

UNIVERSITÉ D'ANGERS
PHYSIQUE APPLIQUÉE et INGÉNIERIE PHYSIQUE

RAPPORT DE STAGE

effectué à l'IRHS-INRAE Angers

06 Mars - 31 Août 2023

**Instrumentation et imagerie pour la
caractérisation des motifs de coloration et la
détection du brunissement chez les pommes à
chair rouge**

YAO K. LETEKOMA

Master II : Photonique, Signal, Image

Encadrants

Anne-Laure FANCIULLINO
Chercheur à L'INRAE

Pierre BOUILLON
Doctorant à L'INRAE d'Angers

Jean-Marc CELTON
Maître de conférences à L'université d'Angers

Etienne BELIN
Maître de conférences à L'université d'Angers

INRAE Pays de la Loire
42 Rue Georges Morel,
49070 Beaucouzé



Je, soussigné (e) LETEKOMA YAO K.,
déclare être pleinement conscient(e) que le plagiat de documents ou d'une partie d'un
document publiés sur toutes formes de support, y compris l'internet, constitue une violation
des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer
toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

Signature :



Remerciement

Je remercie Dieu pour m'avoir donné la force et le courage d'accomplir ce travail. Je tiens à exprimer toute ma reconnaissance à mon Maître de stage à l'INRAE Etienne BELIN, qui a bien voulu m'accorder sa confiance en me permettant de travailler sur ce thème. Sa rigueur de travail et son soutien m'ont été de précieux atouts dans la réalisation de ce travail.

Je tiens à exprimer également, toute ma reconnaissance à mes co-encadrants Pierre BOUILLON, Jean-Marc CELTON et Anne-Laure FANCIULLINO pour leurs disponibilités et leurs apports tout au long de ce travail. Leurs précieux conseils m'ont permis d'améliorer la qualité de ce travail.

Mes profondes gratitude vont à l'IRHS, l'institut qui m'a accueilli et servi de cadre adéquat d'apprentissage durant ces six mois de stage. Je remercie la directrice de l'IRHS d'ANGERS, le personnel administratif ainsi que tous les membres de l'équipe VALEMA.

Enfin, mes remerciements vont également à tous les enseignants de l'université d'Angers qui ont contribué à ma formation le long de cette année universitaire 2022-2023 et tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Résumé

Le stage comporte deux objectifs pour améliorer le phénotypage des pommes à chair rouge :

- la détermination de descripteurs caractérisant les différents motifs de coloration à l'échelle du fruit.
- et la détection du brunissement interne par une méthode non destructive.

Le développement de variétés de fruits à chair rouge est un nouvel enjeu de la sélection, particulièrement chez le pommier. La couleur rouge dans la chair des fruits présente des variations d'intensité et de distribution de la couleur (variations phénotypiques dues à des facteurs génétiques et environnementaux), ce qui rend l'évaluation de ce caractère complexe et subjectif. Des méthodes d'imagerie RGB permettent de segmenter et de classifier les différents niveaux de coloration rouge, de fournir la surface des motifs par coloration, mais ne rendent pas compte de leur répartition spatiale ni de leur morphologie.

Ainsi, l'établissement de descripteurs morphométriques qui caractérisent les différents motifs de chaque niveau de coloration est recherché pour rendre compte de leur variabilité dans les pommes à chair rouge. Nous avons pu répertorier puis implémenter en python près d'une vingtaine de descripteurs de contours et de formes des motifs sur un ensemble de 1536 (données de 2021), et 1743 (données de 2022) images de tranches de pommes. Cependant nous n'avons pu traiter que les résultats d'un certain nombre de descripteurs tels que l'aire, le nombre de trous, le nombre de régions ou motifs connexe et la distance au centre de pomme en miniprojet.

En stage, nous nous sommes penchés, sur l'analyse globale de l'ensemble des données extraites après avoir modifié l'algorithme de segmentation dans la chaîne d'acquisition par l'algorithme Kmeans ($k=4$) à cause des bruits de segmentation.

L'analyse s'est faite à l'aide des ACP (Analyses en Composantes Principales), un puissant outil permettant de gérer les multiples dimensions des données, pour analyser les résultats. Nous avons réussi à séparer les motifs de coloration en différents groupes en fonction des données obtenues. En particulier, nous nous sommes concentrés sur les motifs de la classe 1 (chair peu colorée en rouge), que nous avons redécrits à l'aide de repères (landmarks) ; cela nous a permis de caractériser ces motifs en utilisant des coefficients de Fourier elliptique plutôt que de travailler avec une grande quantité de pixels pour décrire les contours. L'application de l'ACP sur ces coefficients a montré une corrélation significative entre les contours et les données de coloration, en regroupant les motifs selon leur origine (pommes à chair rouge, rouge clair ou blanc). L'ensemble des processus de traitement, depuis la numérisation des pommes jusqu'à l'application des ACP, est regroupé dans un pipeline illustré dans la figure 1.

Bien que ces résultats soient intéressants, nous rencontrons des difficultés pour extraire les informations nécessaires pour identifier les gènes responsables des motifs de coloration. Néanmoins, grâce à ces regroupements et en combinant les données de morphométrie des pommes récoltées en 2021 et 2022 sur un seul site près d'Angers, nous pouvons conclure que les motifs de coloration sont bien associés à des gènes.

La deuxième problématique du thème de ce stage porte sur la détection du brunissement au sein des pommes. Certains fruits à chair rouge souffrent de brunissement en conservation. Cette altération de la chair résulterait de phénomènes oxydatifs générés par une sénescence accrue. Il n'existe aujourd'hui aucune méthode de phénotypage non-destructive capable de prédire la présence de ce type de brunissement.

Tout d'abord, nous avons fabriqué un pendule en utilisant une petite masse pour frapper les pommes. Ensuite, nous avons enregistré les sons émis lors des impacts. Cependant, nous avons remarqué que les fréquences fondamentales des sons enregistrés étaient identiques pour toutes les pommes. Cela indiquait que notre méthode de frappe n'était pas appropriée. Pour remédier à cela, nous avons découvert l'Acoustic Firmness Sensor (AFS), un dispositif qui mesure la fermeté, la fréquence fondamentale et le poids des fruits, dont nous avons mis en place un protocole d'utilisation. En l'absence de pommes présentant des signes de brunissement, nous avons utilisé l'AFS pour mesurer ces paramètres sur des pommes de deux variétés : des pommes saines et des pommes en état de pourriture ou de sénescence (caractéristiques similaires au brunissement interne).

Suite à l'analyse des données collectées, nous avons réussi à distinguer les pommes en état de pourriture des pommes saines. Les pommes en mauvais état ont démontré une réduction de leur masse fraîche ainsi qu'une altération des structures cellulaires contenant de l'eau, telles que les parois et les membranes. Cette altération a eu un impact sur leur fréquence de résonance, entraînant une diminution significative de leur fermeté. Ces constatations soulignent que l'AFS pourrait se révéler utile pour la détection du brunissement dans des études à venir.

Mots-clés : descripteurs de formes et de contours, transformée de fourier elliptique, ACP, landmarks, espace morphologique, fermeté, fréquence fondamentale.

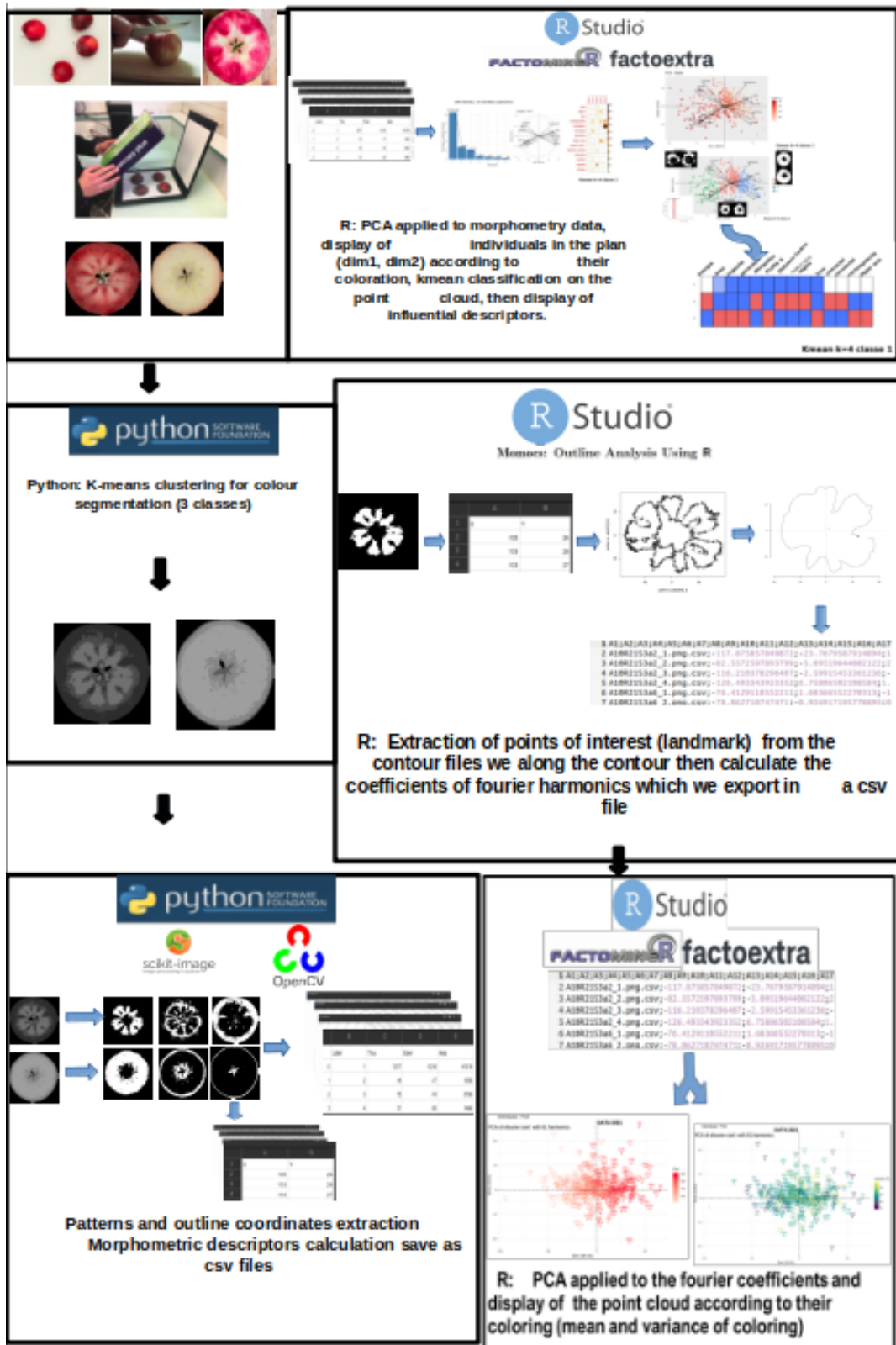


FIGURE 1 – Résumé de l'ensemble de la chaîne de traitements développée en Python et en R.

Summary

The development of red-fleshed fruit varieties is a new challenge in breeding, particularly in apple. The red colour in the flesh of the fruit shows variations in intensity and distribution of colour, which makes the assessment of this characteristic complex and subjective. colour distribution, which makes the assessment of this trait complex and subjective.

RGB imaging methods can be used to segment and classify the different levels of red colouring, to provide the surface area of the patterns per coloration but do not account for their spatial distribution or morphology.

Thus, the establishment of morphometric descriptors that characterize the different patterns at each color level is sought to account for their variability in red-fleshed apples. We were able to identify and implement nearly twenty contour and shape descriptors of patterns on a set of 1536 (2021 data) and 1743 (2022 data) images of apple slices using python. However, we could only process the results of a certain number of descriptors, such as area, number of holes, number of connected regions or patterns, and distance to the center of the apple in the mini-project.

During the internship, our focus was on the overall analysis of the extracted data after modifying the segmentation algorithm in the acquisition chain with the Kmeans algorithm (k=4) due to segmentation noises.

The analysis was carried out using Principal Component Analysis (PCA), a powerful tool for managing multiple dimensions of data and analyzing the results. We successfully separated the color patterns into different groups based on the obtained data. In particular, we focused on patterns from class 1 (lightly red-colored flesh) and re-described them using landmarks.

This allowed us to characterize these patterns using elliptical Fourier coefficients instead of working with a large number of pixels to describe the contours. Applying PCA to these coefficients showed a significant correlation between the contours and the coloration data, grouping the patterns according to their origin (apples with red, light red, or white flesh).

Although these results are interesting, we face challenges in extracting the necessary information to identify the genes responsible for the color patterns. Nevertheless, thanks to these groupings and by combining the morphometric data from apples harvested in 2021 and 2022, we can conclude that the color patterns are indeed associated with genes.

A second issue addressed in the scope of this internship concerns the detection of browning within apples. Initially, we constructed a pendulum with a small mass to strike the apples, recording the emitted sounds. However, the fundamental frequencies of the recorded sounds were the same for all apples, indicating an inaccurate method of striking. Subsequently, we discovered the Acoustic Firmness Sensor (AFS). In the absence of browned apples, we utilized the AFS, a device that measures the firmness, fundamental frequency, and weight of fruits, for which we established a usage protocol. By analyzing various apples of different types, we successfully distinguished between rotten and healthy apples. Rotten apples displayed an increase in water content, which affected their resonance frequency, resulting in a considerable decrease in firmness and weight. These findings suggest that the AFS could be valuable for detecting browning in future research.

Keywords : shape and contour descriptors, elliptical fourier transform, PCA, landmarks, morphological space, firmness, fundamental frequency.

Table des matières

Engagement de non plagiat	I
Remerciement	II
Résumé	III
Summary	V
Table des figures	VII
1 Introduction	1
2 Présentation de l'unité de recherche et de l'équipe d'accueil	2
3 Caractérisation des motifs de coloration	3
3.1 Segmentation via k-means	3
3.2 Morphométrie traditionnelle	5
3.2.1 Descripteurs de forme	5
3.2.2 Descripteurs de contour	10
3.2.3 Autres descripteurs	11
3.3 Choix des Descripteurs pour l'analyse	12
3.4 Analyse en Composantes Principales (ACP)	14
3.5 Représentation des motifs par landmarks	15
3.5.1 Transformées de Fourier et contours fermés	17
3.6 Présentation et analyse des résultats	20
3.6.1 Application de l'ACP au données issues des descripteurs traditionnels	21
3.6.2 Application de l'ACP sur les descripteurs de fourier harmonique	28
3.6.3 Analyse des résultats et discussion des observations	30
4 Détection du brunissement au sein d'une pomme	33
4.1 Prise en main du capteur acoustique afs d'aweta	33
4.2 Analyse et présentation des résultats obtenus grâce à l'utilisation du matériel	37
5 Conclusion	40
5.1 Annexe	41
Bibliographie	42

Table des figures

1	Résumé de l'ensemble de la chaîne de traitements développée en Python et en R.	IV
3.1	Exemple de classification avec KNN, avec illustration du bruit de segmentation dans les classes 1 et 6.	3
3.2	Illustration de l'algorithme Kmeans : de l'initialisation des centroides au regroupement des classes [12].	3
3.3	Illustration de l'application du Kmeans pour $k=4$, $k=5$, et du KNN (à l'extrême droite).	5
3.4	Illustration du point de flexion (intersection des courbes noir et bleue).	5
3.5	Aire d'une région donnée : ensemble des pixels à l'intérieur du contour[4].	6
3.6	Aire de l'enveloppe convexe : ensemble des pixels à l'intérieur du contour bleu [4].	6
3.7	Aire du rectangle englobant.	6
3.8	Illustration du périmètre caractérisant l'objet (région) et l'enveloppe convexe [4].	6
3.9	Ellipse englobant (gauche), à droite l'objet.	7
3.10	Compacité.	7
3.11	Illustration de l'aspect ratio.	8
3.12	Illustration de l'élongation.	8
3.13	Illustration de l'enveloppe convexe (à droite).	9
3.14	Illustration du profil [5]	10
3.15	Courbure $k(p)$	11
3.16	Illustration du nombre d'euler.	11
3.17	Distance au centre.	12
3.18	Les 14 descripteurs retenus.	12
3.19	Corrélation entre les descripteurs établis.	13
3.20	Extraction des landmarks au sein des nuages de points (en noir) représentant le motif à l'aide de droites parallèles (en bleue) et à droite la reconstruction du contour à l'aide des landmarks.	16
3.21	Extraction de landmarks à l'aide de droites étoilées (en vert le rectangle englobant, en rouge les landmarks détectés) et reconstruction du contour à droite.	16
3.22	Vu sur 12 motifs (à gauche) et la représentation par des landmarks (à droite).	16
3.23	Vue d'ensemble de 1536 motifs représentés par des landmarks extraits des données de 2021.	17
3.24	Vue d'ensemble de 1743 motifs représentés par des landmarks extraits des données de 2022.	18
3.25	Trois harmoniques sont ici représentées à quatre emplacements sur le contour d'origine [26].	19
3.26	Influence du nombre d'harmoniques sur la reconstruction du contour à partir des coefficients de ces harmoniques.	20
3.27	Représentation des individus par les données de colorations.	21
3.28	Motifs de la classe 1.	21
3.29	Motifs de classe 1 : Pourcentage de variances expliquées, cercle de corrélation, contribution des descripteurs à chaque axe.	22
3.30	Distribution des individus par coloration suivant les axes principaux 1 et 2.	22
3.31	Motifs de classe 1 : regroupement des individus en 3 groupes dans l'espace construite.	23
3.32	Variable dominant dans chaque groupe.	23
3.33	Motifs de la classe 2.	24
3.34	Motifs de classe 2 : pourcentage de variances expliquées(à gauche), cercle de corrélation (centre), contribution des descripteurs à chaque axe (à droite).	24
3.35	Distribution des individus par coloration suivant les axes principaux 1 et 2.	25
3.36	Motifs de classe 2 : regroupement des individus en 3 régions dans l'espace construit.	25
3.37	Variable dominant dans chaque groupe.	25
3.38	Motifs de la classe 3.	26
3.39	Motifs de classe 3 : pourcentage de variance expliquée, cercle de corrélation, contribution des descripteurs à chaque axe.	27
3.40	Distribution des individus par coloration suivant les axes principaux 1 et 2.	27
3.41	Motif de classe 3 : regroupement des individus en 3 régions dans l'espace construit.	28

3.42	Cercle de Variabilite de l'ACP des coefficients extrait des données de 2021.	29
3.43	Cercle de Variabilite de l'ACP des coefficients extrait des données de 2022.	29
3.44	Répartition des individus suivant les 2 premières dimensions obtenu après l'application de l'ACP sur les coefficients de fourier extraits des données de 2021 coloriés par a_mean.	30
3.45	Répartition des individus suivant les 2 premières dimensions obtenu après l'application de l'ACP sur les Coefficients de fourier extraits des données de 2022 coloriés par a_mean.	31
3.46	Répartition des individus suivant les 2 premières dimensions obtenue après l'application de l'ACP sur les coefficients de fourier extraits des données de 2021 coloriés par L_var.	31
3.47	Répartition des individus suivant les 2 premières dimensions obtenue après l'application de l'ACP sur les coefficients de fourier extraits des données de 2022 coloriés par L_var.	32
4.1	Matériels de l'afs, d'aweta, Nootdorp, Pays-Bas.	33
4.2	Illustration des accessoires de l'AFS, et du positionnement du produit à mesurer.	34
4.3	Interface de l'AFS.	35
4.4	Configuration des paramètres de l'AFS.	36
4.5	Mesure prise par l'AFS.	36
4.6	Pomme rouge de même famille normale.	37
4.7	Pomme rouge de même famille pourris.	37
4.8	Deux pommes rouge de même famille (1 ère mesure).	37
4.9	Deux pommes rouge de même famille (2 ème mesure)	37
4.10	Quatre pommes d'une autre famille (pink lady).	39
4.11	1 ère mesure des pommes No1 et No3 des 4 pommes (pink lady).	39
4.12	2 ème mesure (2 semaines après) des quatre pommes "pink lady" (No1 et No2).	39
4.13	No3 et No 4 des "pink lady".	39
5.1	Protocole d'acquisition avec l'AFS, aweta	41

Chapitre 1

Introduction

L'analyse morphologique appliquée aux fruits est une discipline de recherche qui se concentre sur l'étude de la forme et de la structure des fruits dans le but de mieux comprendre leur développement, leur maturité, leur qualité et leur diversité. Cette branche de la botanique joue un rôle crucial dans de nombreux domaines, notamment l'agriculture, l'horticulture, la sélection des variétés, l'amélioration génétique des cultures et la sécurité alimentaire.

