Pour ce faire, un drone portant un capteur RVB a été utilisé permettant d'obtenir une orthomosaïque très haute résolution (THR, 1.30 $cm.pixel^{-1}$) et un Canopy Height Model (CHM). Ces deux objets sont utilisés dans une procédure de classification orientée objet (GEOBIA) qui a permis de délimiter chaque objet du verger (arbre, bâtit, zone de maraîchage), de les classifier et d'obtenir une carte d'occupation des sols et les différentes métriques de structure des manguiers (hauteur, volume et surface). Le drone a été choisi car il offre de nombreux avantages par rapport à l'imagerie satellitaire ou aérienne notamment sa flexibilité et son faible coût (Pádua et al. 2017). Le potentiel du drone pour obtenir des cartes d'occupation

des sols et estimer la structure des arbres a été en partie discuté dans la Partie I du Chapitre 2. Nous reprenons et approfondissons ici quelques points.

a. Carte d'occupation des sols

Pour les vergers de l'étude nous avons choisi de construire la carte d'occupation des sols en utilisant une approche GEOBIA (Blaschke et al. 2014). Ce type de classification offre de nombreux avantages par rapport à une classification centrée sur le pixel pour les images THR. Elle permet de délimiter le houppier de chaque arbre et d'extraire métriques à l'échelle de l'arbre (variables de structure) et du verger (indices de configuration et de composition, densité de plantation, etc.). La procédure GEOBIA inclut une phase de segmentation afin de morceler l'orthomosaïque en plusieurs centaines d'objets. Nous avons opté une segmentation multi-résolution ('*multiresolution segmentation'*, Baatz and Schäpe 2000) dont le but est de regrouper des pixels voisins de manière à obtenir des objets homogènes selon les valeurs des bandes RVB et du CHM. Chaque objet est ensuite classé dans un processus de classification qui contient en fait deux niveaux : d'abord une classification sur l'ensemble du verger pour obtenir l'espèce des arbres, 6 au totale dont le manguier, puis une classification de chaque manguier selon leur variété ('Kent', 'Keitt', 'Boucodiékhal' ou autre). Chaque classification utilise un algorithme de Random Forest (RF). La procédure GEOBIA a été validée, au moyen de points de contrôle, avec une précision de 0.89 en moyenne signifiant que 89% des objets ont été correctement classifiés. Cette précision reste comparable à celles obtenues dans plusieurs études utilisant des capteurs RVB et le drone pour classifier des espèces arborées (Lisein et al. 2015 ; Nevalainen et al. 2017). Plusieurs pistes ont été évoquées (voir Chapitre 2-Partie I) afin d'améliorer la classification. Une option serait d'augmenter l'information spectrale en optant par exemple pour un capteur multispectral ou hyperspectral (Nevalainen et al. 2017) qui obtiennent de meilleurs résultats sur la classification. Dans notre cas, le choix du RVB est lié à un objectif d'utiliser des outils frugaux. En effet, le drone utilisé dans cette thèse est un drone commercial qui permet d'avoir une cartographie 3D précise pour un coût inférieur à 1000€ (en utilisant des logiciels libres) tandis que les capteurs multispectraux et hyperspectraux restent bien plus cher (premier prix à 5000€). De plus, ces capteurs doivent être le plus souvent combinés à des capteurs RVB (appareil photographique ou GoPro©) pour obtenir une image THR, induisant une complexité du traitement (fusion de données) et l'utilisation de drones professionnels beaucoup plus chers (environ 5000€). Les données RVB et le *structure-from-motion* (SfM) utilisés durant cette thèse sont considérées comme largement suffisantes pour analyser des vergers même complexes (Wallace et al. 2016).

Une des principales contraintes de notre procédé GEOBIA reste le temps de calcul (~15h par carte en moyenne) qui est augmenté par plusieurs étapes manuelles. D'abord, le post-traitement qui permet de

fusionner les objets créés par la segmentation constitue l'étape la plus fastidieuse. En effet, pour obtenir une bonne classification il est nécessaire de créer des objets les plus homogènes possibles (Ma et al. 2015) conduisant à une sur-segmentation des arbres, c'est-à-dire que plusieurs objets composent un seul arbre (Fig. 80). Le post-traitement doit donc à la fois fusionner les bons objets et choisir la classe du super-objet. Il existe plusieurs stratégies de fusion des objets (Ke and Quackenbush 2011), cependant aucune n'est adaptée à des paysages complexes où les arbres ont des formes très variables. Du fait de la diversité des systèmes de notre étude, cette étape de post-traitement se compose d'une chaine de plusieurs règles logiques (ex : objets voisins de classe 'manguier' partagent au moins 70% de leur bordure = fusion) et a été adaptée à chaque type de verger (extensif, diversifié et intensif). Ce post-traitement a été vérifié pour chaque objet 'arbre' dont certains ont été remodelés à la main sur un logiciel SIG afin d'obtenir une carte la plus précise possible. Pour supprimer cette étape de post-traitement il serait nécessaire de parfaitement délimiter chaque objet puis de le classifier. Actuellement aucun des algorithmes mis à disposition par les logiciels de télédétection (ex : eCognition) ne permet d'obtenir de tels résultats et la délimitation des arbres reste un sujet de recherche à part entière (Ke and Quackenbush 2011). Des travaux récents ont montré l'efficacité des réseaux de neurones pour la segmentation et la classification d'images THR mais pour l'instant très peu sont appliqués aux images drones (Weinstein et al. 2019).

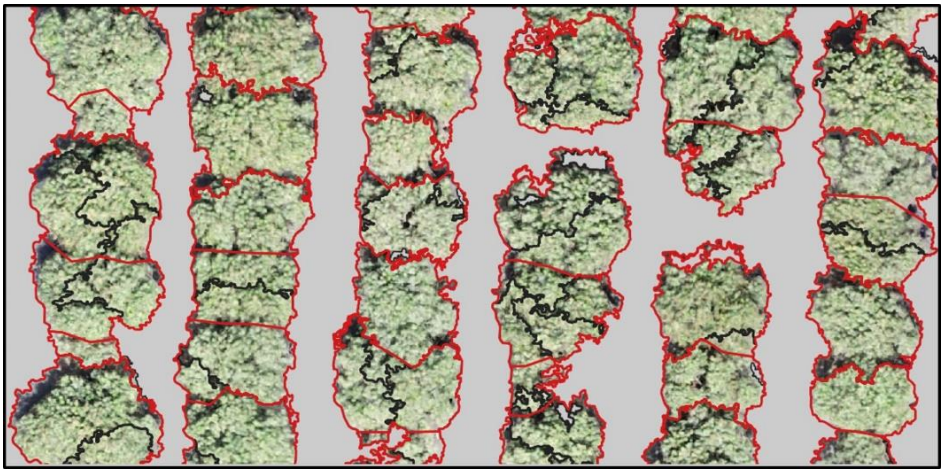


Fig. 80. Exemple de situation de sur-segmentation des houppiers. L'objectif du post traitement est de parvenir à fusionner les objets (en noir) pour former correctement le houppier de l'arbre (en rouge).

L'autre étape manuelle de la procédure GEOBIA est l'entrainement du RF dont la durée était de une à deux heures par verger. Cette étape nécessite d'échantillonner et de classifier manuellement environ 2.5% de la totalité des objets pour obtenir une précision de classification suffisante. Du fait de la variabilité de l'apparence des arbres (variété, vigueur et santé des arbres, etc.) et des conditions de vols (luminosité et vent) l'aspect des arbres d'une même espèce varie d'un verger à l'autre. Dans ces conditions nous avons choisi de construire un jeu d'entrainement par verger ce qui multiplie le temps d'entrainement par le nombre de vergers. En effet, l'algorithme RF obtient les meilleurs résultats pour la classification d'objet mais, une

fois entrainé, il est difficilement transférable d'une carte à une autre (Juel et al. 2015). Ici encore, les réseaux de neurones, de par leur capacité à traiter des classes hétérogènes et leur transférabilité d'une situation à l'autre, pourraient permettre de pallier ce problème (Zhao et al. 2017).

b. Estimation de la structure des arbres

Les algorithmes de SfM permettent d'obtenir des nuages de points 3D avec un simple capteur RVB mais dont la précision égale celle du LiDAR (Wallace et al. 2016). Afin de vérifier la fiabilité de l'estimation de la structure des arbres par le drone nous avons mesuré la hauteur réelle de 144 arbres. Le NRMSE de l'estimation des hauteurs atteint 11% (soit un RMSE de 0.47 m), semblable aux autres études qui se situent entre 6 et 20% (Díaz-Varela et al. 2015; Torres-Sánchez et al. 2015). Cependant la pente entre la hauteur réelle et la hauteur estimée (~0.77) est différente de 1 ce qui est également observé dans les autres études. Le drone sous-estime donc la hauteur réelle, cependant cette estimation est stable ($R^2 = 0.96$) quelle que soit la hauteur des arbres (qui varie entre 1.5 et 10 m) et le système de culture. La mesure de hauteur est donc suffisamment précise pour répondre à l'objectif d'obtenir des variables stables et facilement mesurables reflétant la structure de l'arbre.

Dans certaines situations, le drone n'a pas pu reconstituer parfaitement la structure 3D des arbres. Ceci correspond à 2.1% des manguiers dans l'échantillon de 30 vergers pour lesquels l'information de structure est faussée ou nulle. Il s'est avéré que ce taux était plus élevé dans les vergers à haute densité de plantation (vergers intensifs). En effet, le fort rapprochement entre les arbres empêche la triangulation de parfaitement reconstituer les points 3D. Afin de réduire ce taux d'erreur il est possible d'améliorer le protocole d'acquisition des images en jouant par exemple sur le taux de chevauchement des images ('overlap'), l'orientation du capteur, ou bien la hauteur ou la grille de vol (Dandois et al. 2015). À l'heure actuelle, les paramétrages du vol de drone sont souvent laissés à l'expertise du pilote et il existe peu d'études évaluant l'impact des différents paramétrages sur les données acquises notamment dans les systèmes complexes. L'utilisation de capteurs actifs tels que le LiDAR (mais également des capteurs satellitaires comme les radars) peut également s'envisager. Ces capteurs sont efficaces pour caractériser les milieux complexes car ils ne sont pas sensibles à la densité de couverture végétale (Wallace et al. 2016). Cependant, au vu des coûts d'opération pour les données LiDAR (ou radar satellite), la photogrammétrie reste plus avantageuse pour estimer des variables corrélées à la structure (Wallace et al. 2016). Dans les deux cas, si l'objectif est de réaliser des mesures de structure des arbres, par exemple pour utiliser des modèles d'estimation de la biomasse, une phase de calibration terrain est nécessaire que ce soit pour la photogrammétrie ou pour le LiDAR.

II. Méthodes d'échantillonnage et estimation du rendement du verger

L'estimation de la production à l'échelle de l'arbre pose certes des contraintes techniques, liées aux outils d'analyse d'image, mais il est possible d'obtenir une valeur fiable et précise du nombre de fruits en traitant les sorties des algorithmes par des modèles spécifiques (voir Discussion – Partie I). L'estimation du rendement à l'échelle du verger reste plus complexe : d'une part à cause de la faible corrélation entre des variables facilement mesurables (indices de végétation, structures des arbres, etc.) et la production, du fait des phénomènes d'alternance et d'asynchronisme ; d'autre part à cause des hétérogénéités intra-verger qui complexifient les méthodes d'extrapolation de la production de l'arbre vers le rendement du verger. Lors de cette thèse deux approches d'estimation du rendement ont été explorées (Fig. 32, p. 66). La première porte sur l'utilisation de drone et de la modélisation pour spatialiser les rendements. La seconde est basée sur une méthode d'échantillonnage d'un certain nombre d'arbres pour l'estimation du rendement du verger en fonction de ces caractéristiques.

II.1. Utiliser l'imagerie drone pour l'estimation des rendements

Comme décrit précédemment, le drone permet d'obtenir des informations précises sur l'occupation du sol et permet de délimiter les houppiers des arbres. La carte d'occupation du sol et les variables de structures obtenues par le CHM (i.e., hauteur, volume et surface du houppier de chaque arbre) ont été utilisées pour mettre en place un modèle d'estimation de la production pour chacune des trois variétés étudiées (Chapitre 2 – Partie I). En plus de prendre en compte les variables de structure, une variable proxy appelée '*load index*' a été introduite pour prendre en compte les effets de l'année et des pratiques culturales. En effet, deux arbres d'une même variété peuvent avoir une structure identique tandis que leur production va varier selon les conditions climatiques, les pratiques culturales, et des facteurs endogènes à l'arbre selon des règles complexes. Nous avons donc défini le *load index* (*LI*) comme le rapport évalué visuellement entre la surface du houppier occupée par les fruits et la surface verticale du houppier (Fig. 41, p. 76). Ainsi un *LI* de 0 correspond à la situation où aucun fruit n'est visible tandis qu'un *LI* de 0.5 correspond à 50% de la surface du houppier occupée par des fruits. Le *LI* est donc estimé par une analyse visuelle de l'arbre et permet une estimation rapide de la charge de l'arbre indépendamment de sa structure. En effet, un petit arbre portant 30 fruits peut-avoir un *LI* identique à un gros arbre en portant 1 500. Sa mesure a pu être facilitée par la mise en place d'un abaque (Annexe IV).

La calibration des modèles d'estimation du rendement par drone utilise donc la production des arbres (estimée en utilisant les outils d'analyse d'image, voir Chapitre 1) expliquée en fonction des variables de structure (extraites par cartographie drone) et le *LI*. En 2017, la calibration a été réalisée sur la production de 150 arbres estimée par l'algorithme KNN et des modèles de régression multiples ('*drone-KNN',* Tableau

8, p. 143). En 2018, l'utilisation du réseau de neurones CNN et de la production de l'ensemble des arbres (300 au total) sur deux années (2017 et 2018) ont permis de calibrer de nouveaux modèles ('*drone-CNN*') en utilisant une régression GLM (Generalized Linear Model) avec une distribution Gaussienne (Tableau 18). Les modèles généralisés *drone-CNN* ont été validés sur les mêmes données de validation que pour les modèles *drone-KNN*, soit 60 arbres dont la production réelle a été mesurée sur le terrain en 2017 et 2018 (120 points au total). Les modèles *drone-CNN* obtiennent un NRMSE de 12 à 18%, selon les variétés, contre 20 à 30% pour les modèles *drone-KNN*, soit une réduction de presque la moitié de l'erreur d'estimation. Les meilleures performances des modèles *drone-CNN* peuvent s'expliquer par la réduction de l'erreur dans l'estimation de la production du CNN par rapport au KNN (voir Discussion - Partie I.1).

Dans ces modèles (*drone-CNN* et *drone-KNN*), le *LI*, mesuré en pourcentage, a été transformé en variable catégorielle contenant 4 modalités de valeurs (Fig. 41, p. 76) : nul ($LI = 0$), faible ($0 < LI \leq 15\%$), moyen ($15\% < LI \leq 25\%$ et élevé ($LI > 25\%$). Cette transformation permet d'établir une relation d'extrapolation de la production à l'échelle du verger. En effet, pour estimer le rendement du verger à partir de ces variables, la mesure du *LI* de chaque manguier serait nécessaire ce qui rendrait l'estimation du rendement laborieuse. Le *LI* est donc estimé sur 50 arbres échantillonnés sur un transect (Fig. 42, p. 76) pour intégrer l'hétérogénéité de charge des arbres du verger. À partir de ces 50 arbres, la proportion de chaque catégorie de *LI* (notées a, b, c et d) est estimée pour chaque variété présente dans le verger. La production *P* de chaque arbre (exprimée kg) est ensuite estimée pour les 4 catégories de *load index* (avec pour LI = 0, $P = 0$) en utilisant les modèles du Tableau 18. Puis, la production moyenne $\bar{P}$ de l'arbre est obtenue en faisant la somme des quatre estimations de *P* pondérées par les proportions a, b, c et d de *LI* (voir Eq. 13, p. 141). La production de chaque variété du verger est obtenue en faisant la somme des productions de chaque arbre. Par exemple pour la variété 'Kent' :

Production arbre 'Kent' : $$\bar{P}_{kn,i} = aP_{kn,i/LI=low} + bP_{kn,i/LI=med} + cP_{kn,i/LI=high} + d \times 0 \quad (13)$$

Rendement verger en 'Kent' : $$Y_{kn} = \frac{\sum_{i=1}^{N_k} \bar{P}_{kn,i}}{A} \times 10^3 \quad (23)$$

avec : $\bar{P}_{kn,i}$ la production moyenne pondérée de l'arbre i (variété 'Kent') en kg ; $P_{kn,i}$ la production de l'arbre i pour chaque catégorie de *LI* (*low*, *med*, *high*) en kg ; a, b, c, et d les proportions respectives des catégorie de *LI* faible, moyen, élevé et nul avec : $a + b + c + d = 1$; A la surface productive du verger en ha ; et Y_{kn} le rendement en 'Kent' du verger en kg.ha^{-1}.

Tableau 18. Résultats des modèles GLM (*'drone-CNN'*) d'estimation de la production utilisant le *load index LI* et les structures individuelles des arbres mesurées par drone. La contribution du *LI* (%) correspond au pourcentage du R^2 expliqué par la variable *LI*.

Variété	Modèle*	R^2	RMSE (fruit.arbre^{-1})	NRMSE (%)	Contribution *LI* (%)
'Kent'	*P ~ LI + Height + Area + Vol*	0.77	100.20	11.75	78.3
'Keitt'		0.80	252.06	13.57	61.4
'BDH'		0.68	108.75	13.07	60.0
'Autres'		0.76	164.14	12.06	68.1

* La formule du modèle GLM est la même pour toutes les variétés.

Grâce à cette approche, la production de chaque arbre a pu être cartographiée avec une estimation du rendement pouvant se faire jusqu'à 2 mois avant la récolte. Cette méthodologie représente la première estimation du rendement d'une culture pérenne utilisant l'imagerie drone. Les études sur l'estimation et la spatialisation individuelle de la production des arbres restent rares et principalement basées sur l'imagerie satellitaire (Usha and Singh 2013). La plupart estime la production d'un échantillon d'arbres, et utilisent des techniques d'interpolation pour cartographier les rendements sous forme de carte de chaleur (ex : Aggelopoulou et al. 2010 ; Mann et al. 2011). Ce genre d'étude reste très utile pour analyser l'hétérogénéité de production intra-verger, cependant elles ne cartographient pas réellement la production de chaque arbre. Quelques études se sont penchées sur l'utilisation des images satellites ou aériennes pour cartographier et estimer directement les rendements (Ye et al. 2007; Rahman et al. 2018). Sur mangue, la récente étude de Rahman et al. (2018) a obtenu des résultats intéressants sur l'estimation du rendement à partir d'images satellites (WorldView-3, résolution spatiale de 1.24 m.pixel^{-1}) et de la modélisation. En combinant 18 indices de végétations et la surface du houppier, leur modèle d'estimation de la production, calibré sur trois vergers monovariétaux, atteint un R^2 de 0.70 et un RMSE de 33 fruits. Le faible RMSE obtenu dans cette étude est à relativiser par la faible plage de valeur à prédire dans leur verger (0 à 300 fruits) et par l'homogénéité des arbres étudiés (vergers intensifs) comparés à notre cas d'étude (nombre de fruits par arbre à estimer variant de 0 à 1360 fruits). Ainsi, la méthodologie drone développée lors de cette thèse obtient des performances égales à l'imagerie satellitaire, mais pouvant être appliquées à une plus grande diversité de contextes (système de culture et variété), pour un coût moins élevé (un capteur RVB étant suffisant) et une plus grande flexibilité d'usage.