Les fruits sont des structures essentielles dans le cycle de vie des plantes à fleurs, car ils sont responsables de la production et de la dissémination des graines. Leur forme et leur taille varient considérablement selon les espèces et jouent un rôle crucial dans la manière dont les graines sont dispersées, ainsi que dans leur protection contre les prédateurs. Par conséquent, l'analyse morphologique des fruits est d'une importance capitale pour mieux comprendre la biologie des plantes et les mécanismes évolutifs qui façonnent leur diversité.

Dans notre problématique, notre analyse ne porte pas sur le fruit en lui-même, mais plutôt sur la morphologie des motifs de coloration au sein de la chair de la pomme rouge afin de trouver d'éventuelles informations génétiques. Ce thème fait suite au sujet de mini-projet réalisé de novembre 2022 à février 2023 au cours duquel nous avons développé des descripteurs morphométriques que l'on avait partiellement analysés à l'aide des outils classiques de statistique. Nous voulons cette fois aller au-delà et construire des espaces morphologiques à l'aide de l'analyse en composante principale (ACP), d'abord à partir des données enregistrées grâce à nos descripteurs, puis à partir des coefficients de Fourier elliptiques. Ces derniers décrivent assez efficacement les contours aussi complexes soit-il.

L'analyse morphométrique a fait ces preuves dans [2], où plus de 3300 feuilles provenant de 40 espèces différentes de Passiflore ont été analysées de manière morphométrique, en les collectant séquentiellement le long des tiges. Chaque feuille a été mesurée de deux manières différentes : en utilisant 15 repères ou landmarks homologues ajustés par Procruste pour la vasculature, les sinus et les lobes ; et en utilisant des descripteurs elliptiques de Fourier pour quantifier le contour de la feuille. Les capacités prédictives des repères, des EFD et de l'ensemble combiné des données ont été comparées pour déterminer leur aptitude relative à prédire l'espèce et la position du nœud dans la liane. Les analyses de corrélation des coordonnées des repères x et y ainsi que des coefficients harmoniques des EFD ont mis en évidence des liens étroits entre les caractéristiques morphologiques et des informations sur la relation entre la vasculature et le schéma de la lame des feuilles.

Dans [3] une étude sur huit populations spontanées de *Pistacia atlantica* en Algérie a révélé une diversité morphologique significative intra et inter-populations. Les caractères biométriques des feuilles et fruits ont été étudiés morphologiquement. L'analyse de la variance et des moyennes a montré des différences importantes correspondant à différents bioclimats.

Avec les diversités morphologiques que l'on observe au sein des motifs de coloration des pommes rouges que nous extrayons, nous espérons les regrouper par similarités morphologiques et y trouver des liens génétiques.

Par ailleurs, nous étudions également les signaux acoustiques récupérés à partir des sons émis par les pommes afin de détecter ceux qui sont brunis de manière non destructive. En effet, lorsque l'on conserve des pommes en frigo on s'est rendu compte qu'au bout d'un temps donné certaines commençaient par brunir, ce qui n'est pas très apprécié en commerce. Nous utilisons un outil nommé AFS-Aweta existant permettant de mesurer à partir des signaux émis en cognant sur un fruit afin de mesurer sa fermeté, la fréquence fondamentale et son poids.

Chapitre 2

Présentation de l'unité de recherche et de l'équipe d'accueil

Ce stage, financé par le GIS fruit a eu lieu dans l'Institut de Recherche en Horticulture et Semences (IRHS, regroupant l'INRAE, L'institut Agro Rennes-Angers et l'Université d'Angers, UA) sur le site de de Beaucouze, notamment dans l'équipe VALEMA (Valorization of Epigenetic Marks in plants), sous la supervision de Jean-Marc Celton (maître de conférence UA), d'Etienne Belin (maître de conférence UA), d'Anne-Laure Fanciullino (chercheur INRAE), et de Pierre Bouillon (doctorant). L'IRHS est sous la direction de Marie Agnes Jacques et constitué de plusieurs autres équipes.

L'INRA (Institut national de la recherche agronomique) créé depuis 1946 fusionne le 1er janvier 2020 avec l'IRSTEA pour former l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE).

Premier institut de recherche agronomique en Europe et deuxième dans le monde en nombre de publications en sciences agricoles et en sciences de la plante et de l'animal, l'INRAE mène des recherches finalisées pour une alimentation saine et de qualité, pour une agriculture durable, et pour un environnement préservé.

L'équipe VALEMA, initialement connue sous le nom d'EpiCenter de 2014 à 2019, a été fondée grâce au financement ConnecTalent de la Région Pays de la Loire, attribué à E. Bucher.

Pendant la période allant de 2014 à 2019, l'équipe s'est principalement concentrée sur la recherche du rôle de l'épigénétique dans le développement des plantes et dans les mécanismes de défense contre les pathogènes.

Depuis 2019, l'équipe est sous la direction de Jean-Marc Celton. Les axes de recherche de l'équipe sont centrés sur l'adaptation des plantes pérennes aux contraintes abiotiques et la mémoire épigénétique des stress, et sur l'élaboration des qualités du fruit (en particulier les métabolites spécialisés telles les anthocyanes) dans un contexte de contraintes abiotiques. Le modèle d'étude est le pommier (*Malus domestica*).

Chapitre 3

Caractérisation des motifs de coloration

Nous commençons par aborder la problématique liée à la caractérisation des motifs de coloration. Nous analyserons les données morphométriques en utilisant l'ACP (Analyse en Composantes Principales) afin d'identifier des groupes morphométriques susceptibles de contenir des informations génétiques. Ensuite, nous reconfigurerons les contours des motifs de la classe 1 en utilisant des coefficients de Fourier elliptiques, en vue d'appliquer les mêmes méthodes d'analyse que précédemment.

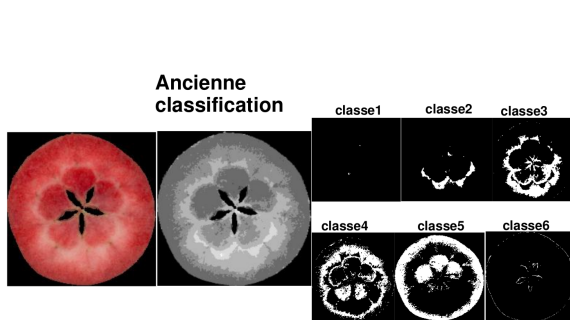


FIGURE 3.1 – Exemple de classification avec KNN, avec illustration du bruit de segmentation dans les classes 1 et 6.

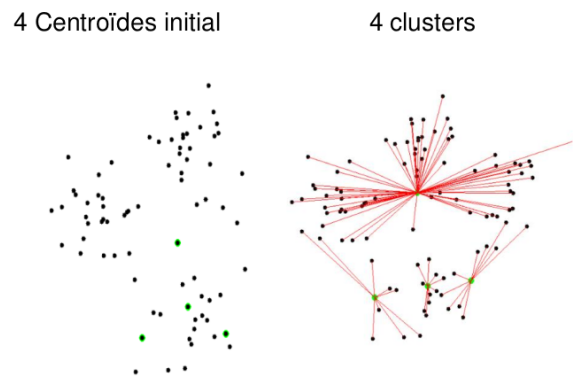


FIGURE 3.2 – Illustration de l'algorithme K-means : de l'initialisation des centroïdes au regroupement des classes [12].

3.1 Segmentation via k-means

L'objectif de ce chapitre est d'étudier les motifs de coloration de la chair de la pomme rouge obtenus à partir des images RGB extrait à l'aide d'un outil existant qui avait aussi permis de classifier ces images, à l'aide de l'algorithme des k plus proches voisins (KNN), en 6 classes de segmentation. Cependant, nous avons constaté que certains motifs ne reflétaient pas fidèlement les zones de coloration, en particulier les zones les plus rouges. Ces incohérences étaient dues au bruit de segmentation, comme illustré dans la figure 3.1, amplifié par le nombre de classes (6) que nous avons fixées.

Dans [11] les auteurs utilisent l'algorithme de segmentation k-mean pour segmenter la couleur de la chair de la fraise pour l'analyse phénotypique et génétique complète des traits morphologiques extraits des images internes et externes des fraises. C'est dans cette optique que nous avons décidé d'adopter une approche de segmentation différente en utilisant aussi K-means. Cette méthode permet de segmenter de manière dynamique les différents niveaux de coloration de rouge des images de la chair de pommes en 4 classes, y compris le fond. En dehors du fond la chair rouge est classifiée en 3 classes de rouge représentant la chair blanche, la chair rouge clair et la chair rouge foncé. Ce qui permet d'éviter d'étendre la chair rouge clair en plusieurs classes potentiellement responsables du bruit de segmentation.

Grâce à cette nouvelle approche d'appréhender les motifs de coloration en regroupant les données pixels en trois classes, cela a permis d'éviter le bruitage des motifs et de les rendre beaucoup plus compact correspondant à nos attentes visuelles représentant ainsi les 3 niveaux de coloration approximés, ce qui faciliterait par la suite les analyses morphologiques.

Le k-means [12] est un algorithme de segmentation largement utilisé dans le domaine de l'analyse d'images. Il est particulièrement adapté pour regrouper les pixels ou les points de données en fonction de leur similarité de caractéristiques. Dans le cadre de notre mémoire sur l'étude des motifs de coloration de la chair d'une pomme rouge, nous avons décidé d'intégrer l'algorithme K-means dans notre approche de segmentation.

L'algorithme K-means fonctionne (voir fig. 3.2) en assignant initialement des centroïdes aléatoires à un nombre prédéfini de classes, dans notre cas, 4 classes, y compris le fond. Ensuite, il itère en alternant entre deux étapes principales. Dans la première étape, chaque pixel ou point de donnée est attribué à la classe dont le centroïde est le plus proche en termes de distance euclidienne. Dans la deuxième étape, les centroïdes sont recalculés en utilisant les moyennes des pixels ou des points de données assignés à chaque classe.

Ce processus d'assignation et de recalcul des centroïdes se poursuit jusqu'à ce qu'une convergence soit atteinte, c'est-à-dire lorsque les centroïdes ne changent plus significativement ou lorsque le nombre d'itération prédéfini est atteint. Une fois l'algorithme K-means terminé, nous obtenons des classes ou des segments de pixels qui représentent les différents niveaux de coloration des images des tranches de pommes.

L'avantage du K-means réside dans sa simplicité et son efficacité computationnelle. Il permet une segmentation automatique des données en fonction des caractéristiques visuelles, ce qui facilite l'analyse et l'interprétation des motifs de coloration dans notre étude. En adoptant cette approche de segmentation dynamique en 4 classes, nous espérons obtenir des résultats plus précis et représentatifs des différents niveaux de coloration de la chair des pommes rouges.

En fait, en dehors du fond de l'image des tranches de pommes RGB, nous avons choisi de classer les niveaux de coloration rouge en 3 classes. Cependant, le nombre de classes $k=4$ a été décidé en utilisant la méthode du coude (elbow method), une technique couramment utilisée pour déterminer le nombre optimal de clusters dans un algorithme de segmentation comme K-means.

La méthode du coude consiste à exécuter l'algorithme K-means pour différents nombres de clusters, en mesurant la variation totale intra-classe à chaque itération. Cette variation représente la somme des distances entre les points de données et leurs centroïdes respectifs dans chaque classe. Ensuite, ces mesures de variation sont tracées sur un graphique en fonction du nombre de clusters.

L'idée est de trouver le point de flexion sur le graphique, qui ressemble à un coude, d'où le nom de la méthode comme nous le montre l'exemple à la fig. 3.4. Ce point correspond au nombre de clusters où l'ajout d'un cluster supplémentaire n'améliore pas significativement la variation totale intra-classe. En d'autres termes, il représente le point où la segmentation commence à se stabiliser et où l'ajout de clusters supplémentaires n'apporte pas d'avantages significatifs.

Dans notre étude, nous avons appliqué la méthode du coude sur 205 images RGB pour déterminer le nombre optimal de classes. Après avoir tracé les mesures de variation, nous avons identifié le point de flexion au niveau du paramètre $k=4$ pour 98/205 images et $k=5$: 107/205 images, où k représente le nombre optimal de classe. Ces deux résultats sont presque similaires et nous permettent de segmenter à $k=5$ ou 4. Mais pour ne conserver et n'étudier que le cas des motifs assez représentés, nous avons gardé le paramètre $k=4$; ce qui signifie que ce nombre de classes est approprié pour notre segmentation et nos objectifs sur les motifs à extraire. En effet, en observant les motifs fournis par une segmentation kmean $k=5$, ceux-ci étaient assez fragmentés et un peu plus bruités que les cas des $k=4$.

En utilisant cette approche basée sur la méthode du coude, nous avons pu choisir le nombre optimal de classes pour segmenter les différents niveaux de coloration rouge dans les images des tranches de pommes, ce qui devrait améliorer la représentation des motifs de coloration dans notre étude.

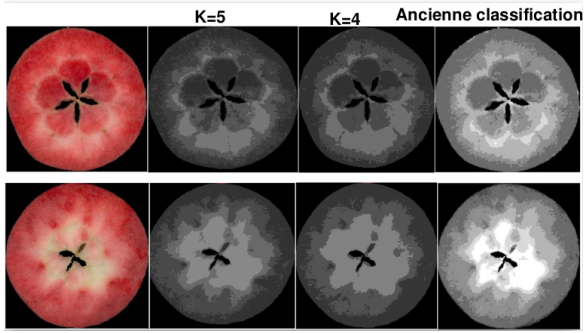


FIGURE 3.3 – Illustration de l'application du Kmeans pour $k=4$, $k=5$, et du KNN (à l'extrême droite).

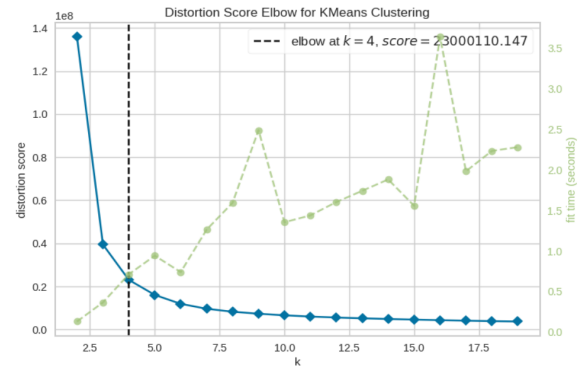


FIGURE 3.4 – Illustration du point de flexion (intersection des courbes noir et bleu).

3.2 Morphométrie traditionnelle

En traitement d'images, un descripteur (ou caractéristique) est une représentation numérique d'une image ou d'une partie d'une image qui peut être utilisée pour effectuer des tâches telles que la reconnaissance d'objets, la détection de visages, la segmentation d'images, etc...

Les descripteurs peuvent être définis comme des valeurs numériques extraites à partir de l'image, qui capturent certaines propriétés ou caractéristiques visuelles de l'image. Par exemple, des descripteurs peuvent être basés sur la couleur, la texture, la forme, la luminosité, etc...

Les descripteurs sont souvent utilisés dans les systèmes de vision par ordinateur pour comparer des images ou pour rechercher des images similaires dans une base de données. Ils peuvent également être utilisés pour réduire la dimensionnalité des données, ce qui peut faciliter le traitement de grandes quantités d'images.

Dans notre travail, nous cherchons à décrire les régions des classes de coloration rouge de la chair de la pomme, et pour cela nous utiliserons les descripteurs dits de formes et de contours.

Les descripteurs de forme et de contour d'objets sont utilisés pour quantifier et caractériser les propriétés géométriques des objets dans une image. Ces descripteurs sont généralement calculés à partir des contours d'objets extraits à partir de l'image. Nous présenterons par la suite les principaux descripteurs de formes et de contours existant dans la littérature.

Nous terminerons avec deux descripteurs particuliers, à savoir le nombre d'Euler qui est un descripteur topologique et la distance de chaque région par rapport au centre de la pomme.

3.2.1 Descripteurs de forme

Les descripteurs de forme sont des caractéristiques géométriques qui permettent de quantifier la forme d'un objet dans une image. Ces descripteurs sont utilisés dans de nombreuses applications d'imagerie telles que la reconnaissance de formes, la segmentation d'objets, etc ...

Les descripteurs ne permettent pas forcément de reconstituer entièrement la forme, mais elles doivent être suffisamment différents pour que les formes puissent être discriminées. Entre autres, on peut citer : la compacité, la solidité, l'enveloppe convexe, l'aspect ratio, l'élongation, rectangularité, la circularité, l'aire, le périmètre, la convexité (concavité), le centroïde, l'allongement, le rapport isopérimétrique, le grand et petit axe, le diamètre de féret.