Dans les modèles drones, le *LI* permet de calibrer les modèles pour prendre en compte les variations inter-annuelles et entre les vergers (du fait des pratiques culturales notamment). La forte contribution du *LI* dans les modèles d'estimation (le *LI* explique entre 60 et 80% du modèle selon les variétés, Tableau 18) montre que sa mesure est indispensable pour chaque nouveau verger et pour chaque année. Cette variable, liée au taux de charge (i.e, le nombre de fruits par unité de volume), est indépendante de la structure de l'arbre (voir Discussion – Partie III). Avec cette méthode, la production du verger est évaluée sans mesure de la structure

de l'arbre au sol, la structure de l'arbre étant estimée par drone. De même, les phénomènes d'alternance et les variations de l'état des arbres font que l'utilisation d'indices de végétation, calculés sur des images satellites ou autre, nécessite également une calibration des modèles d'estimation pour chaque verger et pour chaque année (Rahman et al. 2018). Ainsi la calibration de terrain reste nécessaire pour la spatialisation de la production des plantes pérennes lorsque des images aériennes ou satellites sont utilisées. Cette calibration est permise soit par un comptage visuel du nombre de fruits par arbre avant la récolte, soit par une variable proxy comme le *load index,* introduite ici, ou une estimation de la floraison pour certaines espèces (Aggelopoulou et al. 2011). Cependant ces mesures restent fastidieuses et subjectives même si le *LI* est, par construction, plus rapide à évaluer que le nombre de fruits par arbre.

À l'heure actuelle, seule la télédétection au sol permet une spatialisation directe de la production sans passer par des modèles. Ainsi, le balayage au sol du verger par des capteurs LiDAR et/ou RVB, porté par des plateformes automatisées (robot) ou non (véhicule), permet d'obtenir de très bons résultats pour la cartographie de la production (Stein et al. 2016). Ce genre d'outils a été testé notamment dans des monocultures de mangues et a permis de cartographier la production par arbre avec une erreur de 1.36%. Cependant, ces outils restent lourds à mettre en place et, bien qu'adaptés à des vergers intensifs, ils restent inopérants dans des vergers complexes plurispécifiques et à l'arrangement spatial des arbres aléatoire. Une des pistes envisageables serait de remplacer la mesure visuelle du *LI*, servant à intégrer les variabilités annuelles, en utilisant des outils plus précis et moins subjectifs qu'un opérateur humain. Ainsi, il serait intéressant d'explorer l'utilisation d'outils similaires à celui développé dans cette thèse pour l'estimation de la production de l'arbre (Chapitre 1), pour estimer précisément et rapidement le *LI* ou un nombre de fleurs.

II.2. Combiner les variabilités intra-verger et l'estimation de la production

Les méthodes d'estimation du rendement décrites ci-dessus cherchent à estimer directement la production de tous les arbres en utilisant la modélisation. Une seconde piste, moins 'gourmande' en données, a été explorée durant cette thèse et utilise des méthodes d'estimation du rendement à partir d'un échantillon d'arbres sur lesquels le nombre de fruits a été mesuré (Chapitre 2 – Partie II). En effet, la méthodologie drone déployée au cours de cette thèse reste un moyen très précis pour caractériser les vergers (cartes d'occupation des sols et structures des arbres) mais nécessite un traitement long (~15h) et n'est pas entièrement automatisée. De plus, l'outil drone requis pour cette méthode est encore peu utilisé et accessible pour les producteurs. La seconde méthode considère un échantillonnage adapté en fonction de la variabilité intra-verger et permet de calculer le rendement du verger sans connaissances *a priori* du verger (Anderson et al. 2019), c'est-à-dire sans connaitre l'historique de production (série temporelle ou production de l'année précédente). Connaître l'évolution passée de la production du verger aide à la prédiction de rendement de

l'année en cours mais demeure inaccessible dans la majorité des vergers ouest-africains notamment chez les petits producteurs. Pour identifier la méthode d'échantillonnage la plus fiable pour chaque système de culture, nous avons caractérisé plus finement trois vergers représentant chacun un des types de systèmes de culture (Chapitre 2 – Partie II).

En premier lieu, l'hétérogénéité intra-verger de la production de mangues a été décrite en utilisant l'outil d'estimation de la production à l'échelle de l'arbre (CNN) et en photographiant tous les arbres de chaque verger. Il est apparu que l'hétérogénéité intra-verger de la structure des arbres était différente dans les trois vergers, de même que l'hétérogénéité de la production. Plusieurs études ont montré l'utilité de prendre en compte l'hétérogénéité intra-verger, notamment dans le cadre de l'agriculture de précision (Peeters et al. 2015 ; Anderson et al. 2018), par exemple pour délimiter des zones avec différents niveaux de pratiques (ex : zone de fertilisation azotée). L'hétérogénéité intra-verger peut s'avérer être utile pour mettre au point des stratégies d'échantillonnage pour l'estimation de la production. Les différentes stratégies que nous avons testées utilisent soit aucune connaissance du verger (échantillonnage aléatoire), soit la connaissance du volume des arbres (échantillonnage par groupe de volume). Nous avons montré que l'erreur d'estimation du rendement du verger par échantillonnage dépend de la variabilité de production des arbres (décrite par le coefficient de variation, CV) : plus cette variabilité est forte, plus le nombre d'arbres à échantillonner doit être grand pour atteindre un taux d'erreur acceptable (10%). La variabilité de production des arbres apparaît beaucoup plus forte dans le verger extensif (CV = 116.2%) que dans le verger diversifié (CV = 77.4%) puis intensif (CV = 32.7%). Les résultats des différentes méthodes d'échantillonnages dépendent du type de verger : l'échantillonnage aléatoire de 20 et 10% des arbres constitue la méthode la plus intéressante, en termes d'équilibre entre le nombre d'arbres à sélectionner et l'erreur d'estimation, pour le verger extensif (CV du rendement estimé de 14.1%) et intensif (CV = 4.5%) respectivement, alors qu'un échantillonnage de 20% des arbres par groupe de volume donne de meilleurs résultats pour le verger diversifié (CV = 6.6%). De plus, la prise en compte de la proportion de chaque variété diminue l'erreur d'estimation dans les vergers diversifiés (CV = 6%). Ainsi, à partir de l'estimation de la production d'un certain nombre d'arbres, l'estimation du rendement du verger est possible. La mise en place et l'analyse de stratégies d'échantillonnage pour l'estimation du rendement constituent une première pour des vergers extensifs et diversifiés. Dans des vergers intensifs de manguiers, Anderson et al. (2018) ont déjà travaillé sur la mise en place de stratégies pour l'estimation du rendement. Dans leur étude, une cartographie de la production des arbres est utilisée pour déterminer des zones de production et ainsi établir une stratégie d'échantillonnage en utilisant cette information. Dans notre cas, il s'avère l'information sur le volume des arbres est suffisante pour établir une stratégie d'échantillonnage fiable.

La méthode d'échantillonnage et d'estimation de la production permet une estimation du rendement du verger sans mesure préalable de la structure des arbres. En comparant l'estimation par drone (modèle *drone_CNN*,voir partie précédente) et l'estimation par la méthode d'échantillonnage, on trouve que l'estimation par échantillonnage donne de meilleurs résultats (Tableau 19). La méthodologie *drone_CNN* est pénalisée par la mesure subjective du *LI* qui a un poids important dans l'estimation de la production (Tableau 18). Les bons résultats des méthodes d'échantillonnage et le fait que l'information de structure ne soit nécessaire que dans les vergers les plus complexes - qui ne représentent pas la majeure partie des vergers de la zone d'étude - devrait permettre d'aller vers des outils d'estimation du rendement par une caractérisation simplifiée du verger. En effet, l'échantillonnage, contrairement à la méthode drone, ne prend en compte que la densité de plantation (i.e., le nombre d'arbres par hectare), le type de système (pour déterminer le nombre d'arbres à échantillonner) et, dans le cas des vergers diversifiés, la structure des arbres et le pourcentage variétal (Tableau 12 et 13, p. 170).

La stratégie d'échantillonnage peut encore être améliorée. Dans cette thèse, nous avons uniquement testé le volume des arbres pour les grouper mais d'autres variables peuvent être utilisées comme la vigueur de l'arbre (traduite facilement par des indices de végétation de type NDVI) ou encore la circonférence du tronc (Mann et al. 2011 ; Peeters et al. 2015). De plus, il pourrait être envisagé d'avoir une stratégie adaptative en fonction de la progression de l'échantillonnage. En effet, l'estimation par échantillonnage aléatoire reste dépendante de la manière dont l'opérateur choisi les arbres dans le verger. Les résultats sur la variabilité de production que le CV de la production dépasse les 100% dans les vergers extensifs, majoritaires dans la zone d'étude. Il faut donc sélectionner les arbres de la manière la plus aléatoire possible pour capter l'hétérogénéité du verger. En utilisant un échantillonnage adaptatif, l'hétérogénéité des arbres déjà échantillonné peut être rapidement évaluée et ainsi orienter l'opérateur pour choisir le prochain arbre à estimer (Thompson 2013). Ce type de méthode nécessite une estimation de la production en temps réel ce qui est notamment permise par certains outils de réseau de neurone (Koirala et al. 2019a).

Tableau 19. Comparaison de l'erreur d'estimation du rendement obtenu par la méthode drone (modélisation et utilisation de transect) et la méthode d'échantillonnage.

Type de verger	Rendement réel (kg)	Erreur d'estimation (%)	
		Méthode drone*	Méthode échantillonnage**
Extensif	11 189	35%	14.1%
Diversifié	19 962	24%	6.6%
Intensif	28 072	5.4%	4.5%

* Erreur d'estimation calculée sur l'écart entre le rendement réel et l'estimation à partir de la méthode drone (modèle *drone_CNN*). ** Erreur d'estimation (ou CV dans ce cas) calculée à partir de 1000 simulation d'échantillonnage.

II.3. Extrapolation au bassin de production

Un des objectifs de départ de la thèse était de pouvoir estimer la production de la totalité du bassin de production des Niayes (500 km^2). Si nos travaux de recherche ont permis de résoudre la plupart des problématiques qui ont été soulevées aux échelles de l'arbre et du verger, l'échelle régionale n'a pas pu être totalement finalisée. L'utilisation d'images satellite pour caractériser la zone de production a été explorée (stage de M2 « Cartographie des vergers de manguiers au Sénégal par télédétection orientée-objet multi-échelle ») mais un certain nombre de problèmes méthodologiques rencontrés ont été identifiés.

Sur des images Pléiades (résolution spatiale multispectrale de 2 m et panchromatique de 0.5 m) une classification orientée objet a été utilisée pour segmenter les vergers ou des sous-ensembles de vergers (i.e., élément uniforme dans un verger) de la zone d'étude et tenter de les classifier dans les trois systèmes de culture étudiés. Bien que la segmentation des vergers ait été possible, leur classification dans les trois systèmes de culture reste complexe du fait de la structure qui est similaire entre les vergers extensifs et les diversifiés. La distinction entre ces deux systèmes se base sur la diversité cultivée du verger et particulièrement sur la présence de maraîchage. Or, dans la plupart des vergers diversifiés, le maraîchage est soit difficile à repérer (car sous les arbres), soit absent (le maraichage n'est mis en place qu'une partie de l'année) contribuant ainsi à la confusion avec les vergers extensifs. De plus, pour distinguer les différents systèmes, il est nécessaire d'identifier les espèces arborées. Or, cette étape n'a pu être validée avec les images Pléiades du fait de la ressemblance spectrale entre certaines espèces comme le manguier, les agrumes et l'anacardier (Fig. 81). Ainsi, la distinction des différents types de système s'est avérée infructueuse et seule une carte d'occupation des sols contenant une seule classe '*verger*' a pu être obtenue (Fig. 82). Même l'extraction d'une couche '*manguier*' paraît complexe sur une aussi grande surface (~500 km^2) avec les données dont nous disposions.

La délimitation et la classification des vergers de la zone constitue donc une question de recherche qui reste posée. Plusieurs travaux ont utilisé avec succès la télédétection orientée-objet pour délimiter et/ou classifier des parcelles à l'échelle régionale (Amorós López et al. 2011). Une des solutions pour les arbres fruitiers, serait d'utiliser une série temporelle de plusieurs images prises à des périodes clés de la phénologie des arbres afin de distinguer les espèces arborées qui ont des phénologies différenciées (Peña et al. 2017). Ainsi pour le manguier de la zone d'étude, l'idéal serait d'utiliser des images prises en février, à la période de floraison et donc au moment où les manguiers ont une signature spectrale plus distincte.

Ainsi, à la vue de la difficulté à repérer les vergers, l'obtention par photo-interprétation d'une carte d'occupation des sols, avec la délimitation des vergers et la définition de leur type de système, est en cours d'élaboration. À partir de cette carte, il sera possible de calculer des indices de structures des vergers

notamment la surface du verger et la densité de plantation et, si possible, la surface occupée par les manguiers. L'estimation du rendement de ces vergers devrait donc être possible en utilisant des modèles d'extrapolation prenant en compte le type de verger, sa situation géographique (pédoclimat) et sa structure. Quelques études ont montré l'intérêt des images satellitaires pour mettre en place des modèles de rendement à l'échelle régionale mais qui restent moins performants que ceux à l'échelle des vergers (voir Discussion – Partie I.1 et Rahman et al. 2018).

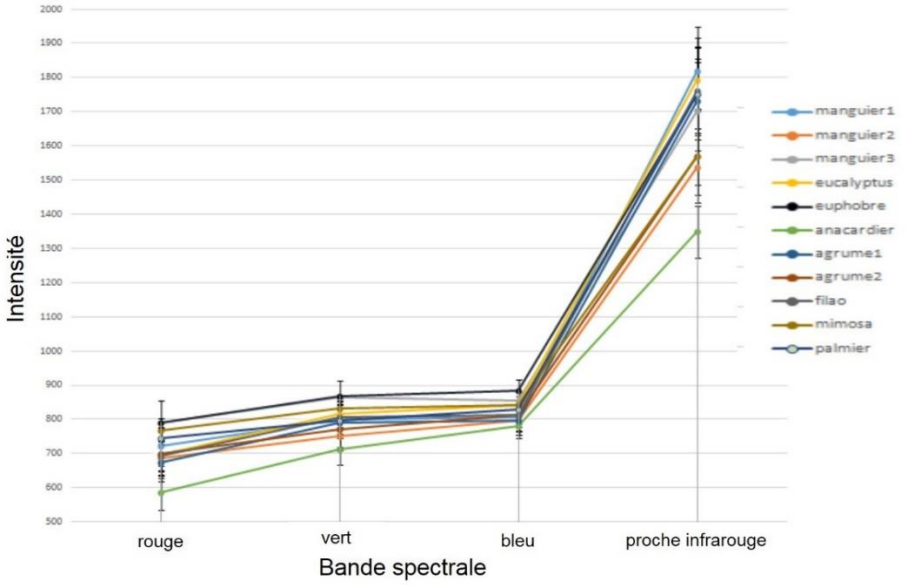


Fig. 81. Analyse des signatures spectrales des espèces arborées de la zone des Niayes (source : M. Smail, étudiant M2 Géomatique, CIRAD HortSys – Tétis).

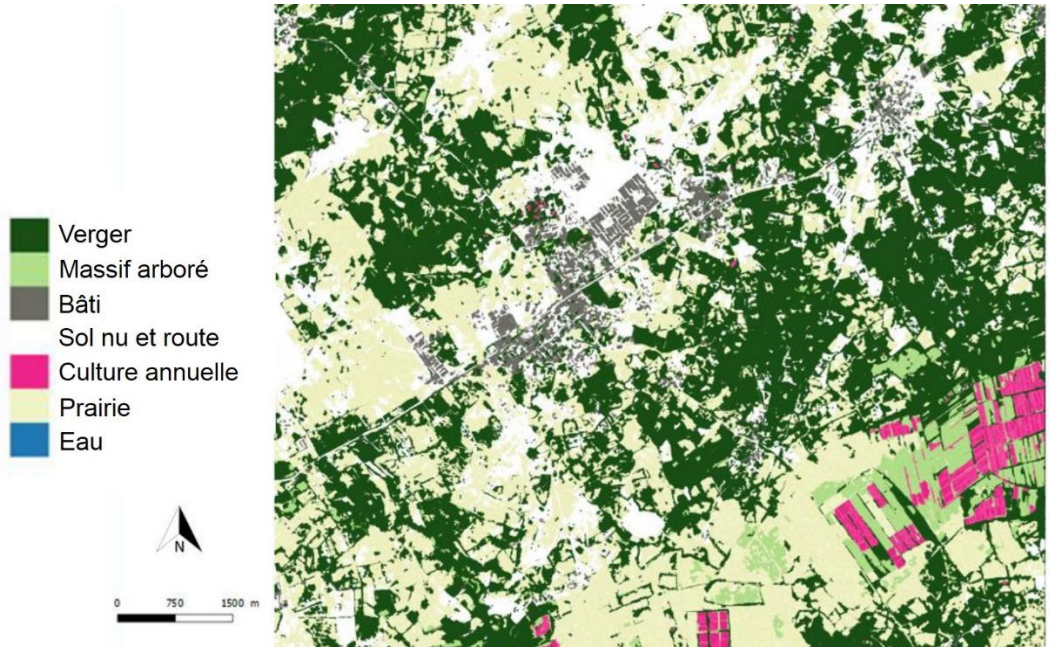


Fig. 82. Carte d'occupation des sols de la zone des Niayes obtenues à partir d'images Pléiades (source : M. Smail, étudiant M2 Géomatique, CIRAD HortSys – Tétis).

III. Facteurs de variabilité du rendement des cultures pérennes à différentes échelles

III.1. Facteurs de variabilité du rendement

Les différents outils mis en place pour l'estimation des rendements aux échelles de l'arbre et du verger nous ont permis de décrire les facteurs de variabilité à ces deux échelles. Pour ce faire nous avons adapté deux méthodologies : l'analyse des '*yield gap*' et le diagnostic agronomique régional.

a. Échelle de l'arbre

Le concept de production gap

La mise en place de l'outil d'estimation de la production de l'arbre nous a permis d'obtenir les données de production pour les 300 manguiers du réseau de vergers et sur deux années consécutives (2017-2018). S'ajoute à cela, un nombre important d'informations obtenues sur la structure de l'arbre (par photogrammétrie drone, voir Chapitre 2 –Partie I), la structure du verger (composition spécifique et densité de plantation), sur les pratiques appliquées au manguier et sur son environnement (via les capteurs de température et des bases de données). Pour analyser ces données nous avons adapté le concept de *yield gap* (écart de rendements) qui permet d'explorer les variables explicatives du rendement en en définissant plusieurs niveaux (van Ittersum et al. 2013).