L'aire

C'est une grandeur qui permet d'évaluer en pixel l'étendue de la forme d'une région donnée dans une image comme l'illustre la figure 3.5.

L'aire convexe [4] d'un objet ou d'une région est l'aire de l'enveloppe convexe (voir figure 3.6) qui entoure l'objet. L'enveloppe convexe est la plus petite région convexe qui inclut l'objet.

On appelle boîte englobante (fig. 3.7) ou rectangle englobant le plus petit rectangle qui circonscrit l'objet d'intérêt ; **l'aire de ce rectangle** [4] peut être utilisée comme descripteur pour caractériser une forme ou une région d'un objet dans une image. Elle s'obtient en multipliant sa longueur par sa largeur, ce qui correspond au nombre de pixels qui la constituent.



FIGURE 3.5 – Aire d'une région donnée : ensemble des pixels à l'intérieur du contour[4].

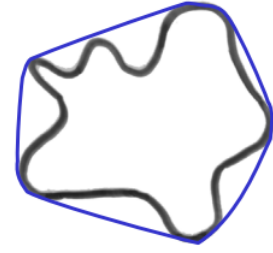


FIGURE 3.6 – Aire de l'enveloppe convexe : ensemble des pixels à l'intérieur du contour bleu [4].



FIGURE 3.7 – Aire du rectangle englobant.

Le périmètre

C'est le nombre de pixels du contour d'une région donnée suivant la connectivité 4 ou 8 [4], ou encore c'est la longueur qui se rapproche du contour de chaque région obtenue en joignant par une ligne les centres des pixels du contour permettant d'évaluer leur étendue au sein de l'image d'une tranche de pomme.

Nous pouvons aussi caractériser la forme d'une région donnée par le périmètre de son enveloppe convexe [4] : c'est le **périmètre convexe** (voir figure 3.8).

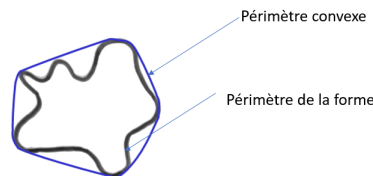


FIGURE 3.8 – Illustration du périmètre caractérisant l'objet (région) et l'enveloppe convexe [4].

Grand axe

Il est considéré comme la plus grande distance reliant deux points du contour au sein d'une région donnée [4]. Il est calculé de deux manières différentes : soit en cherchant le maximum des distances euclidiennes de toutes les lignes reliant deux points du contour, soit à partir du plus grand axe de la plus petite ellipse (voir figure 3.9) qui englobe la région étudiée.

Petit axe

Il est considéré comme la plus petite distance reliant deux points du contour au sein d'une région donnée [4]. Il se calcule avec les mêmes techniques que précédemment en considérant la plus petite distance (voir figure 3.9).



FIGURE 3.9 – Ellipse englobant (gauche), à droite l'objet.

Compacité/circularité

La compacité est définie comme le rapport de l'aire d'un objet par l'aire d'un cercle de même périmètre [4]. Cela permet de mesurer la dispersion du contour de l'objet par rapport à une forme parfaitement compacte qui est le cercle. Elle vaut l'unité pour un cercle (figure 3.10) et une valeur proche de 0 indique une forme très dispersée.

Elle s'exprime par la formule suivante :

$$C = \frac{4\pi \times Aire}{Perimetre^2}$$

Les objets ayant une compacité élevée sont souvent plus faciles à analyser et à traiter que les objets ayant une compacité faible, car ils présentent moins de variations et de bords irréguliers. Elle est utile pour évaluer la forme d'un objet et faciliter son traitement et son analyse.

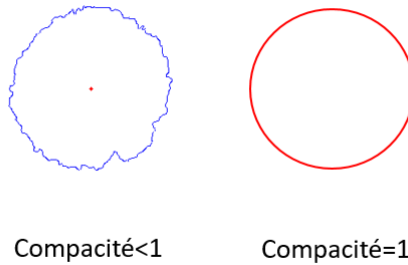


FIGURE 3.10 – Compacité.

Elle peut être évaluée en utilisant les aires du cercle englobant et de la forme [5] :

$$C = \frac{Aire_forme}{Aire_cercle}$$

Ou encore en utilisant la moyenne et l'écart type de la distance radiale entre le centroïde et les points du contour de la forme (variance de cercle) [5] :

$$C = \frac{\sigma_R}{\mu_R}$$

μ_R , σ_R : respectivement la moyenne et l'écart type de la distance radiale entre le centroïde et les points du contour.

D'autres documents [5] utilise la même formule (la première) pour caractériser la **circularité** définie comme un facteur de forme qui permet d'évaluer le rapprochement de la forme d'une région par rapport à un cercle.

$$Circ = \frac{4\pi \times Aire}{Perimetre^2}$$

Le périmètre et l'aire sont deux paramètres mesurant la taille des régions.

Aspect ratio, élongation et rectangularité

Également appelé facteur de forme, l'aspect ratio [4] est le quotient du grand axe par le petit axe de l'ellipse englobant la région considérée.

$$\mathbf{Aspect\ ratio} = \frac{\textit{grand_Axe}}{\textit{petit_Axe}}$$

On l'obtient également en divisant la longueur par la largeur du rectangle englobant la forme.

$$\mathbf{aspect\ Ratio} = \frac{\textit{longueur}}{\textit{largeur}}$$

On peut en déduire la formule de l'**élongation** [5] qui caractérise jusqu'à quel point une forme est allongée ou non (figure 3.11).

$$\mathbf{Elongation} = 1 - \frac{\textit{largeur}}{\textit{longueur}}, \text{Elongation} \in [0,1]$$

Ainsi, une région qui a une forme allongée aura un facteur de forme plus élevé qu'une région qui a une forme plus ronde (d'élongation nulle).

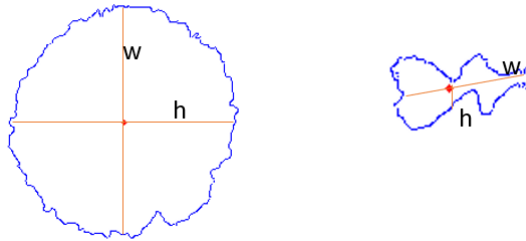


FIGURE 3.11 – Illustration de l'aspect ratio.

Élongation faible

Élongation élevé

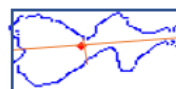


FIGURE 3.12 – Illustration de l'élongation.

Enfin la **rectangularité** [5] représente à quel point une forme est rectangulaire, c'est-à-dire dans quelle mesure elle remplit son rectangle englobant. Elle vaut un pour un objet parfaitement rectangulaire.

$$\mathbf{Rectangularite} = \frac{\textit{Aire_forme}}{\textit{Aire_rectangle_englobant}}$$

Convexité

La convexité [4] d'un objet décrit sa forme en terme de la direction suivant laquelle elle s'incurve. Une surface convexe s'incurve vers l'extérieur, tandis qu'une surface concave s'incurve vers l'intérieur. La convexité est utilisée comme un descripteur pour caractériser la forme d'un objet en reconnaissance de formes. Par exemple,

dans l'analyse d'image d'une main, la convexité ou concavité des doigts peut être utilisée pour identifier les articulations et les mouvements de la main.

Nous la caractérisons à partir de l'enveloppe convexe qui englobe la région étudiée suivant la formule :

$$\text{Convexité} = \frac{\text{perimetre_enveloppe_convexe}}{\text{perimetre_region}}$$

L'enveloppe convexe (voir figure 3.13 : à droite, la région à gauche) d'une image binaire c'est l'ensemble des pixels inclu dans le plus petit polygone convexe qui entoure tous les pixels blancs de la région d'intérêt.



FIGURE 3.13 – Illustration de l'enveloppe convexe (à droite).

3.2.2 Descripteurs de contour

Les descripteurs de contours sont des caractéristiques géométriques qui permettent de quantifier les propriétés des contours des objets dans une image. Les contours peuvent être extraits à partir d'algorithmes de détection de contours tels que les algorithmes de Canny ou de Sobel. Nous aborderons un certain nombre d'entre eux à savoir : le profil, le codage de Freeman, la courbure, l'énergie de flexion et le descripteur de fourier associé aux distances au centre de l'objet.

Profils

Les profils sont obtenus en projetant le contour d'une région donnée sur les axes X et Y du repère (voir figure 3.14) de l'image ; nous obtenons deux fonctions à une dimension, soit

$$Pro_x(i) = \sum_{j=j_{min}}^{j_{max}} f(i, j) \qquad Pro_y(j) = \sum_{i=i_{min}}^{i_{max}} f(i, j), \qquad f(i, j) \text{ représente le contour.}$$

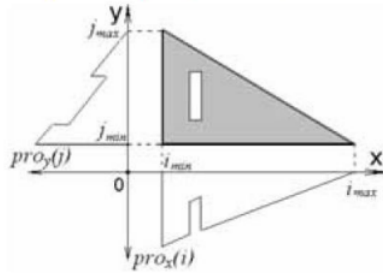


FIGURE 3.14 – Illustration du profil [5]

Descripteurs de Fourier

En traitement d'images, la transformée de Fourier peut être utilisée pour effectuer une analyse fréquentielle de l'image. Cela peut être utile pour extraire des informations sur la distribution de la fréquence dans l'image, ce qui peut être utilisé pour améliorer la qualité de l'image ou pour effectuer des opérations telles que le filtrage.

Dans notre cas, nous aimerions caractériser le contour d'une région en appliquant la transformée de Fourier à la signature de la forme (distance au centre).

Dans [6], la transformée de Fourier est utilisée pour la reconnaissance de formes car c'est un descripteur très robuste face au bruit.

La méthode consiste à mesurer les distances euclidiennes des points du contour de la région à son centre apparent, puis d'appliquer la transformée de Fourier sur ces distances pour trouver les fréquences présentes. Les coefficients obtenus décrivent la forme de manière unique.

Cette méthode de codage de contours est très sensible aux perturbations de l'image et aux variations de résolution. Pour comparer les formes, certaines directions privilégiées peuvent être considérées.

On note aussi que la réduction de dimension des données peut entraîner une perte d'information, il est donc important de choisir un nombre approprié de coefficients pour conserver suffisamment d'informations sur la forme du contour.

Par ailleurs, nous définissons la **rugosité** de la forme qui est l'écart type normalisé entre les distances ; plus cet écart est proche de zéro, plus la forme ou la région étudiée est moins rugueuse et lisse, lorsqu'elle est plus proche de l'unité, la forme est très rugueuse ou très complexe.

Descripteur de courbure

La k-courbure [4] du contour au point (x_i, y_i) est estimée à partir de la k-tangente suivant la formule ci-après :

$$\mathbf{k\text{-courbure}} = \tan^{-1}\left(\frac{y_{i+k} - y_i}{x_{i+k} - x_i}\right) - \tan^{-1}\left(\frac{y_i - y_{i-k}}{x_i - x_{i-k}}\right)$$

En général le calcul ne se fait pas en tout point du contour, c'est pour cela qu'il est considéré comme un descripteur de forme locale.

La figure 3.15 illustre la courbure notée $k(p)$ de la forme d'une région en fonction des points (p) du contour : les zones convexes (respectivement concaves) produisent des courbures positives (respectivement négatives).

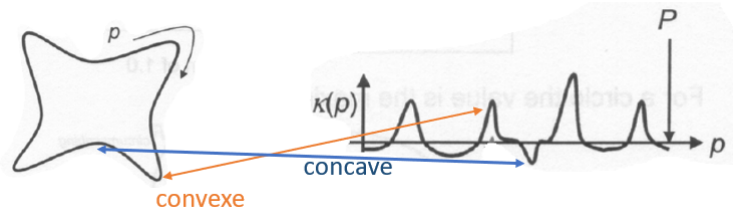


FIGURE 3.15 – Courbure $k(p)$.

Énergie de flexion

L'énergie de flexion E_c [4] associée à un contour peut être définie comme l'énergie nécessaire pour plier une tige à la forme souhaitée et se calcule comme la somme des carrés de la courbure du contour $K(p)$ sur sa longueur L .

$$E_c = \frac{1}{L} \sum_{p=1}^L k(p)^2$$

3.2.3 Autres descripteurs

Le nombre d'Euler

Le nombre d'Euler est un entier associé à toute surface orientable [7]. Il est invariable à la rotation, à la translation, au changement d'échelle ou une déformation continue de la surface.

Il est obtenu comme suit : on trace un graphe planaire connexe sur la surface, ayant s sommets, a arêtes, et f faces, nc : nombre de composantes connexes, nt nombre de trou ; le nombre d'euler est alors l'entier $\mathbf{NE} = \mathbf{s-a+f=nc-nt}$.

Dans le cas d'une image numérique, le nombre d'Euler peut être calculé en comptant le nombre de pixels de certaines configurations de l'objet d'intérêt et du fond d'une image donnée. En effet, elle est basée sur leurs composantes connexes selon la formule suivante :

$E_k = C_k(F) - C_{k'}(B) + 1$ où $C_k(F)$ est le nombre de composantes k -connexes de l'objet, et $C_{k'}(B)$ est nombre de composantes k' -connexes de fond.

Le nombre d'Euler est égal à la différence entre le nombre de composantes connexes de l'objet et le nombre de trous. Ainsi si l'on connaît le nombre de composantes connexes appartenant à l'objet, alors le nombre d'Euler permet de donner automatiquement le nombre de trous, et vice-versa. La figure 4.16 montre trois objets dont le nombre d'Euler vaut respectivement 0, 1 et -1.

6 3 B

FIGURE 3.16 – Illustration du nombre d'euler.

La distance au centre.

Elle permet d'évaluer la répartition spatiale des différentes régions de chaque classe par rapport au centre de la pomme. En effet, elle est calculée en évaluant la distance euclidienne entre le centre de la pomme et le centroïde d'une région donnée.

Elle peut être également utilisée pour caractériser la forme d'une région donnée encore appelée signature de la forme : cette fois-ci c'est la distance euclidienne du contour au centre de masse de la région considérée.

Elle est invariante à la rotation et à la translation.

Pour illustration, la première image de la figure 3.17 nous montre les régions d'une classe donnée, la deuxième montre une droite du centre de la pomme considérée au centre d'une région qui est englobée par un carré ; la troisième image montre la région considérée.



FIGURE 3.17 – Distance au centre.

3.3 Choix des Descripteurs pour l'analyse

Descripteur de forme	Descripteur de contour
Aire	Codage de Freeman
Périmètre, Périmètre Crofton	Courbure
Orientation	Rugosité
Aspect ratio	ProfileX
Elongation	ProfileY
Grand et petit axe	Energie de flexion
Circularité, solidité	
Diamètre équivalent	
Diamètre de Féret	
Compacité	
concavité	
Rectangularité	
Autres Descripteurs	
Distance au centre	Nombre de région
Trou	Nombre d'Euler

FIGURE 3.18 – Les 14 descripteurs retenus.

Nous avons pu définir une vingtaine de descripteurs (voir fig. 3.18) pour pouvoir caractériser les motifs, cependant nous en avons sélectionné 14 (encadrés en rouge) pour analyse à savoir : l'aire, le périmètre, l'aspect ratio, l'élongation, le grand axe, la circularité, la concavité, la rectangularité, la rugosité, le profil, l'énergie, la distance au centre, le nombre de régions et le nombre de trous.

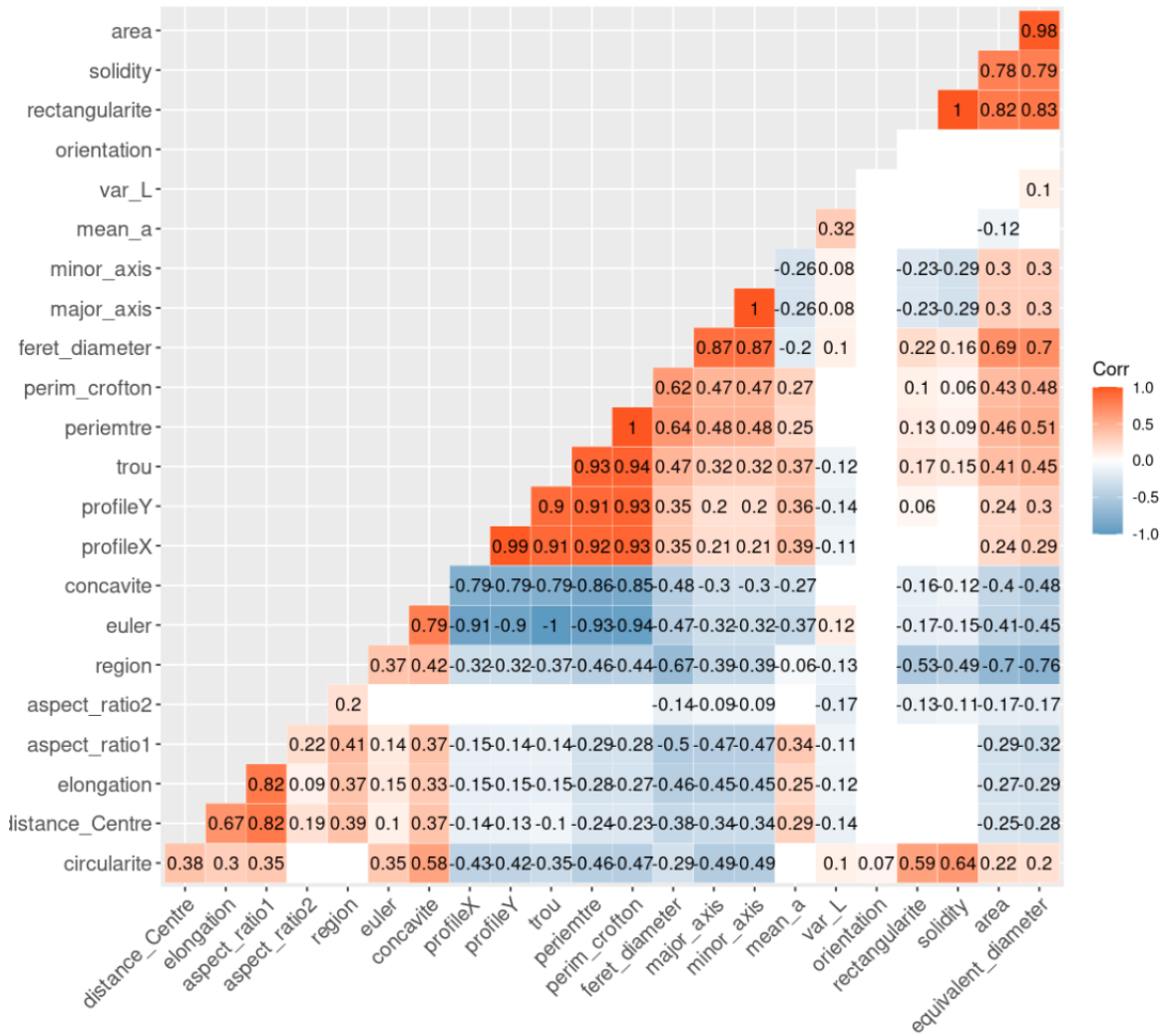


FIGURE 3.19 – Corrélation entre les descripteurs établis.