La méthodologie du yield gap a été majoritairement utilisée sur les cultures annuelles (Lobell et al. 2009 ; Affholder et al. 2013) et très peu sur les arbres fruitiers (de Bie 2004) et uniquement à l'échelle de l'exploitation. Dans ces travaux de thèse, nous avons adapté ce concept en redéfinissant les niveaux de rendement pour y intégrer le caractère pérenne de l'arbre (Fig. 57a, p. 110). Ainsi nous avons défini la production potentielle (P_p) comme la production de l'arbre uniquement limitée par les conditions exogènes (facteurs de croissance ou 'growing-factors') de climat (température, précipitation et radiation solaire) et de sol. La production potentielle correspond donc à la production obtenue avec un développement de l'arbre optimal (génétique x âge x structure), sans stresses biotiques ni phénomène d'alternance. Le calcul de P_p suppose donc que les conditions optimales d'âge, de structure et de génétique (cultivar et porte-greffe) soient connues. Pour de nombreuses espèces de grande culture, l'utilisation de la modélisation donne accès à la mesure du potentiel de rendement dans une situation pédoclimatique donnée (Jeuffroy et al. 2014). Ceci n'est pas le cas pour les plantes pérennes (et particulièrement pour le manguier) pour lesquelles il est impossible d'optimiser certaines variables endogènes de l'arbre tel que l'âge ni d'éviter parfaitement les phénomènes d'alternance. Dans la réalité, la production de l'arbre est donc limitée par l'interaction entre les facteurs limitants ('limiting-factors') que sont sa structure, son âge et sa génétique (voir Etat de l'art –

Partie I.1). Ainsi pour une année productive et une situation pédoclimatique données, et pour une certaine structure de l'arbre, la production est impactée par le climat annuel (stresses abiotiques durant le cycle) et les pratiques culturales visant à augmenter le nombre de fruits et/ou à contrôler les ravageurs. On appelle donc la production atteignable (P_{att}), la production maximale atteinte pour une structure d'arbre donnée. L'écart de production (ou production gap, P_g) est donc la différence entre P_{att} et la production réelle (P_a) et s'explique par les facteurs de réduction ('reducing-factors') qui sont les évènements climatiques annuels, les pratiques et les stresses biotiques.

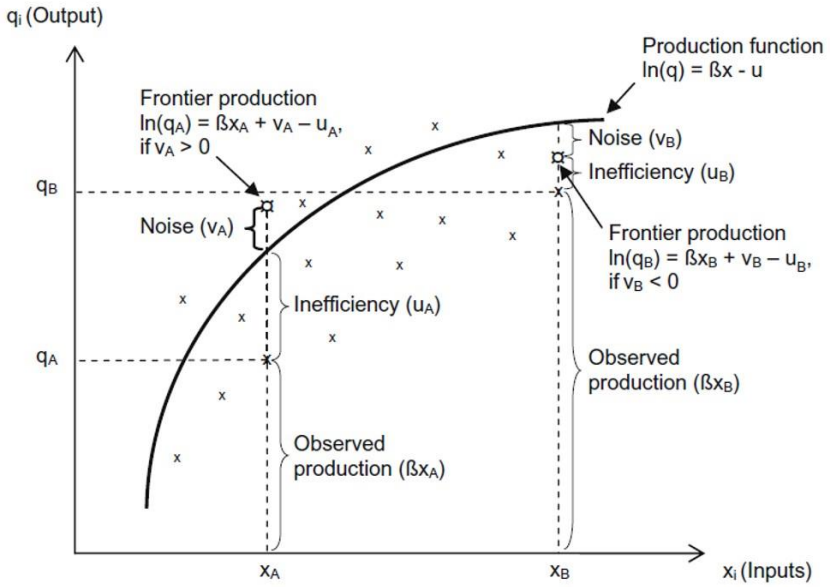


Fig. 83. Principe de l'analyse de frontière stochastique SFA. Les croix (x) représentent la production observée (P_a). La courbe est l'estimation de la frontière qui correspond au lissage de la production atteignable (P_{att}) estimée. Pour chaque valeur de P_a, l'écart à la courbe de frontière permet d'estimer le production gap (P_g) (figure extraite de Neumann et al. 2010).

Conceptualiser les différents niveaux de la production de l'arbre (Fig. 83 et Chapitre 1 – Partie II) et définir P_g ont permis d'adapter les méthodes d'analyses employées dans l'étude du yield gap (analyse de données et courbes frontières). Nous avons opté pour l'utilisation de l'analyse de frontière stochastique (SFA, Battese and Coelli 1995) qui permet de calculer la production atteignable pour une combinaison de variables d'entrées (Fig. 83). Ici, les variables d'entrées sont représentées par les facteurs limitant la production (facteurs endogènes : structure, âge et génétique) et qui définissent P_{att}, c'est-à-dire la variété de l'arbre, sa structure et son âge. Contrairement à l'analyse par courbes enveloppes (ou 'boundary line analysis', Affholder et al. 2013), le SFA est une méthode qui prend en compte plusieurs variables d'entrées (Neumann et al. 2010 ; Silva et al. 2017). En effet, la construction de courbes enveloppes permet uniquement de comparer des variables deux à deux (ex : le rendement en fonction des précipitations, Affholder et al. 2013). À l'inverse, le SFA permet une véritable modélisation de P_{att} en fonction de la combinaison des facteurs limitants (Neumann et al. 2010). Ainsi le modèle SFA peut être vu comme deux régressions

successives (Battese and Coelli 1995) : la première permettant d'estimer P_{att} en fonction d'une combinaison de variables d'entrées, la seconde permettant d'étudier l'effet de variables exploratoires sur la différence entre P_{att} et P_a, c'est-à-dire l'écart de production P_g.

Analyse des écarts de production et limites méthodologiques

Dans notre cas d'étude, l'estimation de P_a (en nombre de fruits) et de la structure individuelle des arbres nous ont d'abord permis de calculer la charge en fruit (*fl*, 'fruit load'). La charge est définie comme le P_a divisé par le volume du houppier (estimé par drone) et exprime le nombre de fruits produits par unité de biomasse (voir Discussion - Partie III.2). En développant des modèles SFA pour chaque variété, nous avons mis en évidence les différences de performances du manguier en fonction de sa variété et du système de culture. Les deux variétés floridiennes ('Kent' et 'Keitt') atteignent des écarts de production entre 41 et 52% par rapport à leur P_{att} (Fig. 58, p. 118). Cependant, la variété Keitt atteint une production (P_a = 262 fruits) et une charge (*fl* = 7 fruits.m^{-3}) plus élevées en moyenne comparée à Kent (*fl* = 5 fruits.m^{-3}) et ce quel que soit le système de culture. À l'opposée, la variété locale 'Boucodiékhal' atteint un écart plus élevé situé à 67% de sa production atteignable (P_{att} = 408 fruits). La variété locale a ainsi la plus faible charge (*fl* = 1.5 fruits.m^{-3}) alors que ces arbres ont les volumes les plus élevés. La comparaison des systèmes de culture a montré des résultats plus contrastés. Les arbres du système intensif obtiennent les mêmes écarts de production que les arbres du système diversifié (P_g = 43%). Ces écarts sont plus faibles comparé au système extensif (P_g = 72%) (Fig. 58, p. 118). Cependant, les productions atteignables sont beaucoup plus élevées dans les vergers diversifiés (P_{att} = 357 fruits) que dans les extensifs et les intensifs (P_{att} = 280 fruits). Ainsi, malgré un plus fort écart de production, la production des arbres dans les vergers diversifiés est plus élevée que dans les vergers intensifs. Par conséquent, les deux systèmes atteignent des charges en fruit significativement identique (*fl* = 6 fruits.m^{-3}).

Les sorties du modèle SFA donne également la significativité de l'effet de chaque facteur de réduction (environnement, pratiques et structure du verger) sur les écarts de production (Tableau 4, p.119). L'environnement semble impacter fortement l'efficacité de production de la variété 'Kent' et peut être un des facteurs explicatifs de sa variabilité interannuelle de production. Les pratiques culturales et la structure du verger ont des effets contradictoires selon les variétés considérées. Par exemple, l'utilisation d'un système de goutte à goutte semble réduire le P_g de 'Kent' alors qu'elle semble l'augmenter chez 'Keitt' et n'a pas d'effet sur 'Boucodiékhal' (Tableau 4, p.119). En réalité, les variables de pratiques sont fortement corrélées entre elles. Les niveaux de pratiques les plus élevés (forte intensité d'application de fertilisant, de produits phytosanitaires, etc.) s'observent uniquement dans les vergers intensifs contenant majoritairement des variétés 'Kent'. Tandis que la variété 'Boucodiékhal' n'est pas présente sous toutes les modalités de pratiques. De plus, les niveaux de pratiques ne sont pas clairement quantifiés. En effet, du fait de la faible

représentativité spatiale des données d'enquêtes, nous avons dû catégoriser les variables de pratiques pour obtenir une donnée par verger. Catégoriser les variables rend leur interprétation plus difficile notamment dans notre cas où nous avons choisi d'intégrer dans le modèle SFA un nombre important de variables à la fois quantitatives et qualitatives. À l'avenir, un choix plus pertinent et moins abondant des variables exploratoires à intégrer dans le SFA (par exemple, en sélectionnant les variables par une analyse de composantes principales, ACP) devrait permettre de clarifier les effets des pratiques sur les écarts de rendement et d'éviter les effets d'autocorrélations.

Néanmoins, cette étude a permis d'évaluer les potentiels de production du manguier dans la zone d'étude. Même dans les situations à fort niveau d'intrants (vergers intensifs), en moyenne 30% de la production atteignable de l'arbre n'est pas atteinte. Ceci peut s'expliquer par les phénomènes d'alternances ou bien les différences de production qui peuvent s'observer au sein même des vergers intensifs. Même si ce groupe reste plus homogène que les deux autres systèmes, des disparités existent notamment au niveau des pratiques et de la densité de plantation qui peuvent impacter la production individuelle de l'arbre (Menzel and Le Lagadec 2017). Nous avons également mis en évidence que les productions atteignables du manguier sont les plus élevées dans les vergers diversifiés. Le P_g de 43% dans ces vergers montre qu'une augmentation de la production est encore possible. Ainsi, en imaginant un abaissement du P_g à 30% de la production atteignable, la production moyenne des arbres augmenterai de 26 fruits par arbre. Ramené à l'échelle du verger, (en utilisant la densité de plantation moyenne de 122 arbres.ha^{-1}) cette diminution du P_g pourrait conduire à une production supplémentaire de 3 100 fruits par hectare en moyenne (environ 1.5 tonnes). Enfin, il est intéressant de souligner le fort écart de production chez la variété locale 'Boucodiéhkal'. Alors que cette variété possède une capacité de production importante, son faible entretien dans la totalité de la zone d'étude (Rey et al. 2007 ; Grechi et al. 2013) en fait une variété avec un fort potentiel de production sous exploité.

b. Étude du rendement des vergers

Diagnostic agronomique régional

En Afrique de l'Ouest, le manque d'informations sur les vergers est une réelle contrainte pour pouvoir étudier les facteurs explicatifs du rendement et comparer les systèmes de culture. Dans cette étude, les méthodes de caractérisation des vergers par drone nous ont permis d'obtenir une quantité importante d'information sur la biomasse de manguiers (densité de plantation, aire moyenne du houppier), la structure des vergers (composition et configuration des arbres) et les rendements. Ces derniers ont pu être évalués de manière précise pendant les deux années d'étude (2017 et 2018), en utilisant la méthodologie drone détaille dans la Partie I du Chapitre 2. S'ajoute à cela les informations sur les pratiques culturales liées au manguier

(obtenues par enquêtes) et les données de température obtenues à l'aide de capteurs. Enfin, l'accès à des bases de données en ligne permet de caractériser l'environnement climatique (WorldClim 2, Fick and Hijmans 2017) et pédologique (AfSoilGrids250m, Hengl et al. 2015).

Afin de pouvoir évaluer l'effet de ces différents facteurs sur le rendement de mangues dans les trois systèmes de culture, nous avons utilisé la méthodologie du diagnostic agronomique régional (voir Chapitre 2 – Partie III) développé par Doré et al. (2008). Cette méthodologie pose les bases pour l'étude de l'élaboration du rendement à l'échelle de la parcelle ou de l'exploitation (Doré et al. 2008). Elle a été largement utilisée en grandes cultures (e.g., Valantin-Morison et al. 2007 ; Doré et al. 2011) et sur quelques cultures semi-pérennes comme la banane (Chopin and Blazy 2013) et le cacao (Jagoret et al. 2017). À notre connaissance, le diagnostic agronomique régional n'a jamais été utilisé pour comparer différents systèmes de culture à base de pérennes. Trois variables de productivité du verger ont été évaluées : la production moyenne des arbres (kg), le rendement du verger ($kg.ha^{-1}$) et le fruit load du verger ($kg.m^{-3}$). Contrairement à l'échelle arbre, le taux de charge ('fruit load') a été calculé ici comme le rendement divisé par la surface totale de houppier du verger c'est à dire la somme des surfaces de cylindre (i.e., biomasse productive) par unité de surface (voir Discussion - Partie III.2).

Impact de la biomasse de manguier sur la productivité du verger

Une première analyse (analyse de composantes principales, ACP) nous a permis de montrer l'effet de la densité de plantation, de la surface de houppier des arbres et du verger (par hectare) sur les variables de productivité. Nous avons démontré que, dans les vergers intensifs de la zone d'étude, le rendement est principalement expliqué par la densité de plantation. En effet, dans ces vergers la surface maximale atteinte par les arbres reste inférieure à 50 m² sans doute du fait des pratiques de taille et de la densité de plantation. Les vergers les plus productifs sont donc les vergers avec la plus forte densité de plantation (qui augmente la surface totale de biomasse productive par rapport à la surface individuelle). Dans la région des Niayes, la densité de plantation du système intensif reste plus faible (maximum observé à 400 $arbres.ha^{-1}$) que dans certains systèmes (au Brésil par exemple) où la densité peut dépasser les 1 000 $arbres.ha^{-1}$ (Menzel and Le Lagadec 2017). Ainsi, la densité observée dans le système des Niayes n'est pas limitante sur la production. En effet, le rendement de plantations en pleine production atteint un palier aux alentours de 800 $arbres.ha^{-1}$ (Nath et al. 2007 cité dans Menzel and Le Lagadec 2017). À ces densités, le rendement du verger peut atteindre les 55 $t.ha^{-1}$ contre 35 $t.ha^{-1}$ pour le maximum mesuré dans notre zone d'étude.

À l'inverse, les vergers diversifiés sont les vergers avec la plus faible densité en manguier (maximum de 200 $arbres.ha^{-1}$) mais une plus forte production moyenne ce qui leur permet d'atteindre des niveaux raisonnables de rendement (7 $t.ha^{-1}$ en moyenne). Ce rendement reste plus élevé que ceux des vergers

extensifs et aussi élevé que dans certains vergers intensifs (ceux à faible densité et les jeunes plantations). La diversification de ces vergers permet d'atteindre des densités de plantation, toutes espèces confondues, aussi élevées que dans les vergers intensifs. Ceci relève d'une stratégie différente des producteurs qui cherchent à diversifier leur source de revenu avec les principales spéculations s'écoulant sur le marché local et régional : les mangues, les agrumes et les productions maraichères (oignon, carotte, tomate, choux, hibiscus) (Fall and Fall 2001). Cependant, à l'heure actuelle peu de données existent pour évaluer l'importance économique du manguier dans ces systèmes diversifiés.

Effets des pratiques et de la diversité

Les équilibres différents entre la densité de plantation et la biomasse productive justifient la nécessité d'utiliser une variable indépendante pour comparer les performances de vergers. Cette étude nous a permis de mettre en évidence l'effet des variables de pratiques et des variables de diversité sur la charge en fruits. Les vergers à forte intensité de pratiques (les vergers intensifs) ont une charge en fruits plus élevée. Du fait de la faible qualité des données sur les pratiques et de l'autocorrélation entre les variables, il est difficile d'identifier exactement quelle(s) pratique(s) a/ont un effet sur la productivité. Sur l'analyse factorielle des données multiples (AFDM), les variables d'irrigation, de fertilisation et d'application des produits phytosanitaires sont celles qui ressortent le plus. On sait également que la disponibilité en eau et la fertilisation, qui sont des facteurs limitants la production de mangues (Crane et al. 2009), sont des stress récurant dans la zone des Niayes (Bationo et al. 2007 ; Fare et al. 2017) bien que peu d'études n'aient été menées à ce jour sur les stresses impactant le manguier dans la zone. Sur variété 'Kent', les travaux de Diatta (2018), ont permis de quantifier la variabilité de la qualité des fruits dans des vergers où les conditions hydriques étaient différentes (Diatta 2018). Cependant, une analyse approfondie de l'effet des pratiques d'irrigation, mais également de fertilisation, sur la quantité et la qualité de la production de mangues reste nécessaire dans les Niayes.

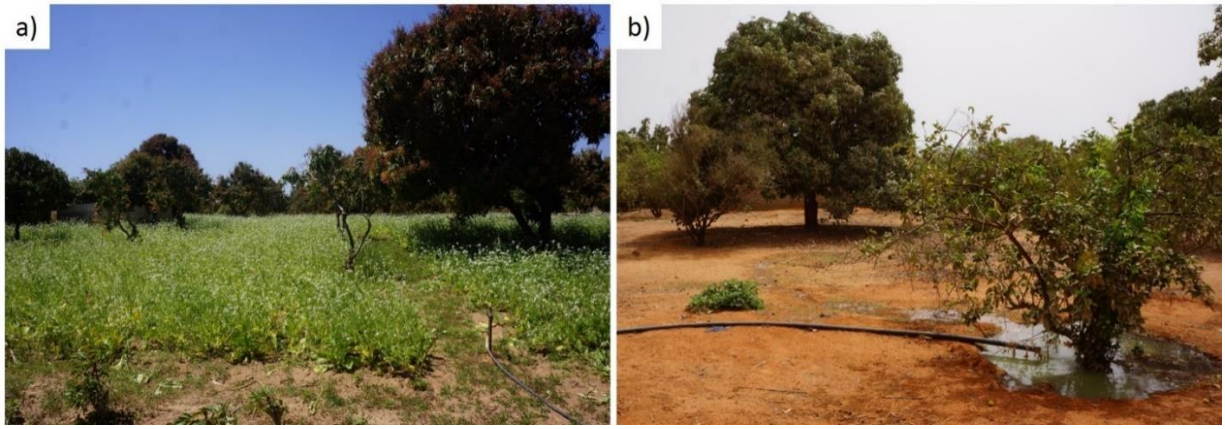


Fig. 84. Dans les vergers diversifiés, le manguier n'est pas irrigué mais profite de l'eau a) d'irrigation du maraîchage ou (b) des agrumes à proximité (source : J. Sarron).

L'effet de la diversité spécifique du verger (caractérisée par l'indice de Shanon, *SHDI*) sur le rendement, la production et la charge en fruits du verger a également été démontré. Au sein du système diversifié, qui est assez homogène en termes de pratiques culturales, les vergers les plus diversifiés (SHDI élevé) ont des performances significativement plus élevées. On peut donc émettre l'hypothèse qu'une plus grande diversité dans ce système de production conduit à une augmentation de la productivité en mangue du verger. Plusieurs facteurs peuvent expliquer cet effet de la diversité. Premièrement, le maraichage, plus largement utilisé dans les vergers diversifiés, pourrait augmenter l'humidité relative du sol et profiter au manguier qui est indirectement irrigué (Rathore et al. 2013). Cet effet peut aussi s'appliquer pour les agrumes qui sont le plus souvent irrigués et dont l'eau d'irrigation peut profiter aux manguiers à proximité (Fig. 84). Deuxièmement, l'augmentation les effets de la biodiversité dans les vergers favorise certaines fonctions favorables (services) au rendement (Demestihas et al. 2017). Par exemple, augmenter la biodiversité peut avoir un effet sur la pression des ravageurs (Simon et al. 2010). Cependant, en Afrique de l'Ouest le principal ravageur de la mangue, la mouche des fruits, n'a qu'un impact très limité sur le nombre de fruits produits par l'arbre et va plutôt poser des problèmes de qualité (Vayssières et al. 2008). Enfin, la diversité en variété de mangue peut améliorer la productivité du manguier. En effet, la présence d'une auto-incompatibilité pollinique a été démontrée chez plusieurs variétés du manguier : la fertilité et le nombre de fruits sont plus élevés en cas de pollinisation croisée entre deux variétés que dans les cas d'auto pollinisation (Pérez et al. 2016).