Certains descripteurs n’ont pas été pris en compte pour analyse à cause des informations redondantes qu’ils apportent. En effet, nous avons calculé la corrélation entre les descripteurs deux à deux afin d’évaluer la redondance des informations : les valeurs colorées en rouge et bleu indiquent de fortes corrélations respectivement positives et négatives et celles colorées en blanc représentent l’absence de corrélation.

En effet, pour **les descripteurs de forme** on a des descripteurs qui fournissent à peu près les mêmes informations ou sont basées sur les mêmes formules.

Nous gardons tout d’abord *l’aire* qui donne l’information sur l’étendue de chaque région. Puis nous éliminons le *périmètre Crofton* au profit du *périmètre* classique, car étant plus précis que le premier qui est généralement utilisé pour trouver approximativement le périmètre d’une forme complexe. En plus il est fortement corrélé à 3 autres descripteurs en plus du périmètre (périmètre, trou, profilX, profilY) comme l’illustre la figure 3.19.

Nous éliminons *l’orientation* à cause du fait qu’il y a des angles positifs et négatifs qui en moyenne ne reflètent pas la réalité à cause des additions et soustractions qui vont s’opérer.

Nous gardons également le *grand axe* qui permet de caractériser la plus grande direction au sein d’un motif, mais pas le *petit axe* vu qu’il est utilisé dans le calcul de *l’aspect ratio*. *L’aspect ratio* permet de distinguer des objets ayant une forme allongée, de ceux ayant une forme plus ronde ou carrée.

Le descripteur *élongation* est aussi pris en compte pour confirmer cet aspect de forme allongée.

En ce qui concerne *la circularité, la compacité* : elles cherchent toutes les deux à caractériser le rapprochement d’un motif à un cercle, donc nous n’analyserons que la *circularité*.

La solidité permet d’évaluer la concavité ou la convexité qu’a un motif, nous ne garderons que le descripteur *concavité*, vu qu’il fera les mêmes opérations.

Et enfin, nous utilisons *la rectangularité*, qui est le seul descripteur qui permet d'évaluer le rapprochement à une forme rectangle.

Pour **les descripteurs de contour** : nous décidons d'utiliser *la courbure*, *la rugosité*, *le profilX*, et *l'énergie de flexion*, car elles décrivent parfaitement le contour de chaque motif.

En effet, le profilY est éliminé car étant fortement corrélé au profilX d'après le tableau de corrélation à la figure 3.19.

Le *codage de Freeman* est éliminé au profit de la *courbure*, car plus simple à analyser au moyen de l'énergie de flexion qui est l'énergie nécessaire pour déformer un matériau en flexion.

Pour terminer nous avons **les autres descripteurs** : *le nombre de régions*, *le nombre de trous* et *la distance au centre* qui sont pris en compte dans les analyses, le nombre d'euler est éliminé au profit du nombre de trous, car étant liés afin d'éviter la redondance d'informations.

3.4 Analyse en Composantes Principales (ACP)

Nos descripteurs morphométriques revêtent une importance cruciale pour comprendre la structure morphologique des motifs de coloration de la chair rouge de pomme. Ils fournissent des mesures quantitatives qui permettent d'évaluer et de caractériser les formes. Cependant, lorsqu'on travaille avec un grand nombre de descripteurs, il devient difficile d'analyser et d'interpréter ces données de manière efficace.

L'Analyse en Composantes Principales (ACP) [8] [9] [10] s'avère être une méthode d'analyse statistique puissante et couramment utilisée pour résoudre cette problématique de multidimensionnalité. L'ACP permet de réduire la complexité des données morphométriques en identifiant les combinaisons linéaires des descripteurs qui capturent le maximum de variance dans l'ensemble des données. Ainsi, elle permet de créer de nouvelles variables appelées composantes principales, qui sont non corrélées entre elles et résumant l'information contenue dans les descripteurs d'origine.

Dans le cadre du stage, nous regardons tout d'abord si les descripteurs morphométriques sont cohérents avec les descripteurs colorimétriques et ensuite, s'ils peuvent nous fournir des informations sur la répartition spatiale des motifs de coloration.

Puis, nous proposons d'appliquer l'ACP sur l'ensemble des descripteurs morphométriques retenus afin de réduire la dimensionnalité des données et de mettre en évidence les structures sous-jacentes.

L'objectif principal est de déterminer si l'ACP peut révéler des patterns morphométriques significatifs et faciliter la discrimination entre différentes familles de pommes. Ce qui permettra plus tard de rechercher s'il existe des gènes responsables.

Dans le cadre de notre étude, nous disposons d'un ensemble de 1536 images de pommes issues de 384 pommiers (génétiquement différents) soit 4 par pommier, au format RGB. La coloration de la chair de chaque pomme a été classifiée en trois classes de manière dynamique à l'aide de Kmeans. De plus, nous avons réussi à extraire les masques correspondant aux motifs de coloration pour chaque classe, et sur chacune d'elles nous avons extrait les 14 descripteurs retenus pour analyse.

Pour obtenir une représentation agrégée par pommier ou arbre, nous avons calculé la moyenne des 14 descripteurs pour toutes les pommes d'un même pommier. Cela nous permet d'avoir une valeur numérique unique pour chaque descripteur et pour chaque pommier, considérant ainsi nos pommiers comme nos individus et les 14 valeurs numériques des descripteurs, comme nos variables.

L'objectif principal de notre analyse est d'appliquer l'Analyse en Composantes Principales (ACP) sur chaque classe individuellement afin de réduire la dimensionnalité des données et de mettre en évidence les principales variations dans les motifs de coloration des pommes. Cette approche nous permettra de comprendre les caractéristiques discriminantes des différentes classes et d'explorer la structure sous-jacente des motifs de coloration, et enfin de sélectionner la classe de coloration la plus intéressante pour plus d'analyses morphologiques.

L'ACP cherche principalement à capturer la plus grande part de la variance totale des données en projetant les observations sur un espace de dimension inférieure. Cela permet de simplifier l'analyse des données en réduisant le nombre de variables tout en préservant au mieux l'information contenue dans l'ensemble initial. L'ACP cherche principalement à capturer la plus grande part de la variance totale des données en projetant les observations sur un espace de dimension inférieure. Cela permet de simplifier l'analyse des données en réduisant le nombre de variables tout en préservant au mieux l'information contenue dans l'ensemble initial.

Avant d'appliquer l'ACP, nous avons standardisé les données des variables pour s'assurer qu'elles sont sur la même échelle. Cela consiste à soustraire la moyenne de chaque variable et à diviser par l'écart type.

Une fois les variables standardisées, il est nécessaire de calculer la matrice de covariance entre les variables morphométriques. La matrice de covariance mesure la variabilité conjointe des variables, tandis que la matrice de corrélation évalue les relations linéaires entre les variables, en prenant en compte les différences d'échelle. Le choix entre la covariance et la corrélation dépendra du contexte de l'analyse et des objectifs visés.

À partir de la matrice de covariance ou de corrélation, on procède à la décomposition en valeurs propres et vecteurs propres. Les valeurs propres représentent la quantité de variance expliquée par chaque composante principale, tandis que les vecteurs propres, également appelés poids, indiquent la contribution de chaque variable à chaque composante principale. Les composantes principales sont en fait des combinaisons linéaires des variables d'origine calculées à partir de la matrice de covariance ou de corrélation ; ils sont ordonnés en fonction de la quantité de variance qu'elles expliquent. La première composante principale capture la plus grande part de la variance, la deuxième composante principale capture la deuxième plus grande part de variance, et ainsi de suite.

Pour déterminer combien de composantes principales retenir, on examine un diagramme en barre, qui représente les valeurs propres en fonction de leur ordre décroissant. On recherche le point d'inflexion où les valeurs propres diminuent de manière significative, ce qui peut indiquer le nombre optimal de composantes principales à retenir. D'autres critères, tels que la proportion de variance expliquée ou des critères d'information, peuvent également être pris en compte pour déterminer le nombre de composantes principales appropriées.

Les données initiales peuvent être projetées sur les composantes principales pour obtenir les coordonnées des observations dans le nouvel espace. Cela permet de représenter graphiquement les observations et d'explorer les relations entre les individus.

3.5 Représentation des motifs par landmarks

Pour une étude plus approfondie des motifs nous avons décidé de redécrire les contours externes ou globaux des motifs de la classe 1, qu'ils soient fragmentés ou non, car ils sont plus représentatifs, solides et variés que ceux des classes 2 et 3 (voir figure 3.28) : il se pourrait que ce soit ceux qui contiennent des informations génétiques.

Et pour ce faire, nous utilisons l'approche par point d'intérêt ou landmarks en anglais. Hao Xu and George W. Bassel l'ont utilisée pour décrire des feuilles de plantes en vue de trouver un lien entre le gène et la forme de feuille [27].

L'extraction de landmarks sur un contour fait référence au processus d'identification et de localisation de points d'intérêt spécifiques le long d'un contour ou d'une bordure dans une image ou une représentation 2D. Ces landmarks peuvent être utilisés dans diverses applications telles que, la détection d'objets, la reconstruction 3D etc ...

Il existe différentes méthodes pour extraire des landmarks à partir d'un contour.

Tout d'abord, il faut extraire le contour de l'objet à l'aide des techniques de segmentation d'images telles que la détection de contours par l'algorithme de Canny ou la méthode du contour actif.

Puis, définir les landmarks que l'on souhaite extraire. Par exemple, pour un contour facial, les landmarks pourraient être les coins des yeux, le bout du nez, les coins de la bouche, etc... dans notre cas, c'est le contour externe des motifs de manière globale.

Ensuite, l'extraction des landmarks proprement dite ; il existe différentes approches pour extraire les landmarks à partir du contour :

- méthode manuelle : sélectionner manuellement les points d'intérêt sur le contour à l'aide d'un logiciel de retouche d'image (comme Fiji Imagej). Cela peut être fastidieux, mais cela donne un contrôle total sur les landmarks souhaités.

- Détection automatique : Vous pouvez utiliser des techniques de traitement d'image et d'apprentissage automatique pour détecter automatiquement les landmarks.

Dans notre approche, nous avons fait d'abord passer des droites verticales entre les points extrémités gauche et droite du motif sur l'axe des abscisses, puis nous avons sélectionné parmi les points d'intersection les points extrémités (haut et bas) qui appartiennent au contour externe. La figure 3.20 présente un exemple de motifs compacts à gauche et à droite, la reconstruction du contour à partir des landmarks extraits.

Une autre technique développée consistait à générer des droites à partir des points du rectangle englobant jusqu'au centre, créant ainsi une sorte de quadrillage étoilé recouvrant le motif. Nous avons conservé tous les premiers points d'intersection les plus proches du rectangle englobant comme landmarks. La figure 3.21 illustre un exemple de motifs compacts à gauche, et à droite la reconstruction du contour par landmarks.

Après avoir extrait les landmarks, il est essentiel de les valider pour s'assurer de leur précision. En comparant les deux reconstructions par landmarks (images à droite des figures 3.20 et 3.21), celle construite par droites verticales parallèles est moins précise sur les côtés : on peut voir que les courbures du motifs sont aplaties. En

revanche, celle construite par des droites étoilées est bien meilleur, car ces droites ont réussi à toucher toutes les zones le long du pourtour du motif. La figure 3.21 illustre 12 motifs et la reconstruction de leur contour par landmarks. Pour les 1536 motifs de classe 1 de coloration extraits sur les données de 2021 et 1743 pour 2022, les formes reconstruites sont présentées dans les figures 3.23 et 3.24.

Ce dernier constitue un résultat que nous adoptons dans la suite de nos travaux.

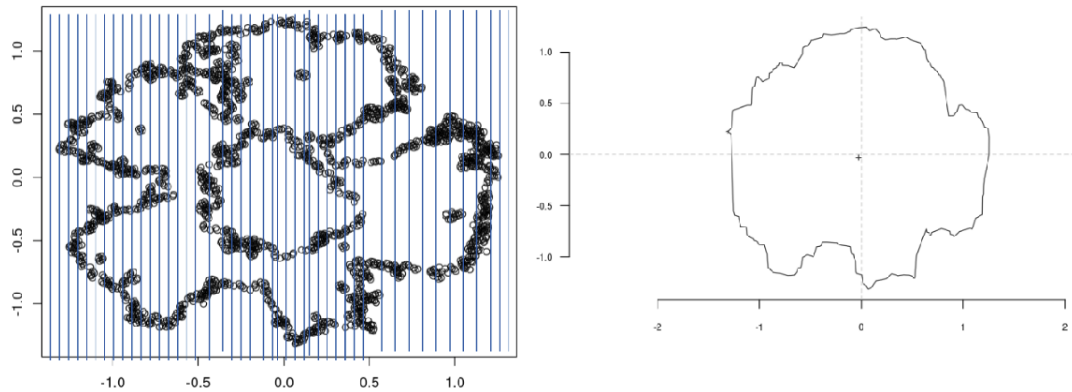


FIGURE 3.20 – Extraction des landmarks au sein des nuages de points (en noir) représentant le motif à l’aide de droites parallèles (en bleu) et à droite la reconstruction du contour à l’aide des landmarks.

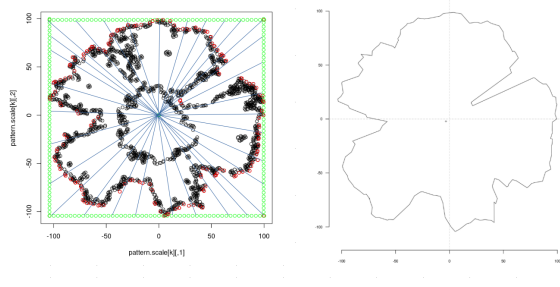


FIGURE 3.21 – Extraction de landmarks à l’aide de droites étoilées (en vert le rectangle englobant, en rouge les landmarks détectés) et reconstruction du contour à droite.

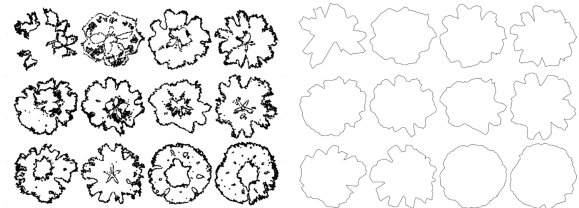


FIGURE 3.22 – Vu sur 12 motifs (à gauche) et la représentation par des landmarks (à droite).

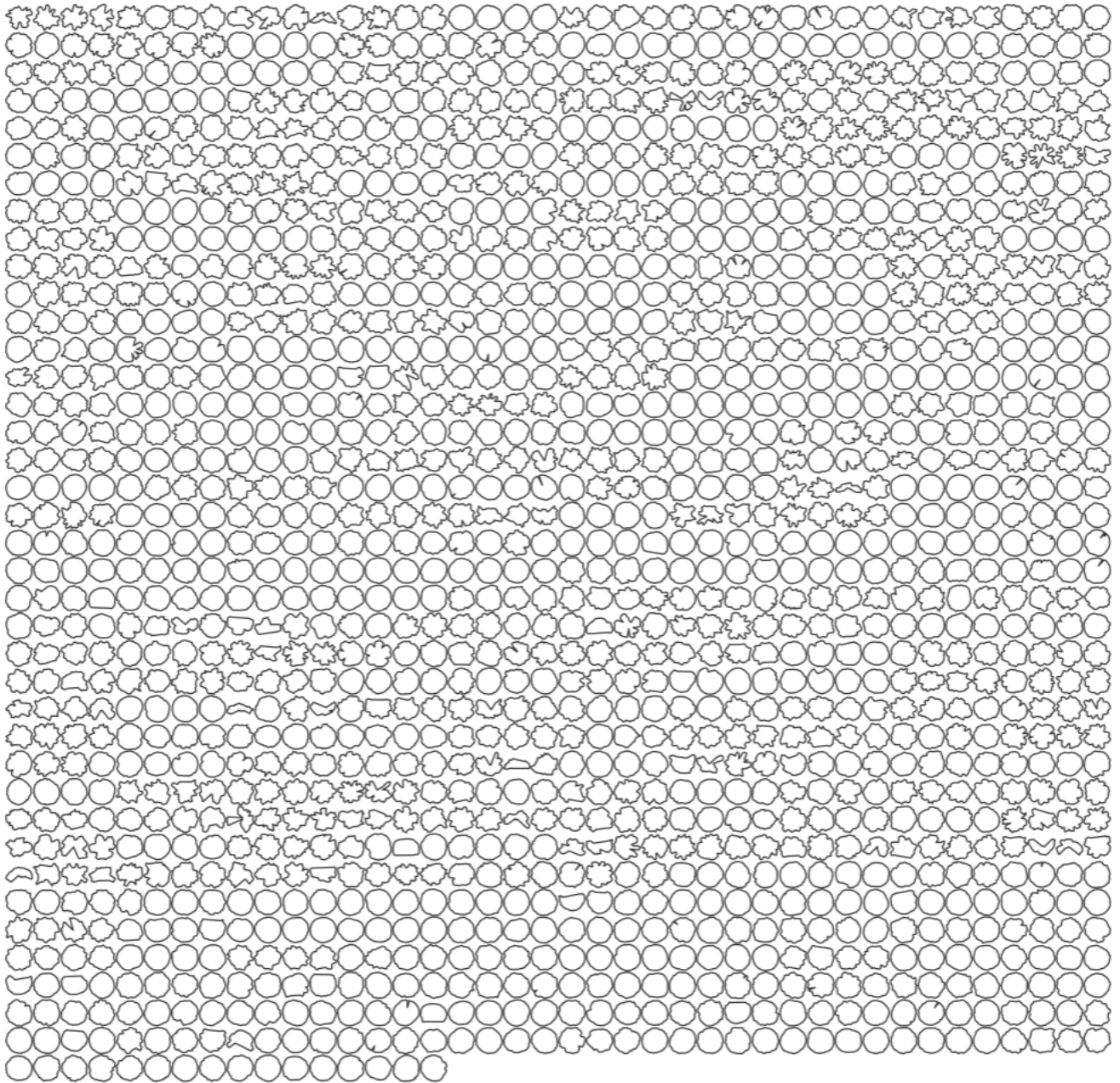


FIGURE 3.23 – Vue d’ensemble de 1536 motifs représentés par des landmarks extraits des données de 2021.