Limites de l'étude : vers la prise en compte de la qualité et du poids des fruits

Dans cette étude et dans la thèse en générale, nous nous sommes focalisés uniquement sur les données de rendement plutôt que sur le calibre et la qualité des fruits. Or dans la zone, la diversité des conduites culturales et des systèmes de culture se reflètent sur la qualité du fruit (Diatta 2018). Les productions ne sont pas écoulées de la même manière selon les systèmes et les variétés. Les exploitations commerciales sont sensibles à la qualité des fruits et au calibre qui déterminent la valeur de leur production pour le marché européen. Les données de qualité et de maturité sont donc très importantes pour ces vergers qui tentent de maitriser ces facteurs (apport d'intrants et suivi intensif des parcelles). À l'inverse, les petits exploitants contrôlent plus difficilement la qualité (accès aux moyens de luttes contre la mouche des fruits limités notamment) et le calibre (principalement du fait des contraintes liées à l'accès à l'eau) des mangues. Cependant, ces exploitants cultivent plusieurs variétés qui leur permettent d'atteindre le marché local voire des marchés de niche (mangue biologique, mangues séchées, confiture…). Dans ce diagnostic agronomique nous n'avons pas pris en compte ces éléments. L'étude de la qualité et du calibre de la mangue permettrait de se focaliser sur les facteurs agronomiques et environnementaux qui impactent la qualité et le calibre des fruits. D'autant plus que la méthodologie du diagnostic agronomique régional a déjà été utilisée avec succès pour décrire les facteurs impactant la qualité des cultures (Valantin-Morison et al. 2007; Doré et al. 2008).

Par exemple, pour la production de banane, Chopin et Blazy (2013) ont analysé à la fois le rendement brut (en $t.ha^{-1}$) et le rendement commercial (en €) incluant la qualité du fruit. Les prochaines études sur mangues devront s'intéresser à quantifier le calibre (taille du fruit) et la qualité des fruits (forme, aspect, couleur… mais aussi qualité intrinsèque du fruit comme sa teneur en sucre).

III.2. La variable de charge en fruit ('*fruit load*')

Pour étudier les variabilités de production à l'échelle de l'arbre et à l'échelle du verger nous avons introduit, en plus des variables de production et de rendement, la variable de charge en fruits (*fl*). Pour analyser la variabilité de production à l'échelle de l'arbre (production gap), nous avons calculé la charge (notée *fl_volume* pour cette discussion) comme étant la production (nombre de fruits) divisée par le volume de l'arbre. Ce volume est estimé par drone et correspond à la somme des hauteurs projetées du nuage de point multipliée par la résolution du pixel (6.40 $cm.pixel^{-1}$). L'utilisation de la variable volume reste pertinente au regard de la résolution des images drones et des capacités de la photogrammétrie pour mesurer précisément les structure de l'arbre (Wallace et al. 2016). Ainsi *fl_volume* a pour unité un nombre de fruits par unité de volume ($fruits.m^{-3}$).

Pour l'analyse des rendements à l'échelle du verger nous avons utilisé une autre mesure de la charge en fruit (notée *fl_surface*). En effet, le *fl_surface* a été mesuré à partir de la surface totale productive du houppier. Cette surface est calculée en utilisant la formule de la surface du cylindre ($A = \pi * h * D$, avec h et D la hauteur et le diamètre du houppier, respectivement) et en utilisant les paramètres calculés par le drone. Ainsi, *fl_surface* correspond au rendement du verger divisé par la surface de houppier rapporté à la surface (voir Chapitre 2 – Partie III). Le *fl_surface* a donc pour unité un poids de fruits par unité de surface ($kg.m^{-2}$).

À l'origine, la charge en fruit est plutôt utilisée pour des travaux à l'échelle de la branche (Léchaudel et al. 2002). Elle se mesure en générale sur l'unité de croissance comme le nombre de fruits produits par unité de surface foliaire. La charge permet d'évaluer la quantité de fruits produits relativement à la capacité photosynthétique (exprimée à travers la surface foliaire) de la branche. Pour cette thèse, nous avons adapté cette variable en l'appliquant à nos deux échelles d'étude (arbre et verger). Il s'agissait de pouvoir exprimer la productivité de l'arbre et du verger relativement au volume (mesurée précisément par drone) disponible pour la photosynthèse. Intuitivement, *fl_volume* apparait pertinent et permet de relier la productivité des arbres de différentes variétés et cultivés dans différentes conditions. L'analyse de la corrélation entre *fl_volume* et le production gap met en évidence la forte interaction entre les deux variables (Tableau 20). La variable *fl_volume* est donc intéressante pour étudier la productivité des arbres. Cependant, de tests de corrélation montre que *fl_volume* est corrélé au volume de l'arbre pour les variétés 'Kent' et 'Boucodiékhal' (r de Spearman = -0.14 et -0.29 respectivement).

La charge en fruit estimée par rapport au volume des manguiers ne peut donc pas être utilisée pour évaluer et comparer la productivité à l'échelle des vergers car elle est corrélée au volume. Nous avons donc opté, pour l'étude des vergers, pour une variable de charge évaluée par rapport à la surface de houppier (*fl_surface*). Ce choix est justifié par le fait que le manguier est un arbre à croissance sympodiale où les fruits sont produits à l'extérieur du houppier (Normand and Lauri 2018). Or, si on modélise la structure d'un manguier par une forme 3D, quel que soit le modèle choisi (cylindre, sphère ou pyramide), l'augmentation du rayon du houppier augmente le volume plus rapidement que la surface. L'intérieur du houppier n'étant pas ou très peu productif, le volume de manguiers croit plus vite que la biomasse efficace à la production à la périphérie. À l'inverse, l'augmentation de la surface augmente d'autant plus le potentiel de biomasse productive. Utiliser la surface « extérieure » du houppier apparait donc comme une alternative intéressante pour calculer la charge en fruit. L'analyse des corrélations entre *fl_surface* et la surface de houppier ne montre aucun lien significatif entre les deux variables (r de Spearman tous non significatifs). Par ailleurs, la variable *fl_surface* est mieux corrélée à la production de l'arbre que *fl_volume* (Tableau 20) ce qui traduit donc bien une meilleure efficacité pour évaluer les performances. On retrouve également une corrélation plus forte entre le *fl_surface* et le production gap qu'entre le *fl_volume* et le production gap (Tableau 20).

Ainsi, même si *fl_volume* et *fl_surface* sont deux variables qui traduisent la même notion, c'est-à-dire une production rapportée à une biomasse, la physiologie du manguier (croissance sympodiale, Normand and Lauri 2018) fait que l'utilisation de *fl_surface* est préférable à celle de *fl_volume.* L'utilisation de *fl_surface* est donc recommandée pour étudier les performances du manguier à l'échelle de l'arbre et du verger.

Tableau 20. Corrélation de Spearman (rho) entre la production du manguier, l'écart de production (production gap) et les deux variables de charge estimées en fonction du volume (*fl_volume*) ou bien de la surface (*fl_surface*) du houppier.

	Production gap	*fl_volume*	*fl_surface*
Production manguier (fruit.arbre^{-1})	-0.67***	0.39***	0.69***
Production gap (%)		-0.77***	-0.88***
fl_volume			0.87***

IV. Perspectives de la thèse

IV.1. De nouvelles méthodologies pour l'étude des cultures fruitières

À l'échelle de l'arbre, nous avons pu mettre en place un outil d'estimation de la production (combinant analyse d'images et modélisation) et développé le concept de 'production gap' pour l'étude des variabilités de production dans différents systèmes de culture. Ces méthodologies restent perfectibles (voir Discussion - Partie I et III) mais elles sont applicables pour la caractérisation d'autres espèces d'arbres fruitiers. La

définition des production gaps permet de pallier au manque actuel de modèles de culture pour les arbres fruitiers (Normand et al. 2015) pour décrypter les variables endogènes et exogènes impactant la production. Dans cette thèse, nous avons été limités par la qualité des données de pratiques agricoles, notamment chez les petits producteurs qui ne réalisent pas un suivi précis de leurs pratiques (Carletto et al. 2015).

L'étude des production gaps n'est possible que si un nombre important de données à l'échelle de l'arbre (production et structure) est acquise. C'est pourquoi, elle est complémentaire de l'outil d'estimation de la production par analyse d'image. À l'heure actuelle, les performances toujours plus pointues des réseaux de neurones permettent d'accéder rapidement à de la donnée fiable (Kamilaris and Prenafeta-Boldú 2018). Dans notre cas, cet outil sert à l'estimation de la production pour étudier les variabilités de productivité. Ainsi, l'acquisition rapide et fiable de données sur le manguier permettrait un suivi temporel et spatial (répartition des fruits dans le houppier) de l'établissement du fruit. Les réseaux de neurones offrent également des perspectives pour l'estimation du calibre et de la maturité des fruits (Wang et al. 2017). Pour cette dernière, des projets de recherche sont actuellement en cours au sein de l'UPR HortSys du CIRAD afin d'optimiser l'estimation de la maturité des fruits de manière non destructive. Enfin, le réseau a montré des capacités pour détecter des fruits à des stades plus précoces (stade 5 cm). La détection des inflorescences a été testée mais son efficacité reste contrastée par rapport aux résultats de la littérature (Wang et al. 2016). L'utilisation de ces données reste à explorer pour, à terme, mettre au point un outil de prédiction de récolte (trois mois avant la première récolte).

À l'échelle du verger, nous avons pu analyser les variabilités intra- et inter-vergers (Fig. 32, Chapitre 2) afin de dégager quelques hypothèses sur la variabilité du rendement dans la zone d'étude. L'étape suivante sera l'application des modèles de rendement des vergers à l'échelle du bassin de production. Pour ce faire, il faudra établir un lien entre les données caractérisant les vergers par télédétection et les rendements observés/estimés sur le terrain. Une cartographie précise des types de vergers dans la zone d'étude devra être élaborée via des outils de télédetection spatiale (classification orientée-objet d'images à très haute résolution spatiale et spectrale). Cette cartographie servira de base pour la conception d'un système d'estimations spatialisées de rendement de la culture du manguier à l'échelle régionale. Ce couplage des données de terrain avec une carte d'occupation des sols issue de la télédétection permettra de caractériser la variabilité des rendements du manguier à l'échelle régionale en prenant en compte la diversité des systèmes de production composant les agrosystèmes de la zone étudiée (Annexe V).

IV.2. Développement d'un outil de gestion et de partage de données pour informer les filières

Les informations sur la production de mangues en Afrique de l'Ouest, à l'instar des toutes les productions horticoles de la région, sont manquantes. Les résultats des travaux de cette thèse (méthodes et algorithmes)

sont à la base de la mise en place d'un outil numérique d'acquisition participative, de gestion et de partage de données de production fruitières (mangue dans un premier temps) pour faciliter la structuration des filières en Afrique de l'Ouest. Ainsi, le Data Manager PixFruit Mangue (Fig. 85), repose sur trois composantes distinctes et complémentaires : 1) une base de données (BDD) pour centraliser l'information enrichie par des données multi-sources, 2) une application mobile « de prise d'informations au champ » pour alimenter la BDD et informer en temps réel le producteur sur le niveau de production de son verger et de ses arbres, et 3) un géoportail Web permettant la consultation des données traitées par l'ensemble des acteurs de la filière.

L'implémentation sur smartphone du réseau de neurones pré-entrainé permettra aux producteurs de quantifier de manière objective, précise et instantanée les productions des manguiers par analyses d'images photographiques. Ces données alimenteront les modèles d'extrapolation des rendements aux différentes échelles spatiales sur une base de données en ligne qui sera source d'informations et d'analyses pour des acteurs de la filière mangue tels que les collectivités territoriales, les agences de développement, les politiques agricoles nationales et les instituts responsables des statistiques agricoles. De plus, cette base de données en évolution constante sera source d'intérêt pour la communauté scientifique toujours en demande d'information précise et couvrant des extensions spatiotemporelles importantes. Le data manager PixFruit Mangue permettra l'orientation de décisions directement aux producteurs, la visualisation d'information de production à diverses échelles spatiotemporelles pour informer les acteurs de la filière, la mise en relation de ces acteurs (chercheurs compris). En amont, les producteurs trouveront une solution efficace et adaptée pour la mesure de leurs productions en champ mais aussi pour les aider dans leurs prises de décision (organisation de la récolte, négociation du prix bord champ, logistique, etc.).

Tout d'abord, déployé dans sa version prototype sur le bassin de production des Niayes (zone d'étude de la thèse), PixFruit Mangue sera étendu progressivement à d'autres bassins dans lesquels de premiers essais sont en cours actuellement (Sud Sénégal, Nord Côte d'Ivoire, La Réunion et Madagascar, etc.). Le projet intégrera progressivement les contraintes métier et les attentes de chaque catégorie d'acteur afin de définir le contour d'un outil de gestion réellement « utile » pour la filière. De plus, cet outil évoluera au fil des recherches menées par les partenaires scientifiques du projet (en partie grâce aux données acquises par PixFruit Mangue). L'amélioration des algorithmes utilisés permettra le perfectionnement des outils développés, pour proposer de nouveaux services comme l'évaluation du calibre et de la maturité du fruit par analyses d'image, ou la prévision de récolte à court terme (quelques mois). Enfin, à terme, cette version PixFruit Mangue pourra être adaptée à d'autres spéculations fruitières comme la banane, le litchi, l'avocat, l'olivier ou les agrumes.

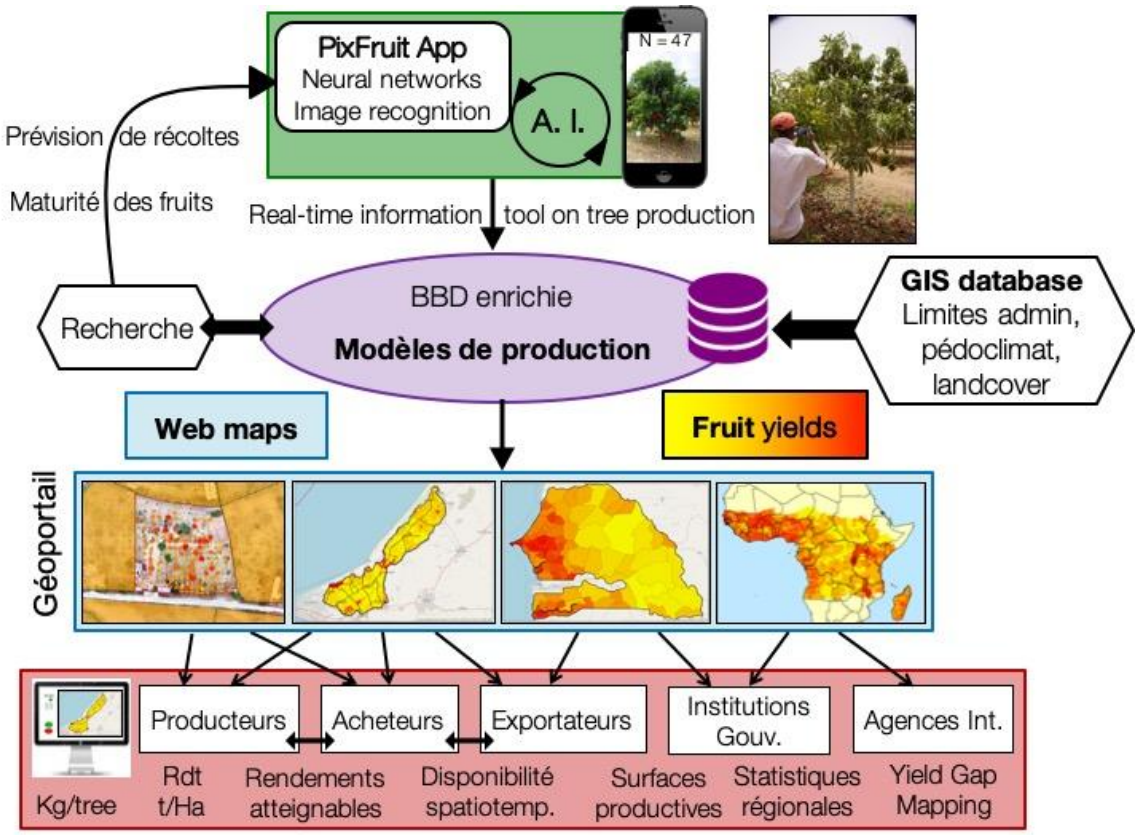


Fig. 85. Schéma fonctionnel de l'outil PixFruit *Mangue*

Conclusion générale

Les objectifs principaux de cette thèse étaient d'estimer et de spatialiser les rendements du manguier en Afrique de l'Ouest. Pour ces travaux de thèse nous avons opté pour des solutions frugales dans le but d'estimer la production des arbres et de caractériser les vergers. Ces solutions sont basées sur l'analyse d'images RVB dont le coût d'acquisition et de traitement est plus faible comparé à d'autre capteurs (ex : LiDAR). Nous avons montré que l'utilisation d'un capteur RVB embarqué par drone permet d'obtenir facilement et de manière fiable une carte d'occupation des sols et la structure des arbres dans des vergers de complexité différente. Dans ces travaux nous avons mis en évidence l'intérêt et les limites de l'utilisation des réseaux de neurones pour l'analyse d'image. Ces algorithmes atteignent de très bonnes performances dans la détection d'objets dont l'aspect varie fortement. Cependant, d'autres critères restent à être évalués tels que le coût de l'entrainement (vérités terrain et annotations notamment), ou les ressources et capacités de calcul nécessaires. L'utilisation des réseaux de neurones pour la segmentation et la classification d'images aériennes constituent un domaine de recherche récent qui reste à être appliqué à des images drones. Cette thèse a permis d'aboutir à un outil pour l'estimation de la production qui ouvre de nouvelles perspectives pour les filières fruitières en Afrique de l'Ouest en favorisant l'acquisition et l'échange d'informations.

Grâce à l'acquisition de données, différentes méthodes d'analyse ont permis d'étudier les facteurs de variation des rendements de mangues. Le concept de '*yield gap*' a été adapté à l'échelle de l'arbre pour définir le 'production gap' c'est-à-dire l'écart entre la production atteignable de l'arbre (i.e., définie par la variété, la structure et l'âge) et sa production réelle. Cette méthode d'analyse permet d'estimer les effets de différents facteurs causant les écarts. L'analyse des production gaps a permis de montrer que les écarts de production étaient les plus élevés dans les vergers extensifs et représentaient 66% de la production atteignable contre 38 et 28% respectivement pour les vergers diversifiés et intensifs. Le grand écart de production observé dans les vergers extensifs pourrait venir des faibles doses d'intrants appliquées dans ce type de verger. Le diagnostic agronomique régional a également été utilisé pour comparer les performances entre les vergers et déterminer les facteurs impactant le rendement. Dans les vergers intensifs, la densité de plantation semble être un facteur majeur positivement corrélé au rendement. Dans les vergers diversifiés, la production de l'arbre compense la faible densité de plantation en manguiers et permet d'obtenir des niveaux de rendement intermédiaires (~ 7 $t.ha^{-1}$).