3.5.1 Transformées de Fourier et contours fermés

L’analyse des contours ne nécessite pas nécessairement que le contour sur lequel les points sont échantillonnés soit structuralement défini [26] : ce qui est extrait est l’information géométrique contenue dans le contour lui-même et prise dans son ensemble. Cependant, lors de la comparaison des formes, les contours doivent correspondre à des caractéristiques structurellement similaires. En particulier, les approches basées sur les séries de Fourier sont suffisamment puissantes pour extraire cette information géométrique. Elles reposent sur l’idée de série de Fourier : décomposer une fonction périodique en une somme de fonctions trigonométriques plus simples telles que le sinus et le cosinus. Ces fonctions simples ont des fréquences qui sont des multiples entiers, c’est-à-dire des harmoniques, les unes des autres. Les harmoniques de basse fréquence fournissent des approximations des tendances à grande échelle de la fonction périodique d’origine, tandis que les harmoniques de haute fréquence s’adaptent à ses variations à petite échelle.

Les séries de Fourier peuvent être utilisées en morphométrie, parmi de nombreuses autres applications dérivées, car les contours fermés peuvent être considérés comme des fonctions périodiques. En effet, si l’on décide de parcourir un contour en commençant par un de ses points, nous repasserons encore et encore par le même point de départ, et donc des fonctions périodiques peuvent décrire ce contour.

On peut citer entre autres : la distance de n’importe quel point du contour au centroïde de la forme, la

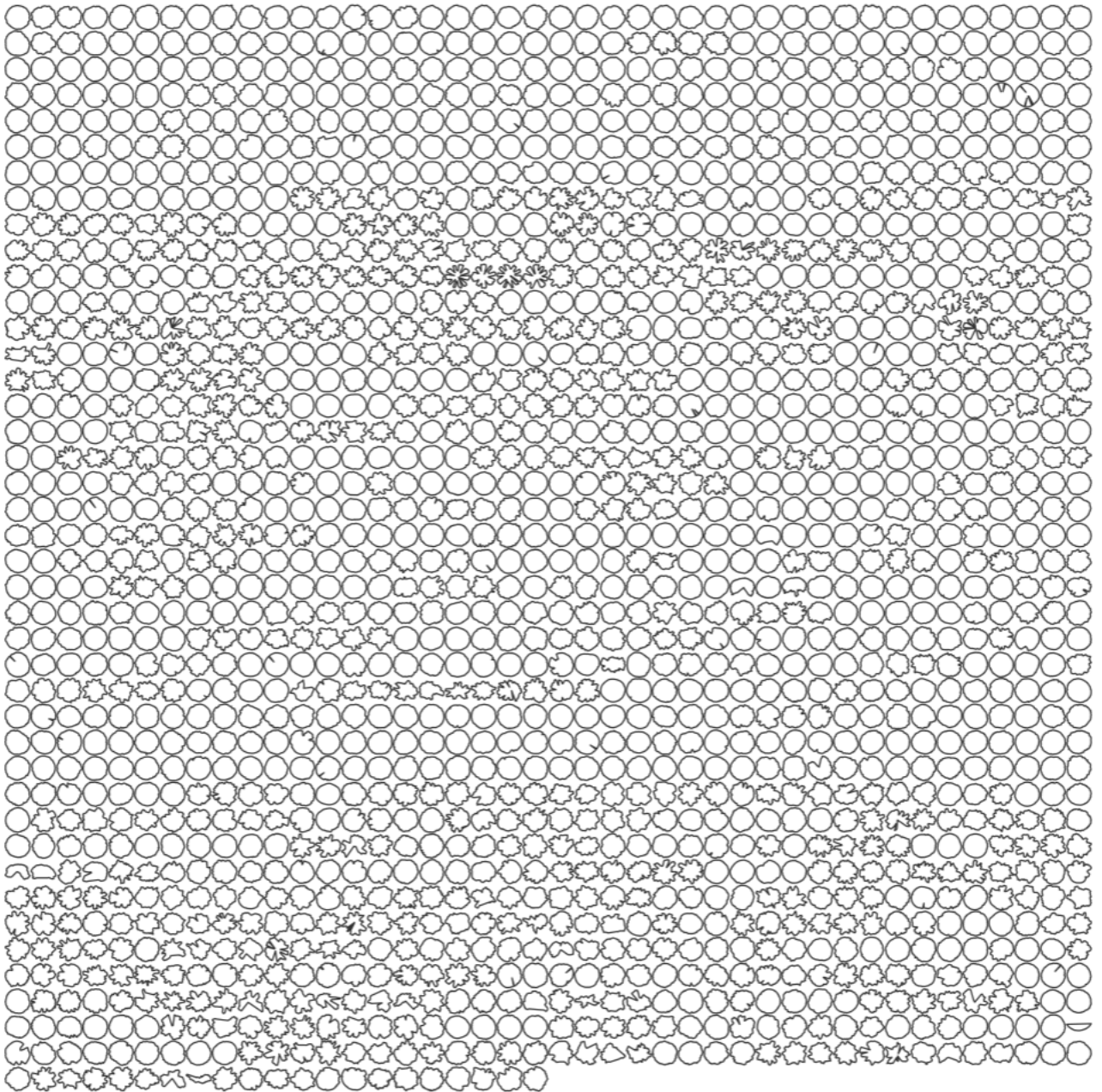


FIGURE 3.24 – Vue d’ensemble de 1743 motifs représentés par des landmarks extraits des données de 2022.

variation de l’angle tangent de tout point du contour ou les coordonnées (x, y) sur le plan. Une ou plusieurs fonctions périodiques sont alors obtenues et peuvent être décomposées (et donc décrites) par des séries de Fourier. Ces trois différentes méthodes ou fonctions sont disponibles dans Momocs [26] (package de R), et sont appelées ci-après "variation du rayon", "angle tangent" et "analyse elliptique". L’**analyse elliptique** est la plus utilisée et la plus adaptée pour décrire les contours, c’est celle que nous allons utiliser dans la suite de nos travaux.

Toutes les décompositions de Fourier aboutissent à une somme harmonique de fonctions trigonométriques pondérées par des coefficients harmoniques. Ils sont généralement normalisés pour éliminer les différences homothétiques, translationnelles ou rotationnelles entre les formes. Deux ou quatre coefficients, selon l’approche utilisée, sont obtenus pour chaque harmonique calculée et peuvent ensuite être considérés comme des variables quantitatives. La fréquence de Nyquist empêche l’utilisation de plus d’harmoniques que la moitié du nombre de points ajustés, ce qui constitue donc leur limite supérieure.

Analyse elliptique de Fourier

La dernière approche présentée ici est due à Giardina et Kuhl (1977) ainsi qu’à Kuhl et Giardina (1982), qui ont développé une méthode permettant d’ajuster séparément les coordonnées x et y d’un contour projeté sur un plan. Cette méthode est devenue très populaire car elle présente de nombreux avantages par rapport aux

autres approches basées sur la transformée de Fourier : il n'est pas nécessaire que les points soient régulièrement espacés, pratiquement n'importe quel contour peut être ajusté (voir Rohlf et Archie 1984; Crampton 1995; Renaud et Michaux 2003) et les coefficients peuvent être rendus indépendants de la position du contour et normalisés en fonction de sa taille.

Soit T le périmètre d'un contour fermé donné, considéré ici comme la période du signal. On pose $\omega = 2\pi/T$ comme pulsation. Ensuite, l'abscisse curviligne t varie de 0 à T . On peut exprimer $x(t)$ et $y(t)$ de la manière suivante :

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t) \text{ avec } a_n = \frac{2}{T} + \int_0^T x(t) \cos(n\omega t) dt,$$

$$b_n = \frac{2}{T} + \int_0^T x(t) \sin(n\omega t) dt$$

de même :

$$y(t) = \frac{c_0}{2} + \sum_{n=1}^{+\infty} c_n \cos(n\omega t) + d_n \sin(n\omega t) \text{ avec } c_n = \frac{2}{T} + \int_0^T y(t) \cos(n\omega t) dt,$$

$$d_n = \frac{2}{T} + \int_0^T y(t) \sin(n\omega t) dt$$

Dans l'analyse elliptique de Fourier, quatre coefficients par harmonique sont obtenus, deux pour x et deux pour y . La figure 3.25 montre 3 harmoniques de base caractérisant le contour d'une forme en coeur qui sont représentées par $4 \times 3 = 12$ coefficients à partir desquels on pourra reconstruire la forme en coeur.

On peut utiliser la première harmonique, celui qui définit l'ellipse qui s'ajuste le mieux, pour normaliser les coefficients harmoniques et les rendre invariants par rapport à la taille et à la rotation. Les coefficients harmoniques peuvent également être normalisés par rapport à la position de la première coordonnée du contour.

Il est à noter que plus on a d'harmoniques, plus on parvient à décrire parfaitement, par des coefficients de Fourier elliptiques, le contour d'une forme, comme l'illustre clairement la figure 3.26 : des exemples démontrent que l'on peut aller jusqu'à utiliser au moins 30 harmoniques pour décrire le contour de l'exemple de forme.

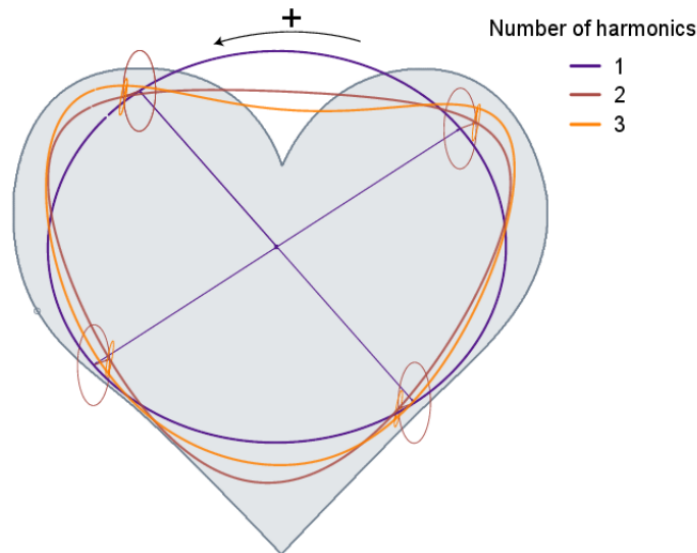


FIGURE 3.25 – Trois harmoniques sont ici représentées à quatre emplacements sur le contour d'origine [26].

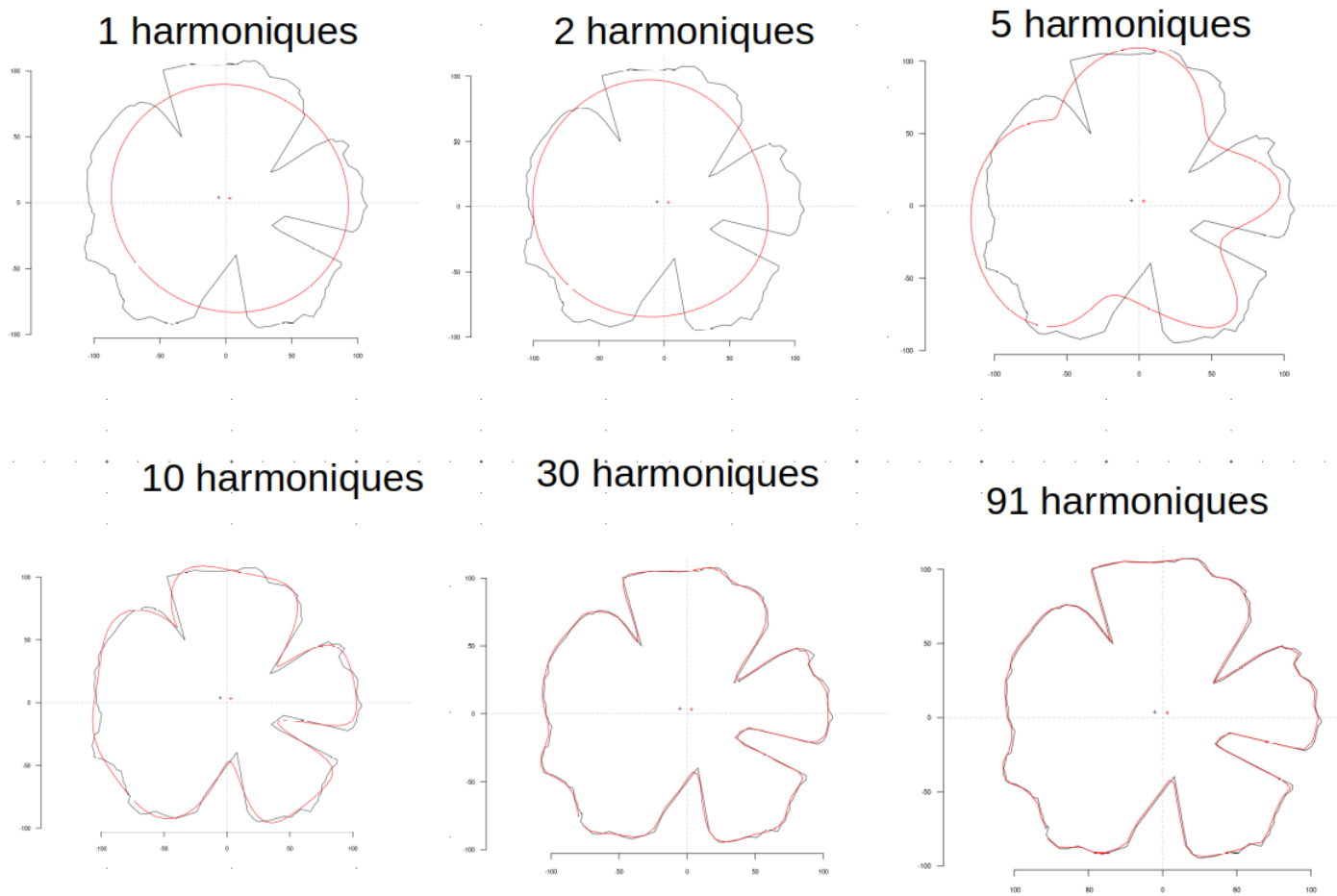


FIGURE 3.26 – Influence du nombre d’harmoniques sur la reconstruction du contour à partir des coefficients de ces harmoniques.

3.6 Présentation et analyse des résultats

Pour étudier les motifs des individus, nous utilisons les données de colorations recueillies par mon collègue doctorant P. Bouillon lors de ses manipulations. Pour toutes les pommes de chaque arbre il calcule l’intensité de coloration de la chair rouge des pommes, qu’il resume d’une part en la moyenne de l’intensité de coloration de la rougeur de la chair des pommes d’un même arbre nommé **mean_a**, et d’autre part leur variance nommée **var_L** en tant que mesure de la distribution des couleurs de la chair.

La figure 3.27 illustre un nuage de points représentant la coloration des individus (pommes) notamment les données de variance **var_L** en fonction des données de moyenne **mean_a** : plus la variance (**var_L**) est grande plus on remarque des motifs à contour accidenté, parfois de forme étoilée ou florale dans la chair vu que la chair est associée à une colocalisation de zones rouges et blanches au sein du fruit ; des valeurs faibles (de **var_L**) indiquent une coloration diffuse centrée autour d’une couleur dominante (soit blanche ou soit rouge clair ou rouge) : dans ce dernier cas on remarque que les motifs de la classe 1 extrait pour une couleur dominante blanche sont plus étendus et compacts ; pour la couleur rouge des motifs de faible étendue et très accidenté et pour le rouge clair des motifs plus ou moins étendus et de contour accidenté également. Des valeurs de **var_L** moyenne correspondent toujours à une colocalisation de zones rouges et blanches au sein de la chair avec une petite dominance de l’une des couleurs rouge ou blanche : ce qui amène à y retrouver parfois des motifs de forme florale de grande étendue pour la classe 1.

En ce qui concerne les données **mean_a** : plus elle est grande plus la chair de la pomme est rouge, plus elle est faible, plus la chair est blanche, une valeur intermédiaire sera celle dont la chair contiendra des motifs.

Les nuages de points se répartissent sous forme d’un arc recourbé vers le bas où la première portion (à gauche) du nuage de points de coloration blanche représente les pommes de chair blanche ayant moins de motifs, la portion intermédiaire (milieu) du nuage de points, de coloration blanche et rouge représente les pommes de chair rouge incluant du blanc contiennent suffisamment de motifs, la dernière portion (à droite) du nuage de points, de coloration rouge représente les pommes de chair rouge avec peu de motifs.

Une fois les motifs ainsi que leurs valeurs morphométriques extraites, toutes ses données de coloration seront

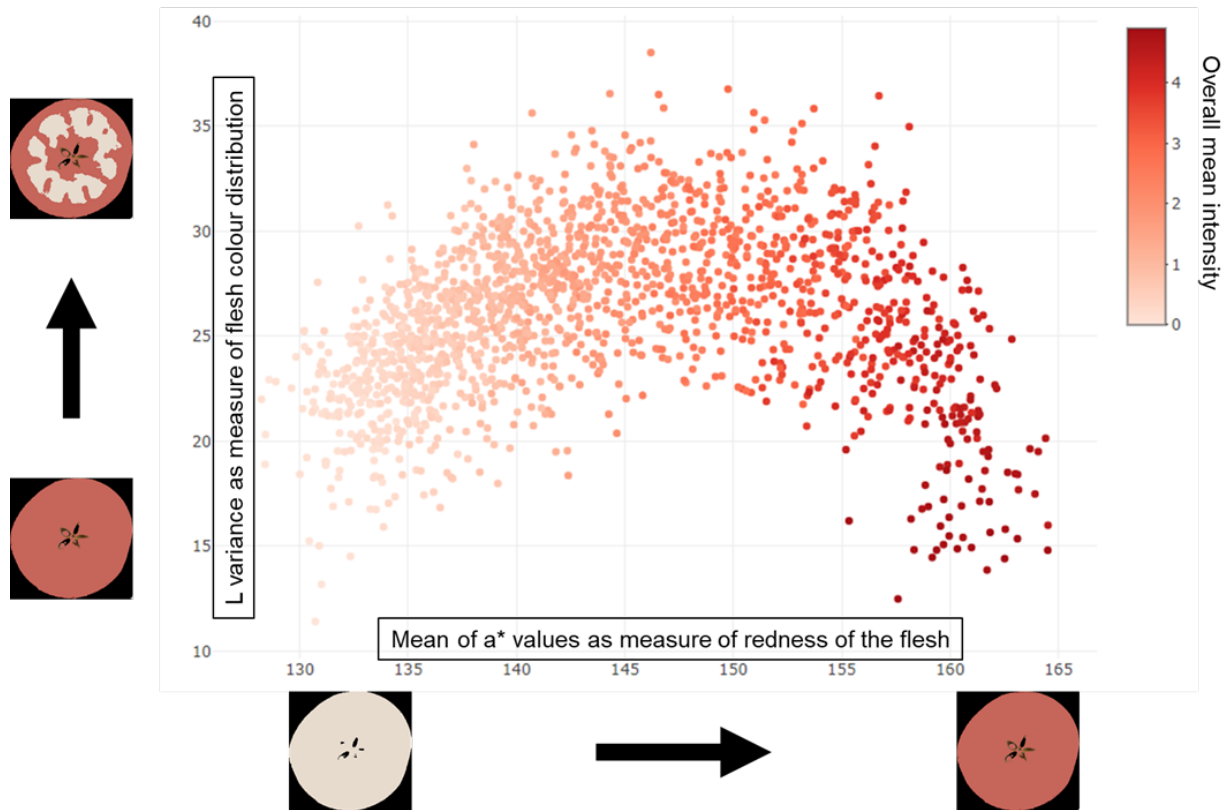


FIGURE 3.27 – Représentation des individus par les données de colorations.

utilisées pour analyser les résultats obtenus.

L'application de l'ACP aux données morphométriques moyennées par arbre (soit 384 individus moyens que nous avons enregistrés sur les données de 2021) et 14 variables (descripteurs morphométriques) nous a permis d'identifier des regroupements de motifs ou groupes morphométriques de forme presque identiques. Nous les représentons avec les données de coloration `mean_a` associées aux arbres d'une part et d'autre part à l'aide de la méthode `elbow` et de `kmeans` nous colorions différemment les différents regroupements détectés.

Ensuite, les informations associées à l'ACP sont utilisées pour identifier les descripteurs morphométriques les plus dominants dans chaque groupe.

3.6.1 Application de l'ACP au données issues des descripteurs traditionnels

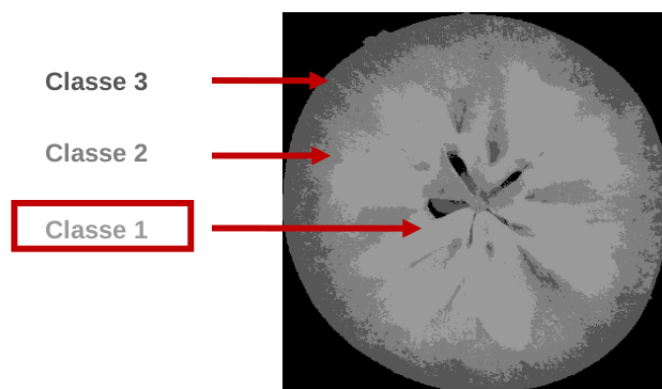
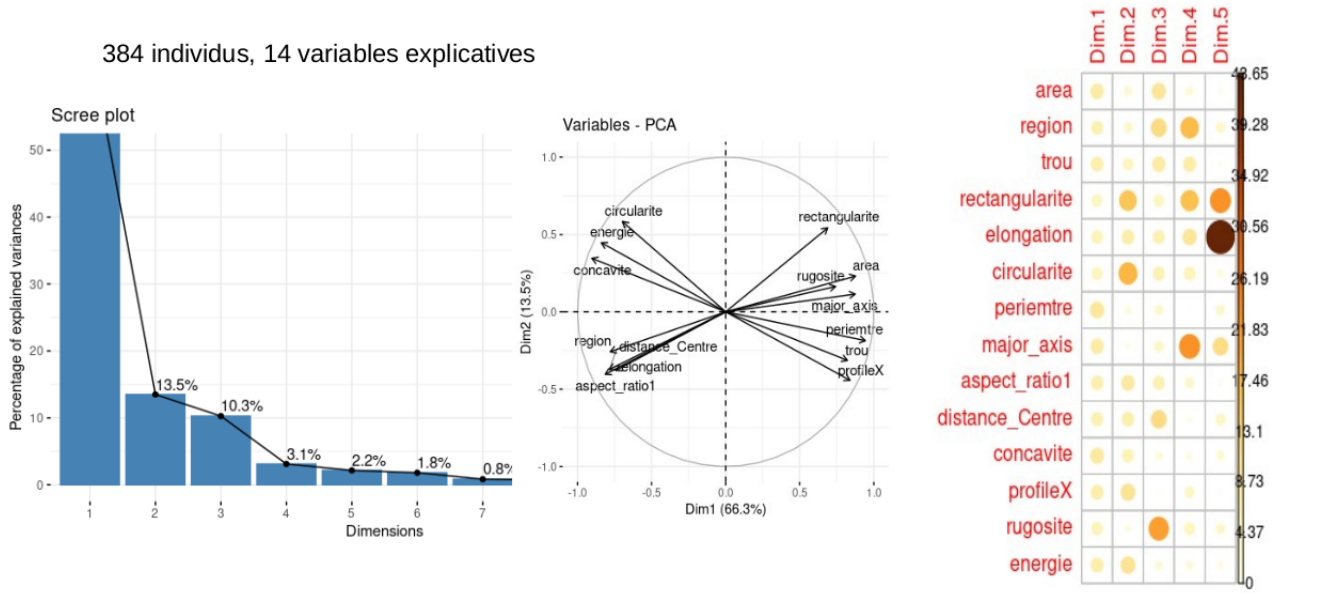


FIGURE 3.28 – Motifs de la classe 1.



Kmean k=4 classe 1

FIGURE 3.29 – Motifs de classe 1 : Pourcentage de variances expliquées, cercle de corrélation, contribution des descripteurs à chaque axe.

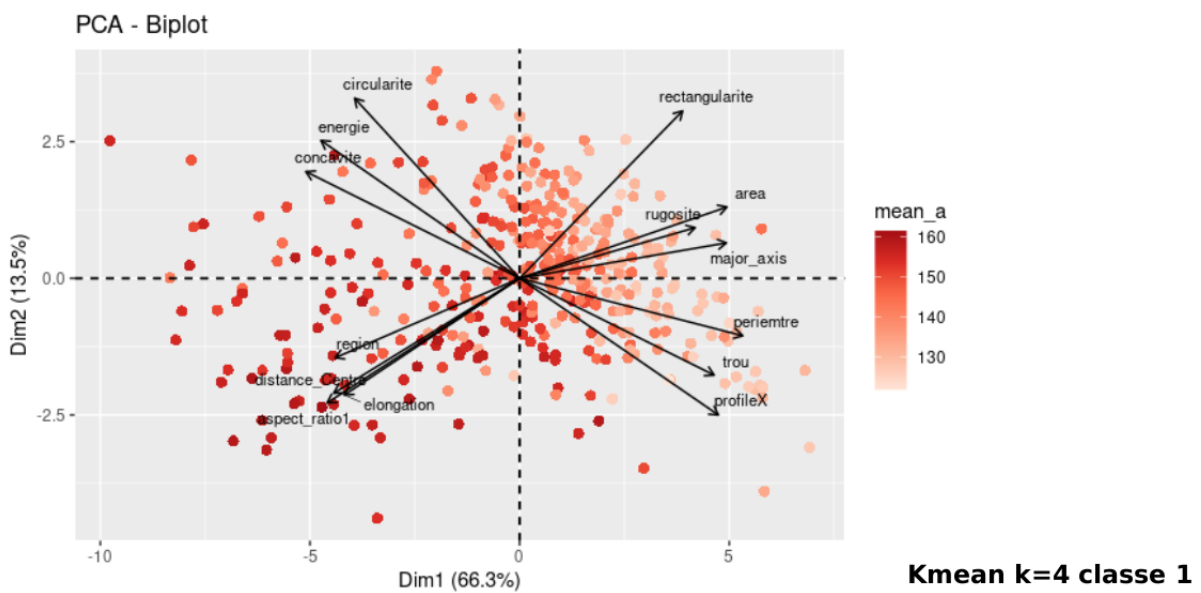


FIGURE 3.30 – Distribution des individus par coloration suivant les axes principaux 1 et 2.

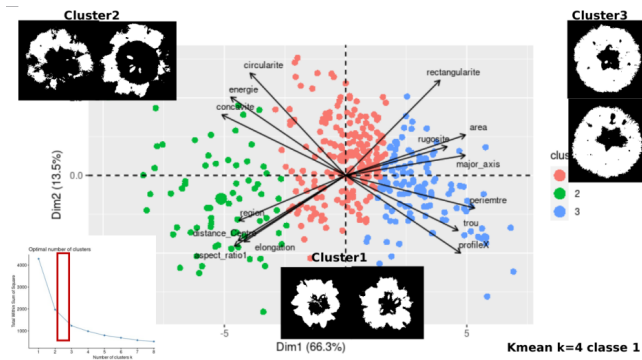


FIGURE 3.31 – Motifs de classe 1 : regroupement des individus en 3 groupes dans l’espace construite.

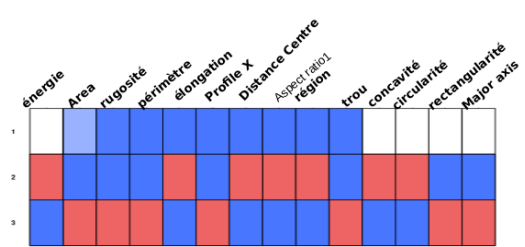


FIGURE 3.32 – Variable dominante dans chaque groupe.

Classe 1

Dans cette partie, nous présentons tous les résultats de l’application des ACP sur les données morphométriques de la classe 1 ; un exemple d’illustration de motif de classe 1 est illustré à la figure 3.28 : on peut voir que le motif est étendu, solide et compact.

Tout d’abord, la figure 3.29 illustre : le pourcentage de variable expliquée par dimension (diagramme en bâton) où la dimension ou l’axe principal 1 contient 66.3 % de variance et le second axe 13.5 % ; cela signifie que la première dimension (Dim 1) explique plus de la moitié de la variance totale des données, ce qui la rend très significative dans la représentation des données. La deuxième dimension (Dim 2) explique environ 13.5% de la variance, ce qui est également considérable, bien que moins important que la première dimension.

En revanche, les dimensions 3, 4 et 5 expliquent une faible proportion de la variance des données. Cela suggère que ces dimensions sont moins informatives et pourraient être négligées dans l’analyse, car elles ne contribuent pas beaucoup à la description des données.

Au centre de la figure(fig. 3.29), nous illustrons le cercle de corrélation qui est un outil graphique utilisé en ACP pour interpréter les relations entre les variables originales et les composantes principales [8].

Les variables qui sont fortement corrélées avec une composante principale contribuent davantage à cette dimension, tandis que les variables avec une faible corrélation ont moins d’influence sur cette dimension.

Chaque variable originale est représentée par un vecteur dans le cercle et la corrélation entre les variables et les composantes principales est indiquée par la position et l’orientation de ces vecteurs dans le cercle ; seul l’axe majeur est légèrement corrélé avec l’axe principal 1.

Les vecteurs (ou variables) proches les uns des autres présentent une forte corrélation entre eux, on peut citer l’élongation et l’aspect ratio puis la rugosité et l’aire (area).

La norme d’un vecteur indique la force de la corrélation. Mis à part la rugosité, les normes des autres vecteurs sont pratiquement les mêmes ce qui montre que les descripteurs associés ont la même importance relative dans la construction des composantes principales.

Le tableau à l’extrême droite illustre les contributions des variables aux différents axes principaux ; en regardant les axes 1 et 2 (en dehors de la rectangularité et de la circularité, qui sont légèrement plus élevées sur l’axe 2), on peut constater que les contributions sont plus ou moins pareilles.

À la figure 3.30 on peut observer la répartition des individus suivant leur coloration (mean_a), les motifs issus de pommes colorées se retrouvent du côté gauche de l’axe vertical ou axe 2 et à droite ceux des pommes peu colorés ; les motifs des pommes colorées seront mieux expliqués par des descripteurs tels que : la circularité, la concavité, région (nombre de régions), distance au centre, aspect ratio et élongation et les motifs de pommes peu colorées seront expliquées par : la rectangularité, area (aire), rugosité, major_axis, périmètre, trou et profileX.

Par ailleurs, nous appliquons une classification kmeans (k : le nombre de groupes ou classes ayant été préalablement extrait par la méthode d’elbow) sur la répartition des individus afin d’évaluer les regroupements qu’on peut y observer : on retrouve 3 groupes de motifs coloriés en vert, rouge et bleu comme l’illustre la figure 3.31 (à droite) le groupe en bleu correspond aux motifs des pommes peu colorées et celles colorées (groupe vert et rouge) se sont réparties en 2 groupes.

En extrayant quelques exemples des groupes (voir figure 3.31), on retrouve des motifs solides et étendus (les vecteurs area et périmètre confirme cela) et de contour plus ou moins circulaires (groupe en bleu) ; des motifs de forme très accidentée (en vert) et des motifs de forme floraux (en rouge), un peu à l’intermédiaire des deux premiers groupes.

Enfin, pour la figure 3.32 quant à elle confirme l’influence des descripteurs dans le regroupement : la couleur rouge dans le tableau correspond à une forte dominance, le bleu à une dominance moyenne et le blanc à une faible dominance des variables au sein d’un groupe.

On voit que le groupe en rouge (ou cluster 1) est très peu influencé (ceux en bleu) ou pas du tout (en blanc) par les variables utilisées du fait de sa position intermédiaire par rapport aux variables. Comme prévu, les variables que nous avons citées plus haut comme explicatives pour les pommes peu colorées sont celles qui dominent dans le groupe en bleu (cluster 3). Pareille pour le groupe en vert (cluster 2).

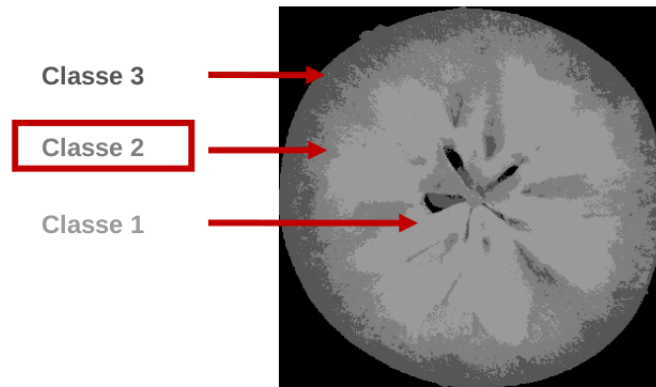
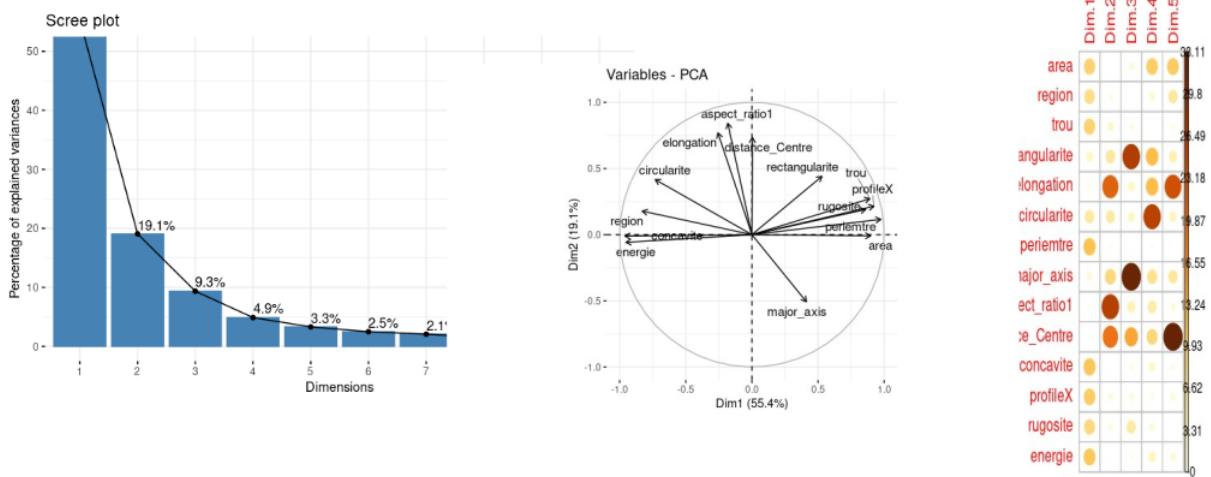


FIGURE 3.33 – Motifs de la classe 2.



Kmean k=4 classe 2

FIGURE 3.34 – Motifs de classe 2 : pourcentage de variances expliquées(à gauche), cercle de corrélation (centre), contribution des descripteurs à chaque axe (à droite).

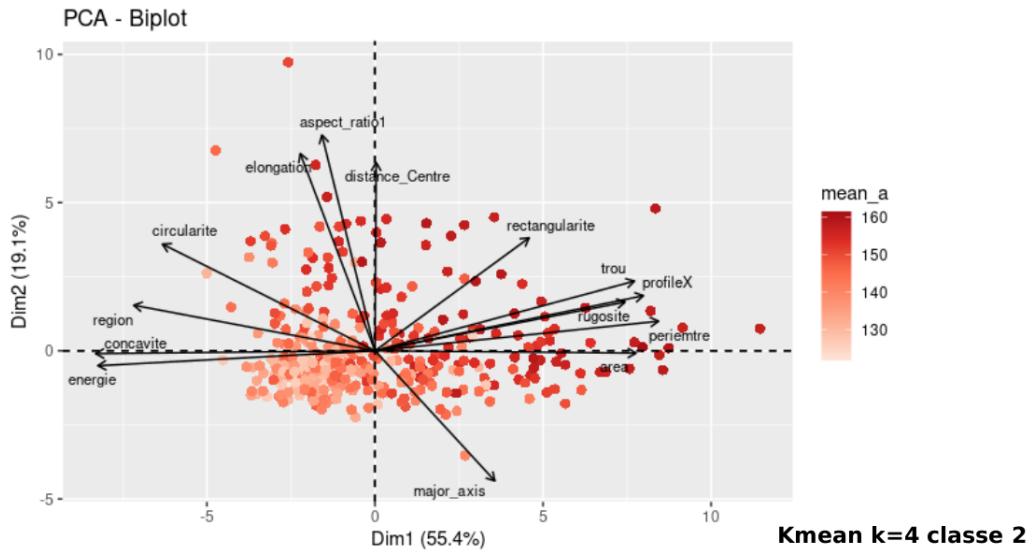


FIGURE 3.35 – Distribution des individus par coloration suivant les axes principaux 1 et 2.