Utilisée aux deux échelles d'étude (arbre et verger), la variable de taux de charge permet de comparer les types de systèmes de culture en comparant la production par unité de biomasse plutôt que par hectare. Grâce aux outils d'analyse d'image développés ici, le taux de charge a été facilement calculé. Aux différentes

échelles, le taux de charge est impacté par les variables environnementales, les pratiques et la structure du verger (composition et configuration).

Les résultats de cette thèse mettent en avant les possibilités méthodologiques pour l'estimation et l'étude des rendements des arbres fruitiers à différentes échelles (Fig. 32, p.66). Cependant, les hypothèses sur l'effet des pratiques culturales et de la diversité cultivée sur le rendement doivent encore être approfondies. Par exemple, augmenter la qualité des données ou réaliser un suivi plus régulier d'un échantillon plus réduit sont des pistes pour améliorer la compréhension de l'élaboration du rendement du manguier en Afrique de l'Ouest. Cette thèse n'a pas non plus abordé les aspects qualité des fruits et revenu économique dans les différents systèmes de culture. Ces travaux constituent cependant une première étape vers l'évaluation agronomique et économique des systèmes à base d'arbres fruitiers en Afrique de l'Ouest.

RÉFÉRENCES

Affholder F, Poeydebat C, Corbeels M, et al (2013) The yield gap of major food crops in family agriculture in the tropics: Assessment and analysis through field surveys and modelling. Field Crops Research 143:106–118. https://doi.org/10.1016/j.fcr.2012.10.021

African Development Bank (2018) West Africa Economic Outlook 2018: Macroeconomic developments and poverty inequality and employment. Labor markets and jobs. African Development Bank, Abidjan, Côte d'Ivoire

Aggelopooulou K, Castrignanò A, Gemtos T, Benedetto DD (2013) Delineation of management zones in an apple orchard in Greece using a multivariate approach. Computers and Electronics in Agriculture 90:119–130. https://doi.org/10.1016/j.compag.2012.09.009

Aggelopoulou AD, Bochtis D, Fountas S, et al (2011) Yield prediction in apple orchards based on image processing. Precision Agric 12:448–456. https://doi.org/10.1007/s11119-010-9187-0

Aggelopoulou KD, Wulfsohn D, Fountas S, et al (2010) Spatial variation in yield and quality in a small apple orchard. Precision Agric 11:538–556. https://doi.org/10.1007/s11119-009-9146-9

Aguiar LAA (2009) Impact de la variabilité climatique récente sur les écosystèmes des Niayes du Sénégal entre 1950 et 2004. Thèse de doctorat, Université du Québec à Montréal

Akohoue F, Segnon AC, Achigan-Dako EG (2018) Diversity in smallholder citrus orchards and cultivation bottlenecks: research avenues for improved production in Benin, West Africa. Ex Agric 54:641–654. https://doi.org/10.1017/S001447971700028X

Altieri MA (2002) Agroecology: the science of natural resource management for poor farmers in marginal environments. Agriculture, Ecosystems & Environment 93:1–24. https://doi.org/10.1016/S0167-8809(02)00085-3

Amorós López J, Izquierdo Verdiguier E, Gómez Chova L, et al (2011) Land cover classification of VHR airborne images for citrus grove identification. ISPRS Journal of Photogrammetry and Remote Sensing 66:115–123. https://doi.org/10.1016/j.isprsjprs.2010.09.008

Anderson NT, Underwood JP, Rahman MM, et al (2019) Estimation of fruit load in mango orchards: tree sampling considerations and use of machine vision and satellite imagery. Precision Agric 20:823–839. https://doi.org/10.1007/s11119-018-9614-1

Andriamampianina L, Temple L, de Bon H, et al (2018) Évaluation pluri-critères de l'agriculture biologique en Afrique subsaharienne par élicitation probabiliste des connaissances d'experts. Cah Agric 27:45002. https://doi.org/10.1051/cagri/2018030

ANSD (2017) Bulletin mensuel des statistiques économiques de juillet 2017. Dakar, Sénégal

Arbonnier M (2009) Arbres, arbustes et lianes des zones seches d'Afrique de l'Ouest, Éditions Quae. Versailles, France

Baatz M, Schäpe A (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl J, Blaschke T, Griesbner G (eds) Angewandte Geographische Informations-Verarbeitung, XII. Wichmann Verlag, Karlsruhe, Germany, pp 12–23

Bargoti S, Underwood J (2017) Deep fruit detection in orchards. In: Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, pp 3626–3633

Bationo A, Kihara J, Vanlauwe B, et al (2007) Soil organic carbon dynamics, functions and management in West African agro-ecosystems. Agricultural Systems 94:13–25. https://doi.org/10.1016/j.agsy.2005.08.011

Battese GE, Coelli TJ (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20:325–332. https://doi.org/10.1007/BF01205442

Bayala J, Sanou J, Bazié HR, et al (2019) Regenerated trees in farmers' fields increase soil carbon across the Sahel. Agroforest Syst. https://doi.org/10.1007/s10457-019-00403-6

Birdal AC, Avdan U, Türk T (2017) Estimating tree heights with images from an unmanned aerial vehicle. Geomatics, Natural Hazards and Risk 8:1144–1156. https://doi.org/10.1080/19475705.2017.1300608

Blaschke T, Hay GJ, Kelly M, et al (2014) Geographic Object-Based Image Analysis – Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87:180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Bompard JM (2009) Taxonomy and Systematic. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 97–169

Borchert R (1991) Growth periodicity and dormancy. In: Raghavendra AS (ed) Physiology of trees. Wiley, New York, pp 221–245

Burke M, Lobell DB (2017) Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proceedings of the National Academy of Sciences 114:2189–2194

Camara C, Bourgeois R, Jahel C (2019) Anticiper l'avenir des territoires agricoles en Afrique de l'Ouest : le cas des Niayes au Sénégal. Cah Agric 28:12. https://doi.org/10.1051/cagri/2019012

Capelli M, Lauri P-É, Normand F (2016) Deciphering the Costs of Reproduction in Mango – Vegetative Growth Matters. Front Plant Sci 7:. https://doi.org/10.3389/fpls.2016.01531

Carletto C, Jolliffe D, Banerjee R (2015) From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. The Journal of Development Studies 51:133–148. https://doi.org/10.1080/00220388.2014.968140

Carlier C, Etchepare M, Ceccon J-F, et al (1992) Efficacy of massive oral doses of retinyl palmitate and mango (Mangifera indica L.) consumption to correct an existing vitamin A deficiency in Senegalese children. Br J Nutr 68:529–540. https://doi.org/10.1079/BJN19920110

Cassman KG (1999) Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. Proceedings of the National Academy of Sciences 96:5952–5959. https://doi.org/10.1073/pnas.96.11.5952

Cemansky R (2015) Africa's Indigenous Fruit Trees: A Blessing in Decline. Environmental Health Perspectives 123:. https://doi.org/10.1289/ehp.123-A291

Challinor A, Wheeler T, Garforth C, et al (2007) Assessing the vulnerability of food crop systems in Africa to climate change. Climatic Change 83:381–399. https://doi.org/10.1007/s10584-007-9249-0

Chopin P, Blazy J-M (2013) Assessment of regional variability in crop yields with spatial autocorrelation: Banana farms and policy implications in Martinique. Agriculture, Ecosystems & Environment 181:12–21. https://doi.org/10.1016/j.agee.2013.09.001

CILSS (2016) Ladscapes of West Africa - A window on a changing world, Comité Permanent Inter-états de Lutte contre la Sécheresse dans le Sahel (CILSS). Ouagadougou, Burkina Faso

Coleman JS, McConnaughay KDM, Ackerly DD (1994) Interpreting phenotypic variation in plants. Trends in Ecology & Evolution 9:187–191. https://doi.org/10.1016/0169-5347(94)90087-6

Connor DJ, Gómez-del-Campo M, Rousseaux MC, Searles PS (2014) Structure, management and productivity of hedgerow olive orchards: A review. Scientia Horticulturae 169:71–93. https://doi.org/10.1016/j.scienta.2014.02.010

Costes E, Lauri PÉ, Regnard JL (2010) Analyzing Fruit Tree Architecture: Implications for Tree Management and Fruit Production. In: Janick J (ed) Horticultural Reviews. John Wiley & Sons, Inc., Oxford, UK, pp 1–61

Courtin F, Guengant J-P (2011) Un siècle de peuplement en Afrique de l'Ouest. Nat Sci Soc 19:256–265. https://doi.org/10.1051/nss/2011146

Crane JH, Salazar-Garcia S, Lin T-S, et al (2009) Crop Production: Management. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 432–483

CSAO/OCDE (2006) Atlas de l'intégration régionale en Afrique de l'Ouest - Les bassins fluviaux transfrontaliers. CEDEAO

Dambreville A, Lauri P-E, Trottier C, et al (2013a) Deciphering structural and temporal interplays during the architectural development of mango trees. Journal of Experimental Botany 64:2467–2480. https://doi.org/10.1093/jxb/ert105

Dambreville A, Normand F, Lauri P-E (2014) Alternating patterns of mango - quantification at the growth unit level of four cultivars. Acta Horticulturae 277–284. https://doi.org/10.17660/ActaHortic.2014.1058.32

Dambreville A, Normand F, Lauri P-É (2013b) Plant growth co-ordination in natura: a unique temperature-controlled law among vegetative and reproductive organs in mango. Functional Plant Biology 40:280. https://doi.org/10.1071/FP12243

Dandois J, Olano M, Ellis E (2015) Optimal Altitude, Overlap, and Weather Conditions for Computer Vision UAV Estimates of Forest Structure. Remote Sensing 7:13895–13920. https://doi.org/10.3390/rs71013895

de Bie CAJM (2004) The yield gap of mango in Phrao, Thailand, as investigated through comparative performance evaluation. Scientia Horticulturae 102:37–52. https://doi.org/10.1016/j.scienta.2003.11.017

de Bon H, Huat J, Parrot L, et al (2014) Pesticide risks from fruit and vegetable pest management by small farmers in sub-Saharan Africa. A review. Agron Sustain Dev 34:723–736. https://doi.org/10.1007/s13593-014-0216-7

Demestihas C, Plénet D, Génard M, et al (2017) Ecosystem services in orchards. A review. Agron Sustain Dev 37:12. https://doi.org/10.1007/s13593-017-0422-1

Diatta J (2018) Vers la maîtrise du rendement et de la qualité de la mangue au Sénégal : déterminants de la floraison, fructification, et impact de pratiques culturales et de facteurs environnementaux. Thèse de doctorat, Université Cheikh Anta Diop

Díaz-Varela R, de la Rosa R, León L, Zarco-Tejada P (2015) High-Resolution Airborne UAV Imagery to Assess Olive Tree Crown Parameters Using 3D Photo Reconstruction: Application in Breeding Trials. Remote Sensing 7:4213–4232. https://doi.org/10.3390/rs70404213

Dong B, Sutton R (2015) Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall. Nature Clim Change 5:757–760. https://doi.org/10.1038/nclimate2664

Doré T, Clermont-Dauphin C, Crozat Y, et al (2008) Methodological progress in on-farm regional agronomic diagnosis. A review. Agron Sustain Dev 28:151–161. https://doi.org/10.1051/agro:2007031

Doré T, Makowski D, Malézieux E, et al (2011) Facing up to the paradigm of ecological intensification in agronomy: Revisiting methods, concepts and knowledge. European Journal of Agronomy 34:197–210. https://doi.org/10.1016/j.eja.2011.02.006

Dupraz C (1999) Adequate design of control treatments in long term agroforestry experiments with multiple objectives. Agroforestry Systems 60:35–48. https://doi.org/10.1007/978-94-017-0679-7_2

Dury S, Vall É, Imbernon J (2017) Production agricole et sécurité alimentaire en Afrique de l'Ouest. Cah Agric 26:61001. https://doi.org/10.1051/cagri/2017047

Duteurtre G, Dièye Faye M, Dièye PN (eds) (2010) L'agriculture sénégalaise à l'épreuve du marché. ISRA, Dakar

Fall M (1986) Environnements sédimentaires quaternaires et actuels des tourbières des Niayes de la Grande Côte du Sénégal. Thèse de doctorat, Université Cheikh Anta Diop

Fall ST, Fall AS (2001) Cités horticoles en sursis? L'agriculture urbaine dans les grandes Niayes au Sénégal. International Development Research Centre, Ottawa

FAO (2008) An introduction to the basic concepts of food security. Rome, Italy

FAO (2018) The State of Food Security and Nutrition in the World 2018. Building climate resilience for food security and nutrition. FAO, Rome

FAO (ed) (2015) Meeting the 2015 international hunger targets: taking stock of uneven progress. FAO, Rome, Italy

FAOSTAT (2017) Food and Agriculture Organization of the United Nations. Rome, Italy. http://www.fao.org/faostat/en/#home

Fare Y, Dufumier M, Loloum M, et al (2017) Analysis and Diagnosis of the Agrarian System in the Niayes Region, Northwest Senegal (West Africa). Agriculture 7:59. https://doi.org/10.3390/agriculture7070059

Félix GF, Scholberg JMS, Clermont-Dauphin C, et al (2018) Enhancing agroecosystem productivity with woody perennials in semi-arid West Africa. A meta-analysis. Agron Sustain Dev 38:57. https://doi.org/10.1007/s13593-018-0533-3

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. Int J Climatol 37:4302–4315. https://doi.org/10.1002/joc.5086

Fifanou VG, Ousmane C, Gauthier B, Brice S (2011) Traditional agroforestry systems and biodiversity conservation in Benin (West Africa). Agroforest Syst 82:1–13. https://doi.org/10.1007/s10457-011-9377-4

Gidoin C, Babin R, Bagny Beilhe L, et al (2014) Tree Spatial Structure, Host Composition and Resource Availability Influence Mirid Density or Black Pod Prevalence in Cacao Agroforests in Cameroon. PLoS ONE 9:e109405. https://doi.org/10.1371/journal.pone.0109405

Godfray HCJ, Beddington JR, Crute IR, et al (2010) Food Security: The Challenge of Feeding 9 Billion People. Science 327:812–818. https://doi.org/10.1126/science.1185383

Godfray HCJ, Garnett T (2014) Food security and sustainable intensification. Phil Trans R Soc B 369:20120273. https://doi.org/10.1098/rstb.2012.0273

Gómez MI, Barrett CB, Raney T, et al (2013) Post-green revolution food systems and the triple burden of malnutrition. Food Policy 42:129–138. https://doi.org/10.1016/j.foodpol.2013.06.009

Gómez-del-Campo M, Centeno A, Connor DJ (2009) Yield determination in olive hedgerow orchards. I. Yield and profiles of yield components in north - south and east - west oriented hedgerows. Crop Pasture Sci 60:434. https://doi.org/10.1071/CP08252

Gongal A, Amatya S, Karkee M, et al (2015) Sensors and systems for fruit detection and localization: A review. Computers and Electronics in Agriculture 116:8–19. https://doi.org/10.1016/j.compag.2015.05.021

González A, Lu P, Müller W (2004) Effect of pre-flowering irrigation on leaf photosynthesis, whole-tree water use and fruit yield of mango trees receiving two flowering treatments. Scientia Horticulturae 102:189–211. https://doi.org/10.1016/j.scienta.2003.12.011

Grechi I, Sane CAB, Diame L, et al (2013a) Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68:447–466. https://doi.org/10.1051/fruits/2013094

Grechi I, Sane CAB, Diame L, et al (2013b) Mango-based orchards in Senegal: diversity of design and management patterns. Fruits 68:447–466. https://doi.org/10.1051/fruits/2013094

Grossman YL, DeJong TM (1994) PEACH: A simulation model of reproductive and vegetative growth in peach trees. Tree Physiology 14:329–345. https://doi.org/10.1093/treephys/14.4.329

Hallé F, Oldeman RAA, Tomlinson PB (1978) Tropical trees and forests: an architectural analysis. Springer, Berlin

Held IM, Delworth TL, Lu J, et al (2005) Simulation of Sahel drought in the 20th and 21st centuries. Proceedings of the National Academy of Sciences 102:17891–17896. https://doi.org/10.1073/pnas.0509057102

Hengl T, Heuvelink GBM, Kempen B, et al (2015) Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS ONE 10:e0125814. https://doi.org/10.1371/journal.pone.0125814

Hernández Delgado PM, Aranguren M, Reig C, et al (2011) Phenological growth stages of mango (Mangifera indica L.) according to the BBCH scale. Scientia Horticulturae 130:536–540. https://doi.org/10.1016/j.scienta.2011.07.027

Himmelstein J, Ares A, Gallagher D, Myers J (2017) A meta-analysis of intercropping in Africa: impacts on crop yield, farmer income, and integrated pest management effects. International Journal of Agricultural Sustainability 15:1–10. https://doi.org/10.1080/14735903.2016.1242332

Hollinger F, Staatz JM (2015) Agricultural growth in West Africa: market and policy drivers. Rome, Italy

Isaac ME, Gordon AM, Thevathasan N, et al (2005) Temporal changes in soil carbon and nitrogen in west African multistrata agroforestry systems: a chronosequence of pools and fluxes. Agroforest Syst 65:23–31. https://doi.org/10.1007/s10457-004-4187-6

Issaka RN, Masunaga T, Kosaki T, Wakatsuki T (1996) Soils of inland valleys of West Africa: General fertility parameters. Soil Science and Plant Nutrition 42:71–80. https://doi.org/10.1080/00380768.1996.10414690

Iyer CPA, Schnell RJ (2009) Breeding and Genetics. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 97–169

Jagoret P, Michel I, Ngnogué HT, et al (2017) Structural characteristics determine productivity in complex cocoa agroforestry systems. Agron Sustain Dev 37:60. https://doi.org/10.1007/s13593-017-0468-0

Jagoret P, Michel-Dounias I, Malézieux E (2011) Long-term dynamics of cocoa agroforests: a case study in central Cameroon. Agroforest Syst 81:267–278. https://doi.org/10.1007/s10457-010-9368-x

Jagoret P, Ngnogue HT, Malézieux E, Michel I (2018) Trajectories of cocoa agroforests and their drivers over time: Lessons from the Cameroonian experience. European Journal of Agronomy 101:183–192. https://doi.org/10.1016/j.eja.2018.09.007

Jamnadass RH, Dawson IK, Franzel S, et al (2011) Improving livelihoods and nutrition in sub-Saharan Africa through the promotion of indigenous and exotic fruit production in smallholders' agroforestry systems: a review. int forest rev 13:338–354. https://doi.org/10.1505/146554811798293836

Jeuffroy M-H, Casadebaig P, Debaeke P, et al (2014) Agronomic model uses to predict cultivar performance in various environments and cropping systems. A review. Agron Sustain Dev 34:121–137. https://doi.org/10.1007/s13593-013-0170-9

Jiménez AR, R. Ceres, J. L. Pons (2000) A survey of computer vision methods for locating fruit on trees. Transactions of the ASAE 43:1911–1920. https://doi.org/10.13031/2013.3096

Jiménez-Brenes FM, López-Granados F, de Castro AI, et al (2017) Quantifying pruning impacts on olive tree architecture and annual canopy growth by using UAV-based 3D modelling. Plant Methods 13:55. https://doi.org/10.1186/s13007-017-0205-3

Juel A, Groom GB, Svenning J-C, Ejrnæs R (2015) Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. International Journal of Applied Earth Observation and Geoinformation 42:106–114. https://doi.org/10.1016/j.jag.2015.05.008

Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: A survey. Computers and Electronics in Agriculture 147:70–90. https://doi.org/10.1016/j.compag.2018.02.016

Ke Y, Quackenbush LJ (2011) A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. International Journal of Remote Sensing 32:4725–4747. https://doi.org/10.1080/01431161.2010.494184

Koirala A, Walsh KB, Wang Z, McCarthy C (2019a) Deep learning – Method overview and review of use for fruit detection and yield estimation. Computers and Electronics in Agriculture 162:219–234. https://doi.org/10.1016/j.compag.2019.04.017

Koirala A, Walsh KB, Wang Z, McCarthy C (2019b) Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO.' Precision Agric 20:1107–1135. https://doi.org/10.1007/s11119-019-09642-0

Koning N, van Ittersum MK (2009) Will the world have enough to eat? Current Opinion in Environmental Sustainability 1:77–82. https://doi.org/10.1016/j.cosust.2009.07.005

Kouassi A, Assamoi P, Bigot S, et al (2010) Étude du climat ouest-africain à l'aide du modèle atmosphérique régional M.A.R. Climatologie. https://doi.org/10.4267/climatologie.445

Läderach P, Martinez-Valle A, Schroth G, Castro N (2013) Predicting the future climatic suitability for cocoa farming of the world's leading producer countries, Ghana and Côte d'Ivoire. Climatic Change 119:841–854. https://doi.org/10.1007/s10584-013-0774-8

Lalel HJD, Singh Z, Tan SC, Agustí M (2003) Maturity stage at harvest affects fruit ripening, quality and biosynthesis of aroma volatile compounds in 'Kensington Pride' mango. The Journal of Horticultural Science and Biotechnology 78:225–233. https://doi.org/10.1080/14620316.2003.11511610

Lauri P-E, Trottier C (2004) Patterns of size and fate relationships of contiguous organs in the apple (Malus domestica) crown. New Phytol 163:533–546. https://doi.org/10.1111/j.1469-8137.2004.01136.x

Lauri P-E, Willaume M, Larrive G, Lespinasse J-M (2004) The concept of centrifugal training in apple aimed at oprimizing the relationship between growth and fruiting. Acta Hortic 35–42. https://doi.org/10.17660/ActaHortic.2004.636.3

Léchaudel M, Genard M, Lescourret F, et al (2002) Leaf-to-fruit ratio affects water and dry-matter content of mango fruit. The Journal of Horticultural Science and Biotechnology 77:773–777. https://doi.org/10.1080/14620316.2002.11511571

Léchaudel M, Lopez-Lauri F, Vidal V, et al (2013) Response of the physiological parameters of mango fruit (transpiration, water relations and antioxidant system) to its light and temperature environment. Journal of Plant Physiology 170:567–576. https://doi.org/10.1016/j.jplph.2012.11.009

Léchaudel M, Vercambre G, Lescourret F, et al (2007) An analysis of elastic and plastic fruit growth of mango in response to various assimilate supplies. Tree Physiology 27:219–230. https://doi.org/10.1093/treephys/27.2.219

Linker R (2017) A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation. Precision Agriculture 18:59–75. https://doi.org/10.1007/s11119-016-9467-4

Linker R, Cohen O, Naor A (2012) Determination of the number of green apples in RGB images recorded in orchards. Computers and Electronics in Agriculture 81:45–57. https://doi.org/10.1016/j.compag.2011.11.007

Lisein J, Michez A, Claessens H, Lejeune P (2015) Discrimination of Deciduous Tree Species from Time Series of Unmanned Aerial System Imagery. PLoS ONE 10:e0141006. https://doi.org/10.1371/journal.pone.0141006

Lobell DB, Cassman KG, Field CB (2009) Crop Yield Gaps: Their Importance, Magnitudes, and Causes. Annual Review of Environment and Resources 34:179–204. https://doi.org/10.1146/annurev.environ.041008.093740

Lowder SK, Skoet J, Raney T (2016) The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. World Development 87:16–29. https://doi.org/10.1016/j.worlddev.2015.10.041

Lu P, K. Chacko E, L. Bithell S, et al (2013) Productivity is negatively related to shoot growth across five mango cultivars in the seasonally wet-dry tropics of northern Australia. Fruits 68:279–289. https://doi.org/10.1051/fruits/2013074

Ma L, Cheng L, Li M, et al (2015) Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. ISPRS Journal of Photogrammetry and Remote Sensing 102:14–27. https://doi.org/10.1016/j.isprsjprs.2014.12.026

Malézieux E (2012) Designing cropping systems from nature. Agron Sustain Dev 32:15–29. https://doi.org/10.1007/s13593-011-0027-z

Malézieux E, Crozat Y, Dupraz C, et al (2009) Mixing plant species in cropping systems: concepts, tools and models. A review. Agron Sustain Dev 29:43–62. https://doi.org/10.1051/agro:2007057

Mann KK, Schumann AW, Obreza TA (2011) Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data. Precision Agric 12:457–472. https://doi.org/10.1007/s11119-010-9189-y

Maschler J, Atzberger C, Immitzer M (2018) Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data. Remote Sensing 10:1218. https://doi.org/10.3390/rs10081218

Mathur P, Ramírez Villegas J, Jarvis A (2012) The impacts of climate change on tropical and subtropical horticultural production. In: Tropical fruit tree species and climate change. Biodiversity International, New Delhi, India, pp 27–44

Matson PA (1997) Agricultural Intensification and Ecosystem Properties. Science 277:504–509. https://doi.org/10.1126/science.277.5325.504

Mbow C, Van Noordwijk M, Luedeling E, et al (2014a) Agroforestry solutions to address food security and climate change challenges in Africa. Current Opinion in Environmental Sustainability 6:61–67. https://doi.org/10.1016/j.cosust.2013.10.014

Mbow C, van Noordwijk M, Prabhu R, Simons T (2014b) Knowledge gaps and research needs concerning agroforestry's contribution to Sustainable Development Goals in Africa. Current Opinion in Environmental Sustainability 6:162–170. https://doi.org/10.1016/j.cosust.2013.11.030

Menzel CM, Le Lagadec MD (2017) Can the productivity of mango orchards be increased by using high-density plantings? Scientia Horticulturae 219:222–263. https://doi.org/10.1016/j.scienta.2016.11.041

Menzel CM, Le Lagadec MD (2014) Increasing the productivity of avocado orchards using high-density plantings: A review. Scientia Horticulturae 177:21–36. https://doi.org/10.1016/j.scienta.2014.07.013

Mitra SK (2016) Mango production in the world – present situation and future prospect. Acta Hortic 287–296. https://doi.org/10.17660/ActaHortic.2016.1111.41

Monselise SP, Goldschmidt EE (1982) Alternate Bearing in Fruit Trees. In: Janick J (ed) Horticultural Reviews. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 128–173

Monteiro F, Catarino L, Batista D, et al (2017) Cashew as a High Agricultural Commodity in West Africa: Insights towards Sustainable Production in Guinea-Bissau. Sustainability 9:1666. https://doi.org/10.3390/su9091666

Mukherjee SK, Litz RE (2009) Botany and Importance. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 1–18

Nevalainen O, Honkavaara E, Tuominen S, et al (2017) Individual Tree Detection and Classification with UAV-Based Photogrammetric Point Clouds and Hyperspectral Imaging. Remote Sensing 9:185. https://doi.org/10.3390/rs9030185

Ngo Bieng MA, Gidoin C, Avelino J, et al (2013) Diversity and spatial clustering of shade trees affect cacao yield and pathogen pressure in Costa Rican agroforests. Basic and Applied Ecology 14:329–336. https://doi.org/10.1016/j.baae.2013.03.003

Noble AE, Rosenstock TS, Brown PH, et al (2018) Spatial patterns of tree yield explained by endogenous forces through a correspondence between the Ising model and ecology. Proc Natl Acad Sci USA 115:1825–1830. https://doi.org/10.1073/pnas.1618887115

Nordey T, Joas J, Davrieux F, et al (2014a) Non-destructive prediction of color and pigment contents in mango peel. Scientia Horticulturae 171:37–44. https://doi.org/10.1016/j.scienta.2014.01.025

Nordey T, Léchaudel M, Saudreau M, et al (2014b) Model-Assisted Analysis of Spatial and Temporal Variations in Fruit Temperature and Transpiration Highlighting the Role of Fruit Development. PLoS ONE 9:e92532. https://doi.org/10.1371/journal.pone.0092532

Normand F, Bello AKP, Trottier C, Lauri P-É (2009) Is axis position within tree architecture a determinant of axis morphology, branching, flowering and fruiting? An essay in mango. Annals of Botany 103:1325–1336. https://doi.org/10.1093/aob/mcp079

Normand F, Lauri P-É (2018) Advances in understanding mango tree growth and canopy development. In: Galán Saúco V, Lu P (eds) Achieving sustainable cultivation of mangoes. Burleigh Dodds Science Publishing, Sawston, Cambridge, UK; Philadelphia, PA, pp 87–120

Normand F, Lauri P-E, Legave J-M (2015) Climate change and its probable effects on mango production and cultivation. Acta Hortic 21–31. https://doi.org/10.17660/ActaHortic.2015.1075.1

Nuske S, Wilshusen K, Achar S, et al (2014) Automated Visual Yield Estimation in Vineyards: Automated Visual Yield Estimation. Journal of Field Robotics 31:837–860. https://doi.org/10.1002/rob.21541

Pádua L, Vanko J, Hruška J, et al (2017) UAS, sensors, and data processing in agroforestry: a review towards practical applications. International Journal of Remote Sensing 38:2349–2391. https://doi.org/10.1080/01431161.2017.1297548

Panthou G, Lebel T, Vischel T, et al (2018) Rainfall intensification in tropical semi-arid regions: the Sahelian case. Environ Res Lett 13:064013. https://doi.org/10.1088/1748-9326/aac334

Payne A, Walsh K, Subedi P, Jarvis D (2014) Estimating mango crop yield using image analysis using fruit at 'stone hardening' stage and night time imaging. Computers and Electronics in Agriculture 100:160–167. https://doi.org/10.1016/j.compag.2013.11.011

Payne AB, Walsh KB, Subedi PP, Jarvis D (2013) Estimation of mango crop yield using image analysis – Segmentation method. Computers and Electronics in Agriculture 91:57–64. https://doi.org/10.1016/j.compag.2012.11.009

Peeters A, Zude M, Käthner J, et al (2015) Getis–Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. Computers and Electronics in Agriculture 111:140–150. https://doi.org/10.1016/j.compag.2014.12.011

Peña MA, Liao R, Brenning A (2017) Using spectrotemporal indices to improve the fruit-tree crop classification accuracy. ISPRS Journal of Photogrammetry and Remote Sensing 128:158–169. https://doi.org/10.1016/j.isprsjprs.2017.03.019

Pérez V, Herrero M, Hormaza JI (2016) Self-fertility and preferential cross-fertilization in mango (Mangifera indica). Scientia Horticulturae 213:373–378. https://doi.org/10.1016/j.scienta.2016.10.034

Perry EM, Dezzani RJ, Seavert CF, Pierce FJ (2010) Spatial variation in tree characteristics and yield in a pear orchard. Precision Agric 11:42–60. https://doi.org/10.1007/s11119-009-9113-5

Pingali PL (2012) Green Revolution: Impacts, limits, and the path ahead. Proceedings of the National Academy of Sciences 109:12302–12308. https://doi.org/10.1073/pnas.0912953109

Qureshi WS, Payne A, Walsh KB, et al (2016) Machine vision for counting fruit on mango tree canopies. Precision Agriculture. https://doi.org/10.1007/s11119-016-9458-5

Rahman M, Robson A, Bristow M (2018) Exploring the Potential of High Resolution WorldView-3 Imagery for Estimating Yield of Mango. Remote Sensing 10:1866. https://doi.org/10.3390/rs10121866

Ramírez F, Davenport TL (2010) Mango (Mangifera indica L.) flowering physiology. Scientia Horticulturae 126:65–72. https://doi.org/10.1016/j.scienta.2010.06.024

Ramírez F, Davenport TL, Fischer G (2010) The number of leaves required for floral induction and translocation of the florigenic promoter in mango (Mangifera indica L.) in a tropical climate. Scientia Horticulturae 123:443–453. https://doi.org/10.1016/j.scienta.2009.10.005

Rathore AC, Saroj PL, Lal H, et al (2013) Performance of mango based agri-horticultural models under rainfed situation of Western Himalaya, India. Agroforest Syst 87:1389–1404. https://doi.org/10.1007/s10457-013-9646-5

Ray DK, Gerber JS, MacDonald GK, West PC (2015) Climate variation explains a third of global crop yield variability. Nat Commun 6:5989. https://doi.org/10.1038/ncomms6989

Reddy YTN, Kurian RM, Ramachander PR, et al (2003) Long-term effects of rootstocks on growth and fruit yielding patterns of 'Alphonso' mango (Mangifera indica L.). Scientia Horticulturae 97:95–108. https://doi.org/10.1016/S0304-4238(02)00025-0

Rey J-Y, Dia ML (2010) Mangue : des vergers villageois aux nouvelles plantations d'exportation. In: Duteurtre G, Dièye Faye M, Dièye PN (eds) L'agriculture sénégalaise à l'épreuve du marché. ISRA, Dakar, pp 257–279

Rey J-Y, Diallo TM, Vannière H, et al (2007) The mango in French-speaking West Africa: varieties and varietal composition of the orchards. Fruits 62:57–73. https://doi.org/10.1051/fruits:2006051

Rey J-Y, Diallo TM, Vannière H, et al (2004) La mangue en Afrique de l'Ouest francophone. Fruits 59:121–129. https://doi.org/10.1051/fruits:2004012

Roudier P, Sultan B, Quirion P, Berg A (2011) The impact of future climate change on West African crop yields: What does the recent literature say? Global Environmental Change 21:1073–1083. https://doi.org/10.1016/j.gloenvcha.2011.04.007

Sall SM, Viltard A, Sauvageot H (2007) Rainfall distribution over the Fouta Djallon — Guinea. Atmospheric Research 86:149–161. https://doi.org/10.1016/j.atmosres.2007.03.008

Sánchez PA (2010) Tripling crop yields in tropical Africa. Nature Geosci 3:299–300. https://doi.org/10.1038/ngeo853

Sané CAB (2016) Relations entre le fonctionnement des agrosystèmes à base de manguiers et les dégâts des termites (Termitidae : Isoptera) dans les régions de Thiès et de Dakar, au Sénégal. Thèse de doctorat, Université Cheikh Anta Diop

Schaffer B, Urban L, Lu P, Whiley AW (2009) Ecophysiology. In: Litz RE (ed) The mango: botany, production and uses, 2nd ed. CABI, Wallingford, UK ; Cambridge, MA, pp 170–209

Schroth G, Läderach P, Martinez-Valle AI, et al (2016) Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. Science of The Total Environment 556:231–241. https://doi.org/10.1016/j.scitotenv.2016.03.024

Sibhatu KT, Krishna VV, Qaim M (2015) Production diversity and dietary diversity in smallholder farm households. Proc Natl Acad Sci USA 112:10657–10662. https://doi.org/10.1073/pnas.1510982112

Simon S, Bouvier J-C, Debras J-F, Sauphanor B (2010) Biodiversity and pest management in orchard systems. A review. Agron Sustain Dev 30:139–152. https://doi.org/10.1051/agro/2009013

Spreer W, Ongprasert S, Hegele M, et al (2009) Yield and fruit development in mango (Mangifera indica L. cv. Chok Anan) under different irrigation regimes. Agricultural Water Management 96:574–584. https://doi.org/10.1016/j.agwat.2008.09.020

Stadlmayr B, Charrondière UR, Eisenwagen S, et al (2013) Nutrient composition of selected indigenous fruits from sub-Saharan Africa: Nutrient composition of selected indigenous fruits from sub-Saharan Africa. J Sci Food Agric 93:2627–2636. https://doi.org/10.1002/jsfa.6196

Stein M, Bargoti S, Underwood J (2016) Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors 16:1915. https://doi.org/10.3390/s16111915

Stephan J, Sinoquet H, Dones N, et al (2008) Light interception and partitioning between shoots in apple cultivars influenced by training. Tree Physiology 28:331–342. https://doi.org/10.1093/treephys/28.3.331

Sukhvibul N, Whiley AW, Smith MK, et al (1999) Effect of temperature on inflorescence and floral development in four mango (Mangifera indica L.) cultivars. Scientia Horticulturae 82:67–84. https://doi.org/10.1016/S0304-4238(99)00041-2

Sultan B, Defrance D, Iizumi T (2019) Evidence of crop production losses in West Africa due to historical global warming in two crop models. Sci Rep 9:12834. https://doi.org/10.1038/s41598-019-49167-0

Ternoy J, Poublanc C, Diop M (2006) Mango value chain Senegal: analysis and strategic framework for sub-sector growth initiatives. USAID

Thompson SK (2013) Adaptive web sampling in ecology. Stat Methods Appl 22:33–43. https://doi.org/10.1007/s10260-012-0222-3

Tittonell P, Giller KE (2013) When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. Field Crops Research 143:76–90. https://doi.org/10.1016/j.fcr.2012.10.007

Torres-Sánchez J, López-Granados F, Serrano N, et al (2015) High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. PLOS ONE 10:e0130479. https://doi.org/10.1371/journal.pone.0130479

Toulon N (2018) Deep learning et agriculture - Comprendre les potentiels et les défis à relever. Chaire AgroTIC, Bordeaux, France

United Nations (2019) Department of Economic and Social Affairs, Population Division. In: World Population Prospects 2019, custom data acquired via website. https://population.un.org/wpp/. Accessed 12 Aug 2019

Usha K, Singh B (2013) Potential applications of remote sensing in horticulture—A review. Scientia Horticulturae 153:71–83. https://doi.org/10.1016/j.scienta.2013.01.008

Valantin-Morison M, Meynard J-M, Doré T (2007) Effects of crop management and surrounding field environment on insect incidence in organic winter oilseed rape (Brassica napus L.). Crop Protection 26:1108–1120. https://doi.org/10.1016/j.cropro.2006.10.005

van Ittersum MK, Cassman KG, Grassini P, et al (2013) Yield gap analysis with local to global relevance—A review. Field Crops Research 143:4–17. https://doi.org/10.1016/j.fcr.2012.09.009

van Ittersum MK, van Bussel LGJ, Wolf J, et al (2016) Can sub-Saharan Africa feed itself? Proceedings of the National Academy of Sciences 201610359. https://doi.org/10.1073/pnas.1610359113

Vandermeer J, van Noordwijk M, Anderson J, et al (1998) Global change and multi-species agroecosystems: Concepts and issues. Agriculture, Ecosystems & Environment 67:1–22. https://doi.org/10.1016/S0167-8809(97)00150-3

Vayssières J-F, Korie S, Coulibaly O, et al (2008) The mango tree in central and northern Benin: cultivar inventory, yield assessment, infested stages and loss due to fruit flies (Diptera Tephritidae). Fruits 63:335–348. https://doi.org/10.1051/fruits:2008035

Wairegi LWI, van Asten PJA, Tenywa MM, Bekunda MA (2010) Abiotic constraints override biotic constraints in East African highland banana systems. Field Crops Research 117:146–153. https://doi.org/10.1016/j.fcr.2010.02.010

Wallace L, Lucieer A, Malenovský Z, et al (2016) Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds. 16

Wang N, Jassogne L, van Asten PJA, et al (2015) Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. European Journal of Agronomy 63:1–11. https://doi.org/10.1016/j.eja.2014.11.003

Wang Z, Verma B, Walsh KB, et al (2016) Automated mango flowering assessment via refinement segmentation. In: 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, Palmerston North, New Zealand, pp 1–6

Wang Z, Walsh K, Verma B (2017) On-Tree Mango Fruit Size Estimation Using RGB-D Images. Sensors 17:2738. https://doi.org/10.3390/s17122738

Weinstein BG, Marconi S, Bohlman S, et al (2019) Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sensing 11:1309. https://doi.org/10.3390/rs11111309

Wheeler T, von Braun J (2013) Climate Change Impacts on Global Food Security. Science 341:508–513. https://doi.org/10.1126/science.1239402

Whelan BM, McBratney AB (2000) The "Null Hypothesis" of Precision Agriculture Management. Precision Agriculture 2:265–279

WHO (ed) (2004) Fruit and Vegetables for Health (Workshop), Report of a Joint FAO/WHO Workshop, 1-3 September 2004, Kobe, Japan.