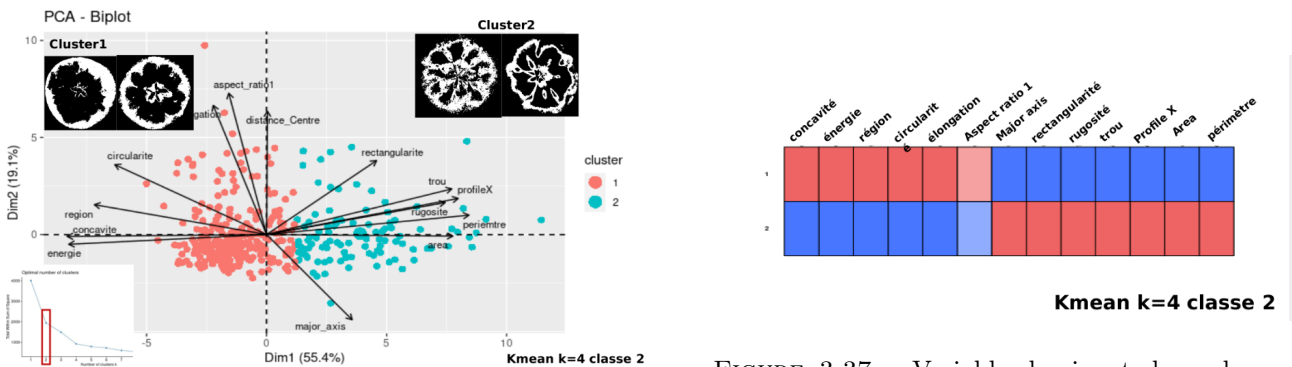


FIGURE 3.36 – Motifs de classe 2 : regroupement des individus en 3 régions dans l’espace construit.

FIGURE 3.37 – Variable dominante dans chaque groupe.

Classe 2

En ce qui concerne les motifs de la classe 2 dont un exemple est présenté à la figure 3.33 ; la figure 3.34 illustre : le pourcentage de variance expliquée par dimension (diagramme de bâton en bleu) où la dimension ou l’axe principal 1 contient 55.4% de variance et le second axe 19.1 % ;

Cela signifie que la première dimension explique plus de la moitié de la variance totale des données, ce qui la rend très significative dans la représentation des données. La deuxième dimension explique environ 19.1 % de la variance, ce qui est également considérable, bien que moins important que la première dimension.

En revanche, les dimensions 3, 4 et 5 expliquent une faible proportion de la variance des données. Cela suggère que ces dimensions sont moins informatives et ne seront pas importantes dans l’analyse, car elles ne contribuent pas beaucoup à la description des données.

Au centre de la figure (fig. 3.34), le cercle de corrélation montre une forte corrélation entre l’axe 1 et la variable area, d’une part, entre les variables concavité et énergie d’autre part, et aussi entre rugosité et profileX.

Le tableau à l’extrême droite illustre les contributions des variables aux différents axes principaux ; en regardant les axes 1, on peut constater que les contributions sont plus ou moins pareilles et l’axe est un peu plus influencé par les variables telles que : élongation, aspect ratio, distance au centre.

A la figure 3.35, la répartition des individus suivant leur coloration montre que les motifs issus de pommes peu colorées se retrouvent du côté gauche de l’axe vertical ou axe 2 et de l’autre ceux des pommes colorées ; les motifs des pommes colorées seront mieux expliqués par des descripteurs tels que : area, profile, rugosité et périmètre et les motifs de pommes peu colorées seront expliqués par : la concavité, énergie et trou.

La classification kmeans appliquée sur la répartition des individus afin d’évaluer les regroupements qu’on peut y observer montre 2 groupes de motifs coloriés en vert et rouge comme l’illustre la figure 3.36 ; (à droite) le groupe en rouge correspond aux motifs des pommes peu colorées et celles colorées (groupe en vert).

Les exemples de ces groupes montrent des motifs solides, de contours externes circulaires (groupes en rouge), le vecteur circularité confirme cela et des motifs de forme accidentée (groupe vert) avec assez de trous, les vecteurs trou, area et périmètre le confirme.

Enfin, la figure 3.37 confirme l'influence des descripteurs dans le regroupement : on voit que les motifs des groupes sont bien influencés par les descripteurs précités, et d'autres descripteurs qui ne sont pas visuellement identifiables sur les motifs.

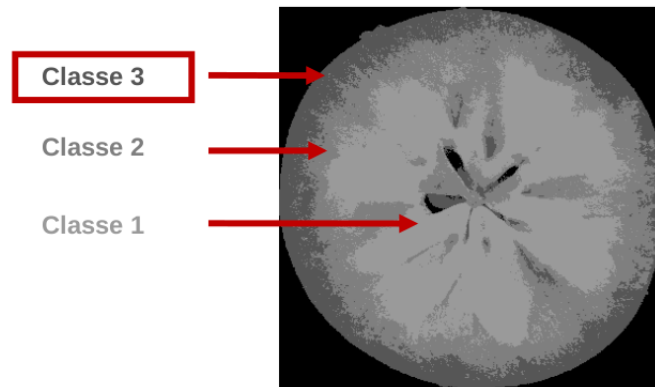


FIGURE 3.38 – Motifs de la classe 3.

Classe 3

Pour la classe 3 (exemple à la figure 3.38), les résultats de l'ACP donnent un pourcentage de variables expliquées par dimension (diagramme de bâton) où la dimension ou l'axe principal 1 contient 48.5% de variance et le second axe 23.6 % (voir figure 3.39) ;

Cela signifie que la première dimension explique 48.5% de la variance totale des données, ce qui la rend très significative dans la représentation des données. La deuxième dimension explique environ 23.6 % de la variance, ce qui est également considérable, bien que moins important que la première dimension.

Au centre de la figure (fig. 3.39), le cercle de corrélation montre une forte corrélation entre l'axe 1 et la variable area et périmètre d'une part et concavité en sens inverse. L'axe 2 quant à lui est fortement corrélé avec la distance au centre.

A la figure 3.40, la répartition des individus suivant leur coloration montre que les motifs issus de pommes peu colorées se retrouvent du côté gauche de l'axe vertical ou axe 2 et de l'autre ceux des pommes colorées ; les motifs des pommes colorées seront mieux décrits par des descripteurs tels que : area, trou, profil, périmètre rugosité et rectangularité et les motifs de pommes peu colorées seront décrits par : la concavité et l'énergie.

La classification kmeans sur la répartition des individus montre 3 groupes de motifs (figure 3.41) coloriés en vert, rouge et bleu ; (à droite) le groupe en rouge correspond au motif des pommes peu colorées et ceux colorées (groupe en vert).

Les exemples de ces groupes montrent des motifs difficiles à cerner vu que généralement, ils se répartissent sur la périphérie des pommes. Seul le cluster 2 ou groupe en vert domine en terme d'aire comme le montre le vecteur area.

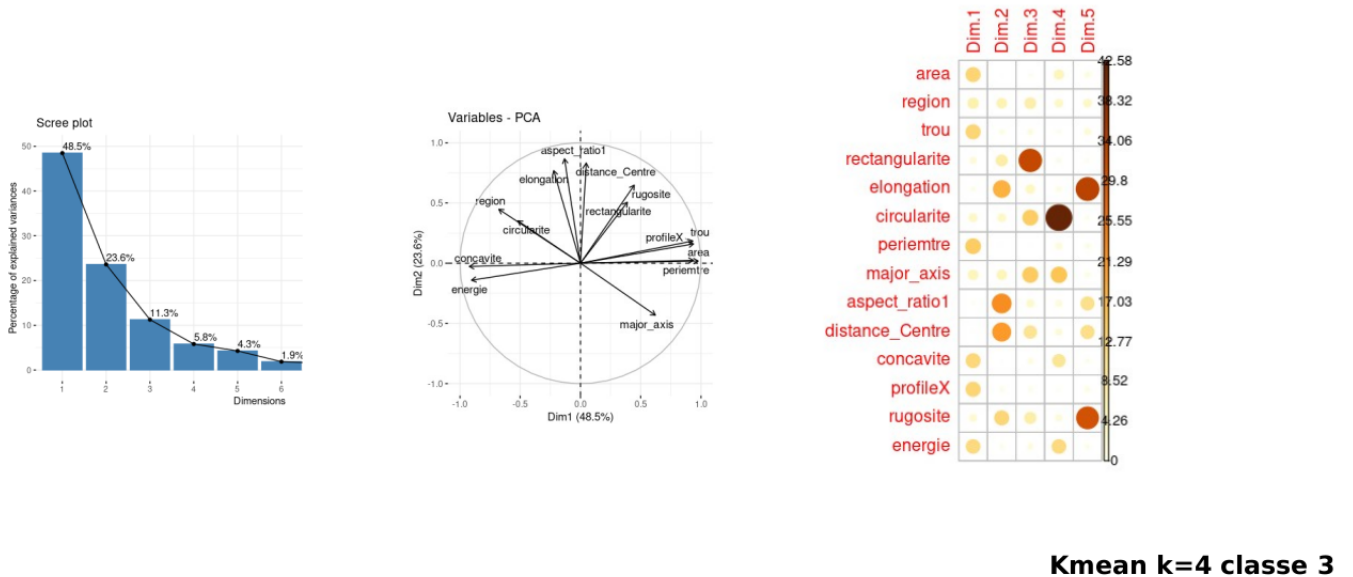


FIGURE 3.39 – Motifs de classe 3 : pourcentage de variance expliquée, cercle de corrélation, contribution des descripteurs à chaque axe.

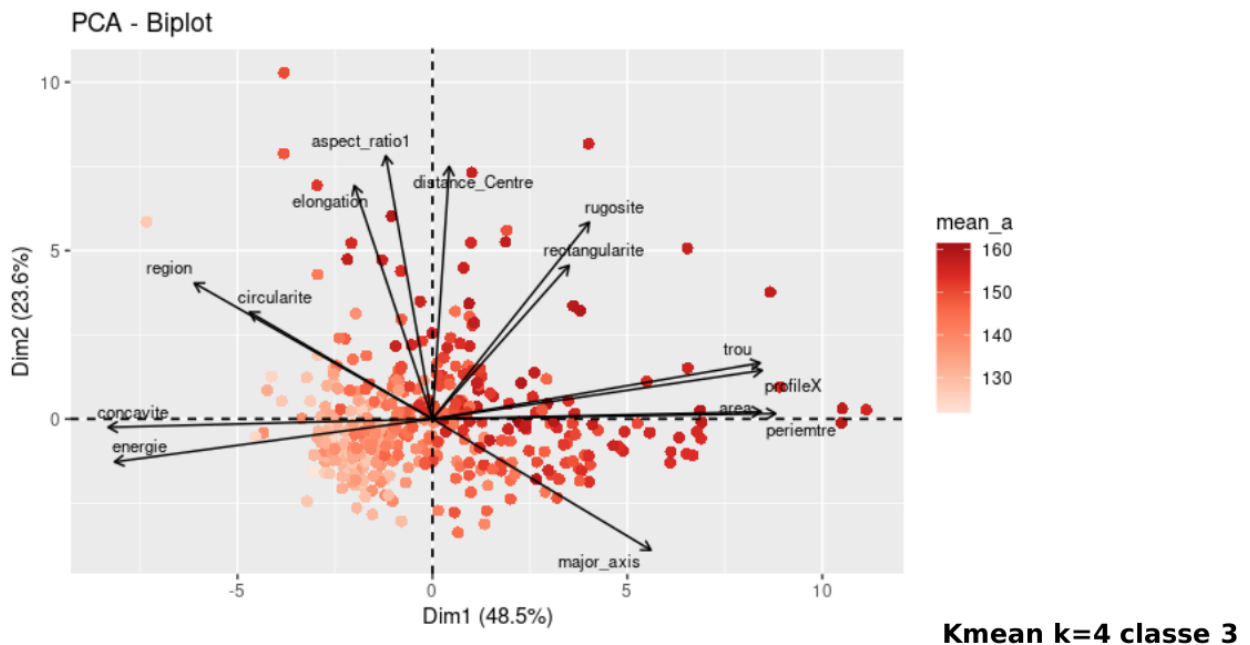


FIGURE 3.40 – Distribution des individus par coloration suivant les axes principaux 1 et 2.

En vue de s’assurer que nous sommes logiques dans les précédents résultats c’est-à-dire l’application de l’ACP sur les données morphométriques extraites des motifs obtenus à partir de la classification avec kmeans k=4 des images RGB de tranches de pommes, nous appliquons la même procédure sur les motifs extraits cette fois avec une classification avec kmeans k = 5.

Nous nous focalisons que sur la classe 1 (pour kmeans k = 5) : nous obtenons pratiquement les mêmes regroupements de motifs avec 3 groupes également. Et pour évaluer la similarité des regroupements pour les 2 types de classifications, nous avons procédé à la caractérisation de chacun des groupes en fonction de leurs descripteurs. Pour ce faire, nous avons utilisé la fonction 'catdes' du package 'factomineR' de R, qui nous a permis d’explorer et de mettre en évidence les caractéristiques spécifiques telles que la moyenne et l’écart type associées au données des descripteurs de chaque cluster ou groupe.

Les valeurs obtenues sont presque identiques pour les deux types de données, indiquant des similitudes importantes dans la répartition des caractéristiques morphologiques entre ces deux classifications. Sauf les descripteurs tels que trou dans le cluter 1 et 2 (ou groupe 1 et 2), area dans le groupe 2 élongation et distance au centre dans le groupe 3 ne sont pas identiques.

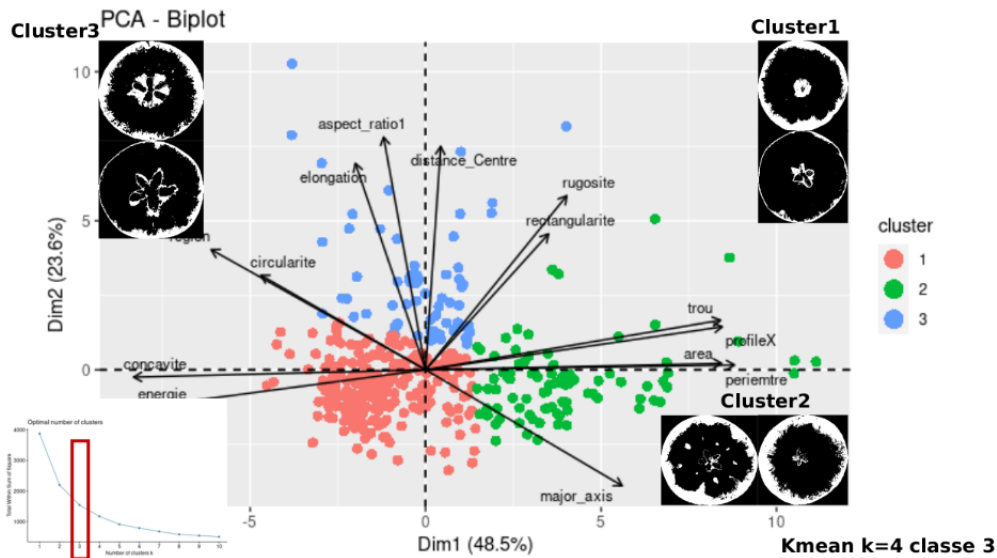


FIGURE 3.41 – Motif de classe 3 : regroupement des individus en 3 régions dans l’espace construit.

Cette constatation suggère que le choix du nombre de classes de classification (k-means $k = 4$ et $k = 5$) n’affecte pas de manière significative les résultats obtenus dans notre étude sur les motifs de classe 1.

3.6.2 Application de l’ACP sur les descripteurs de fourier harmonique

Dans cette étude, nous avons entrepris d’appliquer l’Analyse en Composantes Principales (ACP) sur les coefficients de Fourier elliptiques calculés à partir des contours des motifs de la classe 1, obtenus par le biais de l’approche des landmarks.

En réalisant l’ACP sur les coefficients de Fourier elliptiques, notre objectif est de construire un espace morphologique afin de représenter et caractériser les motifs de manière plus concise, tout en préservant leurs propriétés morphologiques distinctives. Ainsi, chaque motif est représenté par un ensemble de coefficients de Fourier elliptiques correspondant aux fonctions elliptiques spécifiques de leurs contours.

L’application de l’ACP nous a permis de réduire la dimensionnalité de l’ensemble de données des coefficients de Fourier elliptiques, tout en préservant les informations essentielles concernant la forme des motifs. Les principales composantes identifiées dans l’ACP sont des combinaisons linéaires des coefficients de Fourier elliptiques qui expliquent la variance présente dans nos données.

En utilisant les 2 premières composantes principales, nous avons construit un espace morphologique dans lequel chaque motif peut être représenté par ses coordonnées. Cette représentation réduite de l’information morphologique devrait permettre de comparer facilement les motifs, de détecter des similitudes et des différences significatives, ainsi que de classifier les motifs en fonction de leurs caractéristiques morphologiques. Cependant, on s’est rendu compte que les motifs ayant des contours très accidentés ou parfois étoilés n’apparaissent pas dans l’espace formé : en effet, les coefficients étaient extraits simultanément par les mêmes paramètres notamment le nombre d’harmoniques qui était limité (à 35) par les motifs de périmètre faible, en gros on avait peu de coefficients pour décrire ceux qui ont un grand contour.

Donc on a entrepris d’extraire les coefficients de fourier motifs par motifs avec une valeur appropriés pour chaque taille de contour, le motif pour lequel on a extrait le plus grand coefficient a 91 harmoniques soit $91 \times 4 = 364$ coefficients. Nous les avons tous enregistrés dans un même fichier csv, chaque ligne correspond aux coefficients d’un motif donné : la première colonne correspond au nom et numero de la pomme dont est issu le motif, puis viennent tous les coefficients dans l’ordre : A1, A2,.....,A91, B1,.....,B91, C1,.....,C91, D1,.....,D91.

ceux dont le nombre de coefficient n’atteignent pas 91 sont complétés par des valeurs nulles. Chaque colonne de coefficients joue le rôle de variable pour l’ACP.

L’ACP une fois appliquée, on se rend compte que ce sont les tous premiers coefficients (A1,B1,C1,D1) qui contribuent le plus à la construction des axes ; c’est donc eux qui permettent de discriminer le plus les individus sur les dimensions 1 et 2 comme on peut le voir sur les figures 3.42 et 3.43.

Les premiers résultats obtenus en utilisant l’Analyse en Composantes Principales (ACP) sur les coefficients de Fourier elliptique extraits des contours des motifs de colorations de la pomme rouge semblent prometteurs

pour la détection d'au moins deux groupes génétiques distincts. Ces groupes peuvent être distingués en fonction de la forme de leurs motifs : ceux dont les motifs présentent des contours accidentés et ceux ayant des contours circulaires, mais tous appartenant à la même classe de coloration.