Ye X, Sakai K, Manago M, et al (2007) Prediction of citrus yield from airborne hyperspectral imagery. Precision Agric 8:111–125. https://doi.org/10.1007/s11119-007-9032-2

Zhao W, Du S, Emery WJ (2017) Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. IEEE J Sel Top Appl Earth Observations Remote Sensing 10:3386–3396. https://doi.org/10.1109/JSTARS.2017.2680324

ANNEXES

Annexe I: Fiche d'enquête auprès des producteurs

I- INFORMATIONS GENERALES :

N° Verger : …….
Nom personne enquêtée : …….
Fonction de la personne enquêtée (propriétaire, responsable production, gardien, etc.) :……..
Nom propriétaire du verger :……
Profession propriétaire du verger :…….
Localisation (commune) : …….
Superficie exploitation : …….
Type de production :
Conventionnel **Bio**

Type d'exploitation :
Exportateur/Industriel **Familiale**
Certification (Ecocert, GLOBALG.A.P., etc.) :……..

II- INFORMATION VERGER :

Superficie verger (si différente de superficie exploitation) :……..
Type de limite (mur, clôture, haie, aucune, etc.) :……..
Principales cultures de rentes du verger :……

Détail des productions :

Cultures fruitières

Espèces	Superficie (ha) ou nombre de pieds	Irrigation (oui/non)
Manguier		

Cultures maraîchères et annuelles

Espèces	Superficie (ha) ou % du verger	Irrigation (oui/non)

Elevages

Espèces	Objectifs (vente/autoconsommation)

Rotation culture maraîchère/annuelle type :…….
Déplacement des cultures maraîchère/annuelle dans l'espace : OUI - NON
Système d'irrigation culture maraîchère :
Micro-aspersion Goutte-à-goutte
Jet d'eau

Origine de l'eau :

Puit avec pompe solaire Puit avec motopompe Puit avec seau
Forage SDE Autre :……

Nombre de puit/forage : …….

III- INFORMATION MANGUIER :

Année de plantation/âge des arbres : …….
Porte-greffe : …… **Variétés :** ……
Densité : ……

Taille :

Fréquence taille d'entretien **(**nulle, bisannuelle, annuelle, etc.) **:**……
Date/période taille d'entretien :……..
Taille de régénération : OUI - NON Eclaircissage pré-récolte : OUI - NON
Suppression inflorescence ou petit fruit : OUI - NON

Gestion du sol :

Enherbement verger :

Enherbement Sol nu
Maraîchage total Maraîchage sur certaines zones
Fréquence désherbage / broyage : ……
Travail du sol : OUI – NON Si oui, outils utilisé :……..

Gestion du rang :

Enherbement Désherbage chimique Désherbage mécanique
Fréquence désherbage / broyage : ……

Irrigation du manguier :

Système d'irrigation :

Micro-aspersion Goutte-à-goutte
« Bassines » Pas d'irrigation

Date de la première irrigation : …….. **Date de la dernière irrigation** : …….
Fréquence des apports : …….
Dose par apport : ……. **Quantité totale :** ……
Problèmes liés à la nutrition hydrique du manguier : ……

Fertilisation :

Type (chimique/organique)	Produits/composition	Dose	Localisation apport (rang, inter-rang, foliaire)	Fréquence/date d'apport

Origine des fertilisants organiques :

Produit sur place Acheté Les deux

Gestion des bioagresseurs

Bioagresseurs rencontrés (liste) :
-
-
(Mouches des fruits, thermites, anthracnose, thrips, oiseaux, etc.)

Traitements phytosanitaire :

Date	Bioagresseur ciblé	Produit / Substance active	Dose

Protection des troncs :
Aucune Huile Chaux

Nettoyage des fruits tombés : OUI – NON
Inducteurs floraux : OUI – NON

IV- RECOLTE

Opérateur récolte :
Ouvrier/saisonnier Producteur personnellement
Banabana Exportateur

Suivi du rendement : OUI – NON
Si oui, méthode (balance, caisse, etc.) :…….

Variété	Début récolte	Fin récolte	Nombre de passage	Destination (export, transformation, marché local)	Rendement 2017	Prix de vente
Kent						
Keitt						
BDH						
Autres						

Remarques particulières saison et récoltes 2017 (accidents climatiques, faible rendement, retard floraison, etc…) :

Annexe II: Poster conférence ESA 2018

Estimating and modelling productivity of mango trees from endogenous factors

A methodology set up and validated for African orchards

J. Sarron[1,2,3], C.A.B. Sané[4], J. Diatta[2], E. Faye[1,2,3], E. Malézieux[1,3]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]Centre pour le Développement de l'Horticulture, ISRA, Dakar, Sénégal
[3]Univ Montpellier, F-34090 Montpellier, France
[4]Université Cheikh Anta Diop (UCAD), Dakar, Sénégal

INTRODUCTION

- In West Africa, mango (*Manguifera indica L.*) trees are cultivated in a variety of cropping systems (extensive, diversified and intensive). At the tree scale, yield is highly variable as a result of cultivar, variability in tree structures and cropping practices.
- However, heterogeneity of mango productivity depending on endogenous factors has not been described and quantified yet.
- The aim of this study was to quantify the yield of mango trees and study the effects of endogenous factors (cultivar, load index and structure) on tree production. Models for yield estimation from these factors were fitted and validated.

MATERIALS AND METHODS

1. **Sampling of 209 trees** - depicting in-field structure heterogeneity of three main cultivars: 'Kent', 'Keitt' and 'Boucodiékhal' (BDH) - were imaged using RGB camera. Yield (in kg) was estimated using machine vision system coupled with models (*Sarron et al. 2018*). Actual yield was measured in the field on 64 additional trees by mango experts for model validation.



3 classes load index: "low", "medium" and "high"

2. **Tree load index** (area of visible fruits compared to the overall crown area) was evaluated in the field by visual estimation.

3. **Tree structure** was assessed using 12 variables (length, area, vol.)

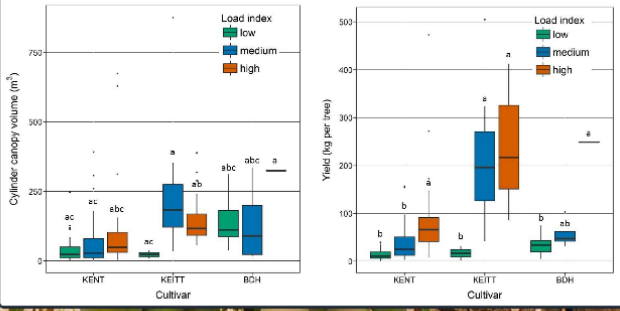


Variable	Assessment
perimeter	In field
tree_height	On images
canopy_height	On images
canopy_width	On images
volume_cylinder area_cylinder	Cylinder model using canopy height and width
volume_sphere_height area_sphere_height	Sphere model using canopy height as radius
volume_sphere_width area_sphere_width	Sphere model using canopy width as radius
volume_ellipse area_ellipse	Ellipse model using canopy width (minor axis) and height (major axis)

4. **Structure effects on yield** were analysed using Pearson correlation tests whereas cultivar and load index effects were analysed using two-way ANOVA with Tukey tests. For each cultivar, model was fitted between yield and independent variables (i.e. load index and structure variable most correlated to yield) and were validated by comparing actual and estimated yield.

RESULTS

- **For all trees, structure variables were all positively correlated with yield** (p-value < 0.001). We focused on cylinder canopy volume which reached the highest correlation (r = 0.76). Cultivar has a significant effect on volume: 'Kent' average volume (63.2 m³) is significantly lower than 'BDH' and 'Keitt' (165.5 m³). Load index was highly correlated to canopy volume for 'Keitt' and 'BDH' cultivars.

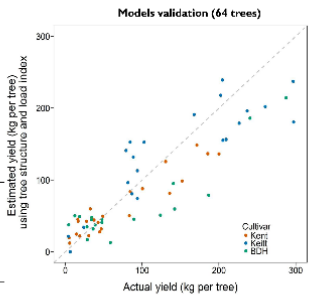


- **Cultivar and load index have significant effects on yield.** Cultivar average estimated yield was 43.5, 206.0, and 55.1kg per tree for 'Kent' 'Keitt' and 'BDH' respectively. For a given cultivar, a higher load index gives significantly higher yield except for Keitt-medium/-high and for Kent-low/-medium. For 'Keitt', increase of load index has no effect on yield as it also reduces canopy volume. 'Kent' trees have too low yield and volume (compared to other cultivars) to display a significant effect of load index.

- Yield predictive models reached R^2 > 0.70. There is no differences between estimated and actual yield (t-test p-value = 0.40).

Performance of built models for yield estimation:

Cultivar	Model variables	R^2	RMSE %
All	*cylinder crown volume, load index*	0.71	0.41
'Kent'	*sphere crown volume, load index*	0.81	0.39
'Keitt'	*canopy height, load index*	0.92	0.32
'BDH'	*cylinder crown volume, load index*	0.90	0.56



CONCLUSION AND PERSPECTIVES

- The yield of three mango cultivars was estimated and the effects of endogenous factors (fruit load, cultivar and structure) on productivity were quantified. Further studies will investigate the effect on yields of other variables (phenological effect, temperature, etc.).
- This work demonstrated that mango tree yield could be accurately estimated from cultivar, load index and one structure variable. Future work will use high resolution UAV cartography to allow a remote measure of tree structure for yield estimation at the orchard scale.

Sarron, J., Sané, C.A.B., Borianne, P., Malézieux, É., Nordey, T., Normand, F., Diatta, P., Niyang, Y., Faye, É. : Is machine learning efficient for mango yield estimation when used under heterogeneous field conditions? In: 30th International Horticultural Congress (IHC), submitted (2018)

Annexe III: Poster PixYield

PixYield – L'analyse d'images photographiques comme outil pour l'estimation de rendement : le cas du manguier au Sénégal

J. Sarron[1,2,3], C.A.B. Sané[4], P. Borianne[3,5], P. Diatta[2], E. Faye[1,2,3]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]Centre pour le Développement de l'Horticulture, ISRA, Dakar, Sénégal
[3]Univ Montpellier, F-34090 Montpellier, France
[4]Université Cheikh Anta Diop (UCAD), Dakar, Sénégal;
[5]CIRAD, UMR AMAP, F-34398 Montpellier, France

INTRODUCTION

- L'estimation **précise** des rendements des cultures pérennes est un enjeu stratégique pour le développement des **filières fruiticoles** dans les pays du Sud, à la fois en terme de **sécurité alimentaire** (autonomie) mais aussi **économique** (maitrise, contrôle et optimisation des volumes produits).
- L'analyse d'image photographique par **machine-learning** est ici testée sur manguier (*Manguifera indica* L.) pour développer un outil opérationnel permettant 1) d'informer les producteurs des **niveaux de production** de leur système de culture et 2) de quantifier la **variabilité spatiotemporelle des rendements** afin d'identifier les facteurs physiologiques, structuraux, agronomiques, et environnementaux explicatifs de cette variabilité.

MÉTHODES

- Suivi photographique de 300 arbres répartis dans:

3 variétés (Kent, Keitt, BDH)



X 2 faces opposées

- À 3 stades phénologiques clés:
  - Floraison ⟶ **Taux floraison**
  - Fruits établis ⟶ **Nb fruits verts**
  - Fruits matures ⟶ **Nb fruits matures**

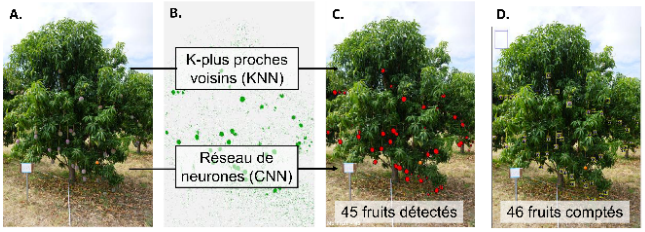
- Élaboration et entrainement des algorithmes de machine-learning:

Apprentissage supervisé à partir d'annotations expertes sur les images (1.2 M de pixels pour KNN, 3000 boîtes pour CNN; voir D.)

Sélection des variables (R, G, B, texture variance, texture moyenne, et Luma) pour KNN et paramétrage du CNN (Non-Maximal-Suppression à 0.25, seuil de sélection à 0.7, 10 000 itérations).

300 arbres X 2 faces X 3 phénologies = **1800 images par an**

- Mise en œuvre pour le comptage d'organes (fruits matures):



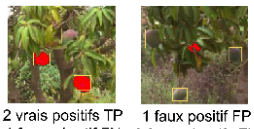
- Évaluation des performances:

Validation sur des annotations expertes (7000 boîtes pour KNN et CNN)

$\text{Precision} = \frac{TP}{TP+FP}$ $\text{Rappel} = \frac{TP}{TP+FN}$

$\text{F1-score} = \frac{2 \times P \times R}{P+R}$ = mesure d'efficacité

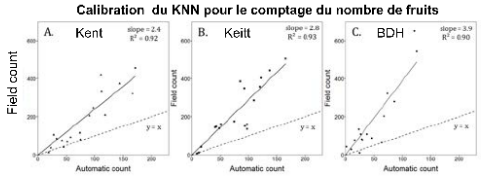
$\text{Accuracy} = \frac{TP}{TP+FP+FN}$ = % de mangues détectées

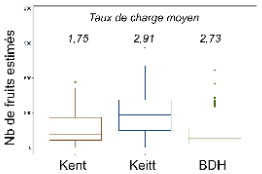


RÉSULTATS et DISCUSSION

- Dans le meilleur des cas (aspect des fruits homogène), le **KNN** obtient un **F1-score de 0.71** et réussi à **détecter 64% des mangues** présentes sur les images. Ces résultats se dégradent fortement lorsqu'il s'agit de détecter des fruits plus hétérogènes – différentes variétés, mangues de taille, forme et couleurs différentes (Sarron *et al.* 2018). Le **réseau de neurones CNN** s'est révélé plus performant pour le comptage de fruits mûrs sur images photographiques, obtenant un **F1-score de 0.95** et **détectant 84% des mangues** visibles sur les images.
- Pour pouvoir utiliser le **machine-learning pour le comptage précis** de mangues (incluant les mangues intra-canopée non visibles sur les images), une **calibration par comptage au champ** par variété est donc nécessaire (Cf. relations linéaires des comptages de fruits ci-contre.)

- Les résultats du KNN sont dépendants de la qualité du jeu d'entrainement. Le CNN, plus robuste, nécessite cependant un jeu d'entrainement important.
- Le KNN calibré a permis d'obtenir le comptage de **48 708 fruits répartis sur 300 arbres**, avec 115, 334, et 173 fruits par arbre en moyenne pour les variétés Kent, Keitt, et BDH respective.





- Le CNN (plus performant et plus rapide: il est capable de **détecter des mangues non repérées par les experts**) est en cours de calibrage statistique pour son utilisation dans la zone des Niayes au Sénégal.

PERSPECTIVES

- Le développement du machine-learning pour l'estimation de la production à l'échelle de l'arbre permettra **l'analyse des paramètres explicatifs de la variabilité des rendements mesurés** (en cours sur 2017-2018) et l'étude des relations entre stades phénologiques (vers des prévisions de récolte?).
- Le déploiement du CNN sous forme d'une **application smartphone** fournirait un outil pertinent pour les producteurs, et ces résultats pourront servir d'inputs à des modèles de rendement à l'échelle du verger pour **étudier la variabilité de production intra et inter-vergers**.

Ce travail a été réalisé dans le cadre de l'Action Incitative CreSi Créativité Scientifique et Innovation du CIRAD et avec l'appui de l'Institut de Convergence Agriculture Numérique #DigitAg (ANR-16-CONV-0004).

Annexe IV: Annexes article Remote Sensing, Chapitre II – Partie I

Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV

Julien Sarron [1,2,3], Éric Malézieux [1,2], Cheikh Amet Bassirou Sané [4] and Émile Faye [1,2,3,*]

[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]HortSys, University Montpellier, CIRAD, F-34090 Montpellier, France
[3]Centre pour le Développement de l'Horticulture, ISRA, 14000 Dakar, Senegal
[4]Université Cheikh Anta Diop (UCAD), Fac. Sci. Tech., BP 5005 Dakar, Senegal

This supplementary material supports the main text as follows:

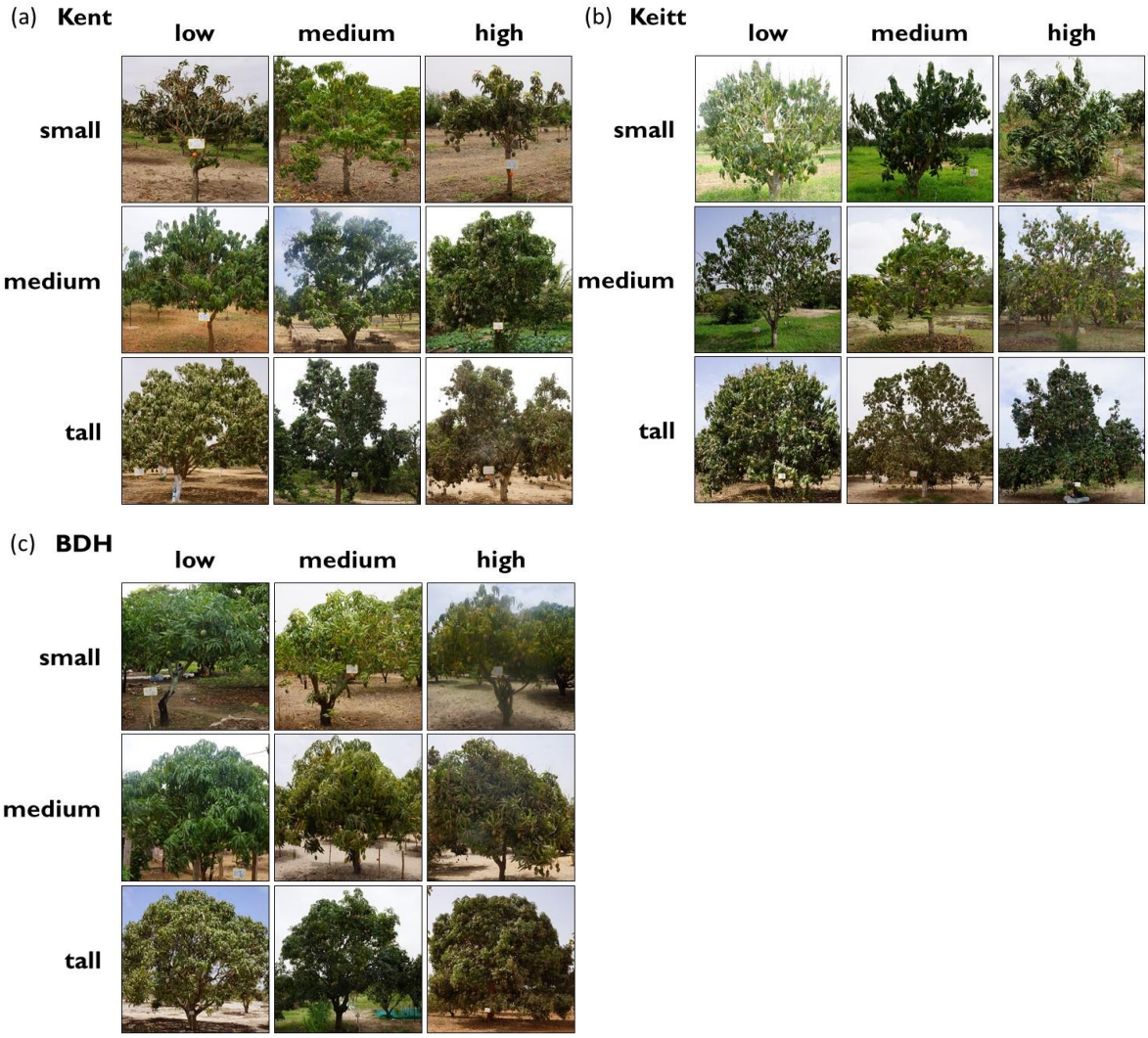


Figure S1. Image abacus used by expert in the field to estimate load index for (a) 'Kent', (b) 'Keitt', (c) and 'BDH' cultivar. Load index categories (low, medium and high) are displayed in column and different tree heights (small, medium and tall) are represented in line.

Table S1. Mean fruit weight and standard deviation (SD) for the three variety in Niayes region.

Variety	Number of fruit measured	Fruit weight (g)	
		Mean	SD
Kent	30	406.9	91.9
Keitt	30	586.6	151.8
BDH	30	442.3	131.8

Table S2. Description of the 150 calibration trees: cultivar; number of fruit detected by the KNN-based machine vision and ground estimated production; load index; and tree structure parameters (tree height, crown area and volume).

#	Cultivar	Number of fruit	Production (kg)	Load index	Tree height (m)	Crown area (m^2)	Crown volume (m^3)
1	Kent	49	60.1	high	3.6	9.5	23.3
1	Kent	57	69.9	high	4.9	15.6	52.8
1	Kent	43	52.8	high	4.9	13.1	48.8
1	Kent	118	144.8	high	4.6	12.7	43.1
1	Kent	26	31.9	medium	4.0	13.1	35.5
1	Kent	37	45.4	high	4.2	10.1	32.9
1	Kent	48	58.9	high	4.1	12.7	39.6
1	Kent	24	29.4	high	3.3	6.4	14.7
1	Kent	65	79.8	high	4.2	12.9	37.0
1	Kent	50	61.4	high	4.8	13.9	43.0
2	Kent	4	4.9	low	3.4	9.3	18.7
2	Kent	16	19.6	medium	4.1	12.0	34.6
2	Kent	61	74.8	medium	4.9	33.2	114.3
2	Kent	51	62.6	medium	5.8	49.7	182.0
2	Kent	8	9.8	medium	3.1	5.7	12.3
2	Kent	75	92.0	high	3.9	8.7	24.1
2	Kent	11	13.5	medium	4.6	6.7	18.2
2	Kent	19	23.3	low	5.2	17.7	62.8
2	Kent	7	8.6	low	4.9	27.4	90.7
2	Kent	57	69.9	high	3.2	6.2	16.6
3	Kent	7	8.6	low	3.1	3.5	7.1
3	Kent	24	29.4	medium	3.2	3.5	4.9
3	Kent	28	34.4	medium	3.1	6.9	16.0
3	Kent	8	9.8	medium	2.6	2.3	2.7
3	Kent	12	14.7	medium	2.7	7.3	6.6
3	Kent	43	52.8	high	3.1	8.5	19.0
3	Kent	3	3.7	medium	1.7	1.2	0.2
3	Kent	56	68.7	medium	3.9	12.6	29.8
3	Kent	25	30.7	medium	3.3	5.7	12.6
3	Kent	31	38.0	medium	3.4	13.1	26.4

4	Kent	7	8.6	medium	1.7	2.6	1.9
4	Kent	6	7.4	medium	1.6	1.0	0.3
4	Kent	3	3.7	low	1.9	3.9	2.2
4	Kent	25	30.7	high	2.5	6.9	11.0
4	Kent	22	27.0	medium	2.1	4.6	5.8
4	Kent	18	22.1	medium	2.4	5.1	8.0
4	Kent	30	36.8	high	2.8	8.2	12.6
4	Kent	22	27.0	high	2.6	5.4	7.6
4	Kent	26	31.9	medium	2.8	7.0	12.9
4	Kent	25	30.7	high	2.4	3.4	4.6
5	Keitt	167	281.1	high	5.7	29.7	111.8
5	Keitt	42	70.7	medium	6.4	29.5	141.9
5	Keitt	124	208.8	high	5.2	19.3	64.7
5	Kent	114	139.9	high	6.4	25.6	115.7
5	Keitt	93	156.6	high	5.8	22.7	91.0
5	Keitt	48	80.8	high	4.5	6.0	18.3
5	Kent	37	45.4	low	4.6	11.7	42.1
5	Keitt	76	127.9	high	5.4	19.1	66.8
5	Kent	49	60.1	medium	4.5	13.0	43.7
5	Keitt	16	26.9	medium	4.4	11.6	38.7
6	Keitt	57	96.0	high	5.1	11.9	29.6
6	Keitt	129	217.2	high	7.6	37.7	157.3
6	Keitt	76	127.9	medium	5.3	14.1	37.3
6	Kent	20	24.5	low	4.7	5.8	17.8
6	Keitt	37	62.3	medium	5.4	21.9	51.5
6	Keitt	64	107.7	high	5.1	13.7	51.8
6	Kent	81	99.4	medium	6.4	22.5	90.3
6	Keitt	75	126.3	medium	5.7	18.5	61.8
6	Kent	102	125.2	medium	5.5	17.7	64.5
6	Keitt	43	72.4	high	3.3	8.6	17.3
7	BDH	50	110.9	medium	8.0	34.4	206.0
7	Kent	118	144.8	high	6.5	11.4	49.6
7	BDH	23	51.0	low	7.6	44.0	269.8
7	Kent	24	29.45	medium	5.1	6.1	19.6
7	Kent	2	2.5	low	3.5	2.8	5.9
7	Keitt	1	1.7	low	2.7	3.9	4.6
7	BDH	90	199.6	medium	7.8	47.9	268.1
7	BDH	2	4.4	low	3.8	7.4	9.5
7	BDH	48	106.4	low	6.9	37.7	190.2
7	BDH	10	22.2	low	4.2	8.2	24.0
8	BDH	29	64.3	low	3.6	15.1	44.2
8	Kent	42	51.5	medium	2.7	8.4	13.3
8	BDH	18	39.9	medium	5.4	13.7	43.8

8	Kent	33	40.5	high	2.7	5.2	10.8
8	Kent	18	22.1	medium	3.1	7.0	17.4
8	BDH	26	57.7	low	3.2	20.2	41.4
8	BDH	16	35.5	low	2.8	19.0	39.2
8	BDH	39	86.5	medium	3.4	19.0	45.6
8	BDH	34	75.4	low	3.5	25.1	61.9
8	Kent	22	27.0	medium	3.0	4.3	10.3
9	Kent	19	23.3	low	4.2	11.2	31.5
9	Kent	7	8.6	low	3.9	6.2	19.3
9	Kent	75	92.0	medium	6.5	16.6	75.5
9	BDH	43	95.3	low	4.4	25.9	81.2
9	Kent	5	6.1	low	3.4	5.5	11.2
9	BDH	63	139.7	low	6.7	53.4	275.8
9	Kent	75	92.0	medium	6.5	20.6	81.9
9	BDH	126	279.4	medium	9.0	90.2	573.8
9	BDH	121	268.3	medium	9.3	93.6	655.2
9	Keitt	187	314.8	medium	8.1	48.3	266.7
10	Keitt	300	505.1	medium	11.2	86.1	575.4
10	Keitt	245	412.5	high	9.2	58.1	261.2
10	Keitt	190	319.9	high	6.8	28.2	122.3
10	Kent	284	348.5	high	9.9	53.4	328.0
10	Kent	494	606.1	high	9.8	56.1	387.5
10	Keitt	211	355.2	high	9.2	44.9	258.2
10	Kent	112	137.4	high	7.7	33.7	149.1
10	BDH	68	150.8	medium	5.5	19.1	82.9
10	Keitt	157	264.3	medium	8.0	48.2	249.9
10	Keitt	51	85.9	medium	6.4	23.8	100.8
11	Keitt	110	185.2	high	3.3	7.8	18.9
11	Kent	73	89.6	high	3.6	10.2	30.8
11	Keitt	204	343.4	high	4.8	34.3	106.6
11	Keitt	144	242.4	high	4.6	24.5	79.6
11	Kent	70	85.9	medium	4.5	21.5	61.0
11	Kent	93	114.1	high	4.4	19.3	51.9
11	Kent	151	185.3	high	4.0	24.1	45.3
11	Kent	68	83.4	high	3.3	9.2	23.6
11	Kent	164	201.2	medium	4.5	27.6	87.3
11	Kent	76	93.3	high	3.6	10.2	28.6
12	Kent	81	99.4	high	3.4	10.2	28.6
12	Kent	104	127.6	high	5.4	19.4	71.2
12	Kent	15	18.4	medium	2.8	3.1	7.0
12	Kent	154	189.0	high	5.4	22.6	88.2
12	BDH	24	53.2	low	4.1	21.3	67.9
12	Kent	95	116.6	high	4.7	13.5	45.0

12	Kent	168	206.1	high	5.8	19.9	89.1
12	BDH	74	164.1	medium	5.2	29.8	100.1
12	Kent	40	49.1	medium	4.2	11.6	35.6
12	Keitt	74	124.6	high	6.4	28.8	131.4
13	Kent	45	55.2	high	5.2	5.1	55.6
13	Kent	18	22.1	medium	1.7	2.2	2.0
13	Kent	51	62.6	high	0.8	5.8	0.7
13	Kent	75	92.0	medium	4.6	11.7	36.5
13	Kent	63	77.3	medium	3.8	9.2	26.7
13	Kent	101	123.9	high	5.2	18.4	71.9
13	Kent	98	120.2	high	4.7	21.9	71.1
13	Kent	37	45.4	high	0.1	2.5	0.0
13	Kent	152	186.5	high	4.6	20.0	66.1
13	Kent	90	110.4	high	3.6	11.3	30.0
14	Keitt	94	158.2	high	4.7	10.4	33.7
14	Kent	137	168.1	high	6.2	22.6	95.6
14	Keitt	12	20.2	medium	2.1	2.2	3.2
14	BDH	24	53.2	low	5.3	24.6	97.8
14	Kent	17	20.9	medium	3.6	5.6	15.6
14	Kent	63	77.3	medium	4.9	8.3	31.5
14	BDH	143	317.1	high	6.9	63.6	344.7
14	Kent	172	211.0	high	6.4	29.9	133.7
14	Keitt	95	159.9	high	7.3	21.5	103.8
14	Keitt	9	15.2	medium	5.0	4.6	15.1
15	Keitt	126	212.1	medium	7.8	29.7	155.2
15	Keitt	99	166.7	high	5.2	13.3	43.4
15	BDH	28	62.1	low	6.8	54.8	239.1
15	Keitt	192	323.2	medium	8.2	43.8	232.5
15	Keitt	122	205.4	high	6.6	20.0	75.0
15	Keitt	85	143.1	high	6.6	40.2	3.6
15	Keitt	138	232.3	high	7.1	26.0	104.5
15	Keitt	157	264.3	high	8.4	33.2	151.0
15	Keitt	97	163.3	high	6.1	6.9	23.1
15	Keitt	84	141.4	medium	6.6	24.7	110.6

Annexe V: Poster conférence Global CSA

Integrating innovative digital tools to estimate yields of perennial crops at different spatial scales in the tropics



Emile Faye[1,2,3], Julien Sarron[1,2,3], Eric Malézieux[1,3,*]
[1]CIRAD, UPR HortSys, F-34398 Montpellier, France
[2]Centre pour le Développement de l'Horticulture, ISRA, Dakar, Sénégal
[3]Univ Montpellier, F-34090 Montpellier, France
*Corresponding author: eric.malezieux@cirad.fr

cirad

#DigitAg

Context and objectives

For most crops, estimation of yield is a strategic and key information for both growers and stakeholders of the value chains involved in market supply and exports. Climate has already a strong effect on yields and their variability but climate change may dramatically impact yields and increase yield variability and uncertainty some areas. Information on yields and production is barely available nor existing for fruit tree production in the tropics, such as mango or litchi, making it difficult to improve the efficiency and resilience of the production system especially for smallholders. As a result, we developed specific digital tools to estimate mango system productivity at different scales.



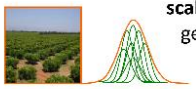
Methodology and results

This innovative toolbox brings together:

1. Fruit load and production estimation **at the tree scale**, based on image recognition process **using deep neural networks** on RGB infield images.



2. Orchard structure and production estimation **at the orchard scale**, using **UAV photogrammetry** and geographic object-based image analysis.



3. Production estimation **at the regional scale** using remote sensing technology and expert knowledge for orchard cropping systems classification.

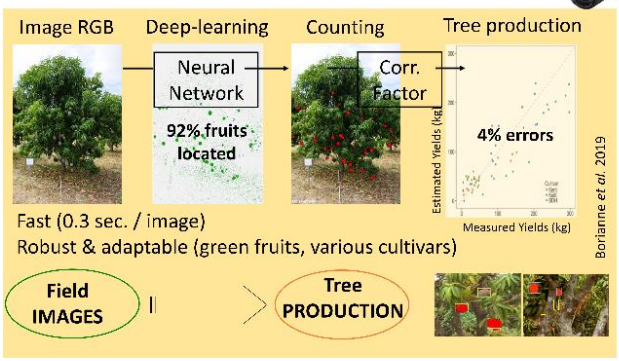


Take home message

An integrative digital toolbox was designed for different tree crop species, value chains, and local conditions to estimate tree crop yields using image analysis, neural networks, and remote sensing to implement empirical models. The methodology was set up, adapted, and performed for mango production in Senegal. Yields' spatial heterogeneity could be related to agricultural practices and environment (orchard composition and climate) at different scales. **The toolbox provides new perspectives for finer forecasting and analysis of crop yields for perennials in tropical value chains that include numerous smallholders.**

1. Tree scale





2. Orchard scale



3. Regional scale



Mango orchard yield mapping in the Niayes region in Senegal

Conception : Cirad - septembre 2020

This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004 and ANR-10-EQPX-20.

Résumé

L'Afrique de l'Ouest fait face à des changements démographiques et climatiques qui mettent sous tension l'agriculture. Les arbres fruitiers, malgré les nombreux services écosystémiques qu'ils fournissent (alimentation, biodiversité, fertilité, etc.), pâtissent de faibles rendements. Dans ce contexte d'insécurité alimentaire, il s'avère indispensable d'améliorer les moyens de suivi de la production agricole pour faire face aux enjeux du développement. Les objectifs de cette thèse étaient de mettre en place des outils d'estimation du rendement du manguier (*Mangifera indica L.*) à différentes échelles spatiales afin d'identifier les facteurs explicatifs de la variabilité intra- et inter-verger observée en Afrique de l'Ouest. Ces outils ont été construits et déployés dans la région des Niayes au Sénégal pour analyser les rendements d'un réseau de 30 vergers répartis dans trois types de systèmes de culture (extensif, diversifié et intensif). À l'échelle de l'arbre, un outil d'analyse d'images numériques par réseau de neurones assisté par modèle a été mis au point pour estimer la production de manguiers à l'approche de la récolte. Le concept de '*yield gap*' a été adapté pour calculer le '*production gap*', c'est-à-dire l'écart entre la production atteignable de l'arbre (i.e., définie par sa structure et sa variété) et sa production réelle. Cette méthode a mis en évidence les effets de la variété et du système de culture sur les écarts de production. À l'échelle du verger, l'imagerie drone a été utilisée pour construire une carte d'occupation des sols (position des arbres) et déterminer la structure (hauteur, volume) des arbres de chaque verger. Ces informations permettent de modéliser et spatialiser la production des manguiers au sein d'un verger, permettant ainsi l'étude des variabilités du rendement intra et inter-vergers. Les distributions spatiales des caractéristiques et de la production des arbres apparaissent plus hétérogènes dans les vergers extensifs et diversifiés. À l'échelle régionale, l'utilisation du diagnostic agronomique a permis de déterminer les principaux facteurs d'élaboration du rendement du manguier. Si la densité de manguiers est une variable déterminante du rendement dans les vergers intensifs, les vergers diversifiés sont caractérisés par une production de l'arbre plus importante. Le climat, les pratiques culturales et la diversité cultivée sont également des facteurs explicatifs de la variabilité de la production. Cette thèse ouvre de nouvelles voies méthodologiques pour pallier au manque d'information pour l'analyse de la production des arbres fruitiers tropicaux. Les méthodes et outils numériques mis au point permettent l'identification des facteurs impactant l'élaboration du rendement du manguier (climat, pratique et diversité cultivée), une étape décisive vers l'amélioration durable des pratiques et de la production de mangues en Afrique de l'Ouest.

Mots clés : arbre fruitier, agroforesterie, système de culture, télédétection, agriculture numérique, Sahel.

Abstract

West Africa faced many challenges including demographic growth and climate change that put agriculture under pressure. Fruit trees, despite the numerous ecosystem services they provide (food, biodiversity, soil fertility, microclimate, etc.), suffer from low yields. In this context of food insecurity, it is essential to improve tools for agricultural yield monitoring to face development challenges. The objectives of this thesis were to develop tools for yield estimation of mango (*Mangifera indica L.*) at different scales and to identify drivers of yield variabilities in West Africa. These tools have been deployed to analyse yields of 30 orchards depicting three cropping systems (extensive, diversified, and intensive) in the Niayes region in Senegal. At the tree scale, an image analysis tool using deep learning algorithm combined with models allowed to accurately estimate the production of mango trees before harvest. The '*yield gap*' concept was adapted to compute the tree '*production gap*' as the difference between the attainable production of the tree (i.e., determined by the cultivar and the structure) and its actual production. This statistical method evidenced that the mango production was influenced by the cultivar and the cropping system. At the orchard scale, drone imagery was used to build land cover map and estimate the structure parameters (tree height, crown area, crown volume) of all trees in the orchard. These data allowed to model and spatialize the individual production of each mango tree. Yield estimation allowed the study of the variability of yields within and between orchards. Spatial distribution of tree structure and production appear more spread out in the extensive and diversified orchards. At the regional scale, regional agronomic diagnosis allowed to identify important drivers impacting orchard mango yield in the study area. While mango planting density is a key factor for yield in intensive orchards, the tree production is favoured in diversified orchards. The climate, the management practices, and the species diversity also induce yield variabilities. This thesis opens new methodological paths to compensate the lack of data for fruit tree yields analysis in West Africa. Finally, the study of factors impacting mango yields (climate, practices, and species diversity) will allow to sustainably improve practices and production of mango in West Africa.

Keywords: tree crop, agroforestry, cropping system, remote sensing, digital agriculture, Sahel