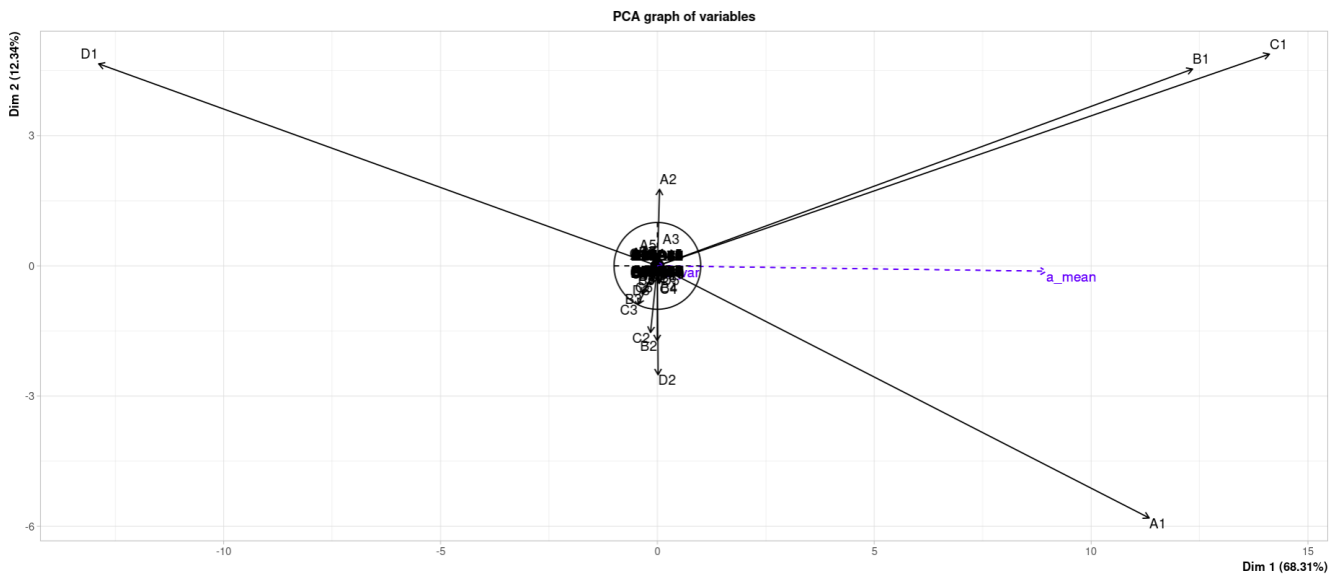


FIGURE 3.42 – Cercle de Variabilité de l'ACP des coefficients extrait des données de 2021.

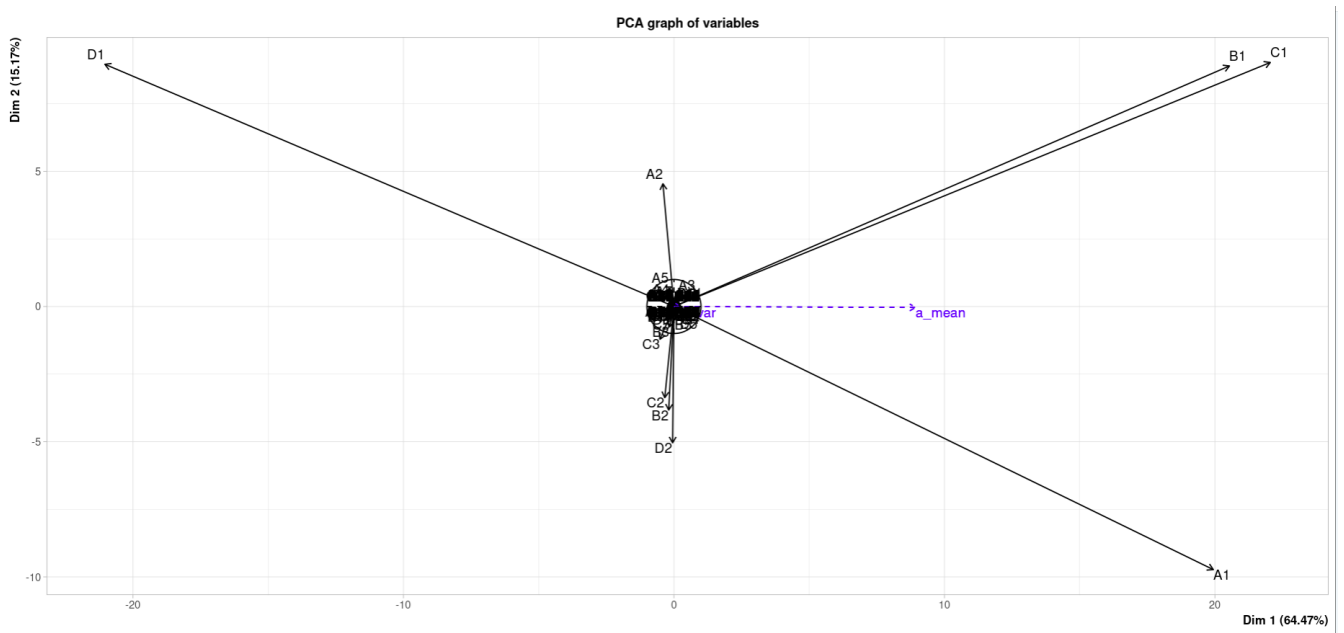


FIGURE 3.43 – Cercle de Variabilité de l'ACP des coefficients extrait des données de 2022.

3.6.3 Analyse des résultats et discussion des observations

Pour une représentation plus élégante des individus nous utilisons l'ACP du package de factomineR (de R) sur les coefficients de fourier qui fournit un outil plus élégant de visualisation des individus. Cette fois nous ajoutons les données de coloration `a_mean` (moyenne de la coloration des pommes d'un même arbre), et comme nous le montre les figures 3.44 et 3.45, on arrive à séparer à gauche et à droite de l'axe principe 2, respectivement les individus peu colorées et des individu colorées.

Ceci confirme bien que les contours sont bien liés à la coloration et que les coefficients de fourier sont efficaces dans la description des contours. De plus, les contours issus de pommes peu colorées sont généralement plus compacts, circulaires et solides, tandis que ceux des pommes colorées sont plus accidentés.

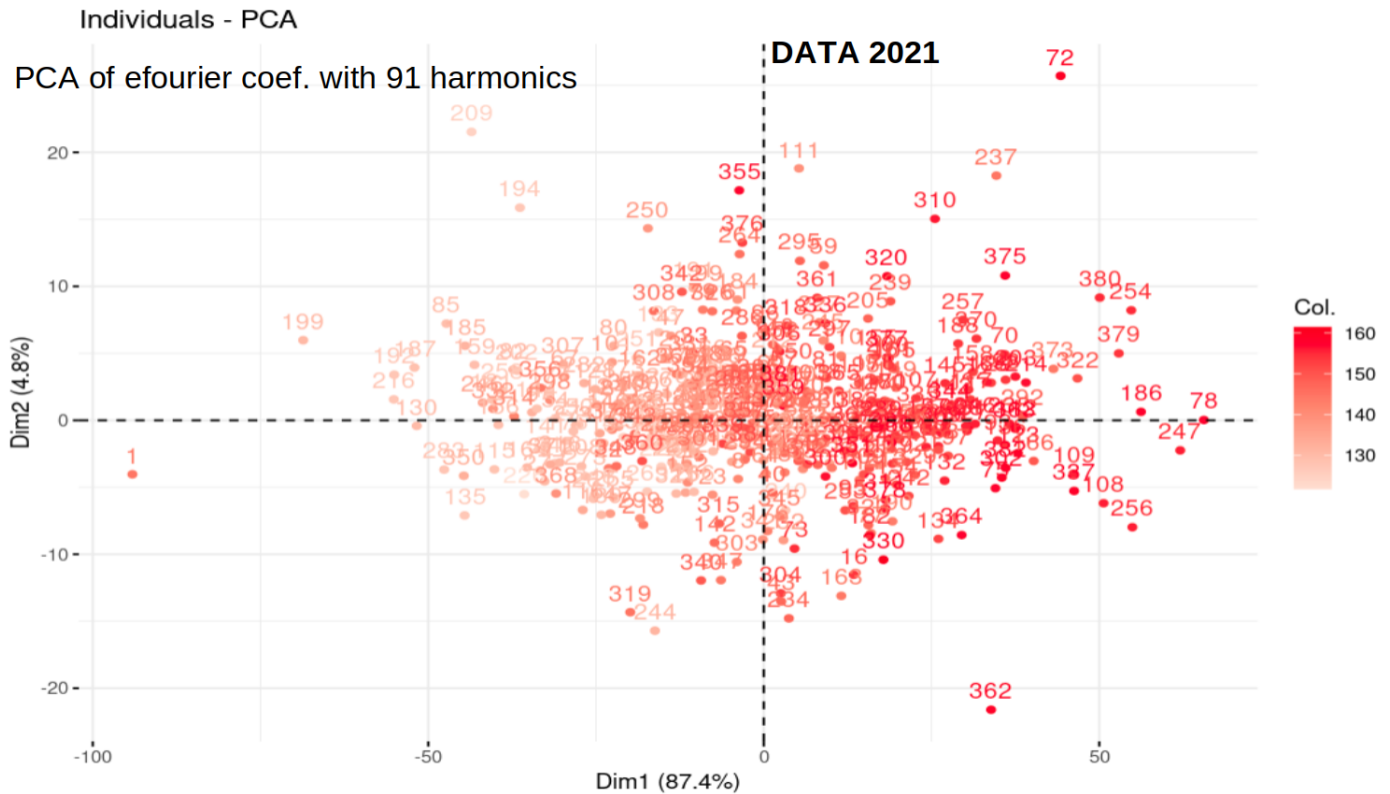


FIGURE 3.44 – Répartition des individus suivant les 2 premières dimensions obtenu après l'application de l'ACP sur les coefficients de fourier extraits des données de 2021 coloriés par `a_mean`.

Cette fois nous représentons les individus suivant les colorations `L_var` (variation de la coloration au sein des pommes d'un même arbre).

On espère observer une cohérence entre les valeurs de `L_var` et les descripteurs de fourier, ce que nous confirme bien les graphiques 3.46 et 3.47. Tout comme nous les avons présentés plus haut à l'aide du graphique 4.27, où l'on identifiait 3 groupes de motifs suivant que la variance de coloration `L_var` soit grande, intermédiaire ou faible représentée respectivement en légende par les couleur jaunes, vert et violet. On arrive à retrouver ces 3 groupes : les motifs de variance faible (points en violet) apparaissent comme attendus sur les flancs gauches et droites : à gauche ce sont ceux des pommes très peu colorées et à droite ceux très colorées.

Les motifs de variance intermédiaire (points en vert), sont ceux des pommes moyennement colorées (rouge clair) au sein desquels nous retrouvons des motifs floraux de grande étendue (de contour un peu accidenté).

Les motifs de variance élevée (points en jaune) que l'on identifie difficilement au centre sont ceux dont les motifs sont floraux et de faible étendue (de contour un peu accidenté).

Ainsi on a réussi à montrer que les coefficients de fourier sont efficaces pour décrire les contours et qu'ils sont bien liés aux valeurs colorimétriques des pommes.

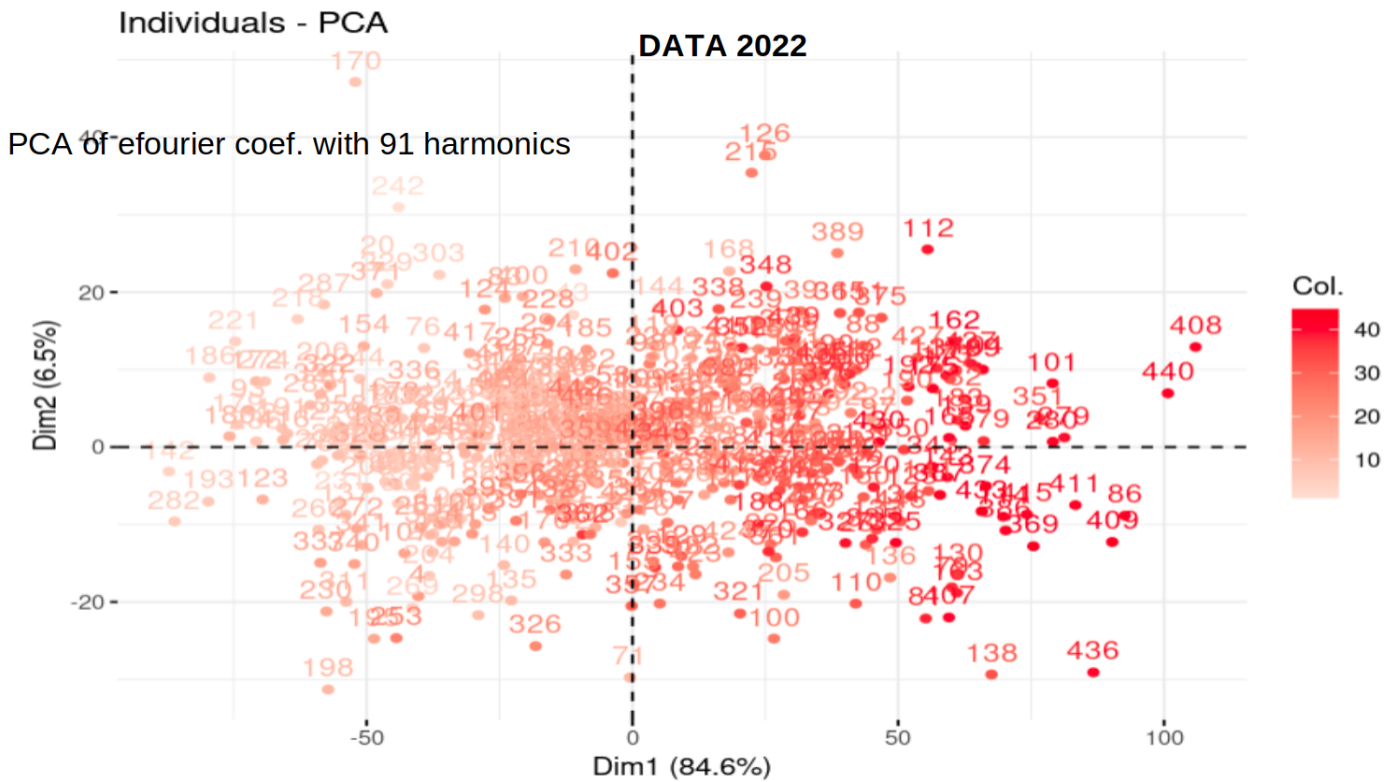


FIGURE 3.45 – Répartition des individus suivant les 2 premières dimensions obtenu après l’application de l’ACP sur les Coefficients de fourier extraits des données de 2022 coloriés par a_mean .

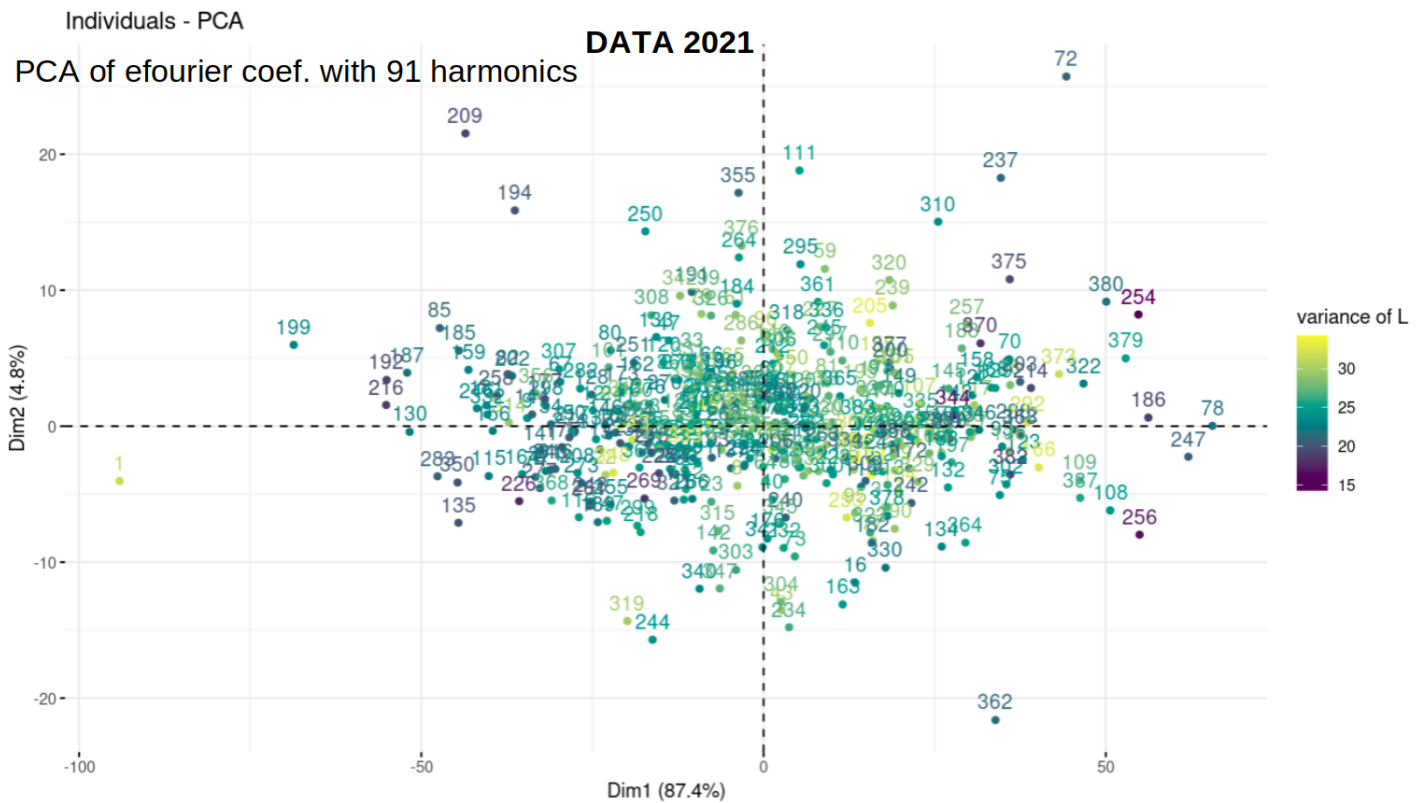


FIGURE 3.46 – Répartition des individus suivant les 2 premières dimensions obtenue après l’application de l’ACP sur les coefficients de fourier extraits des données de 2021 coloriés par L_var .

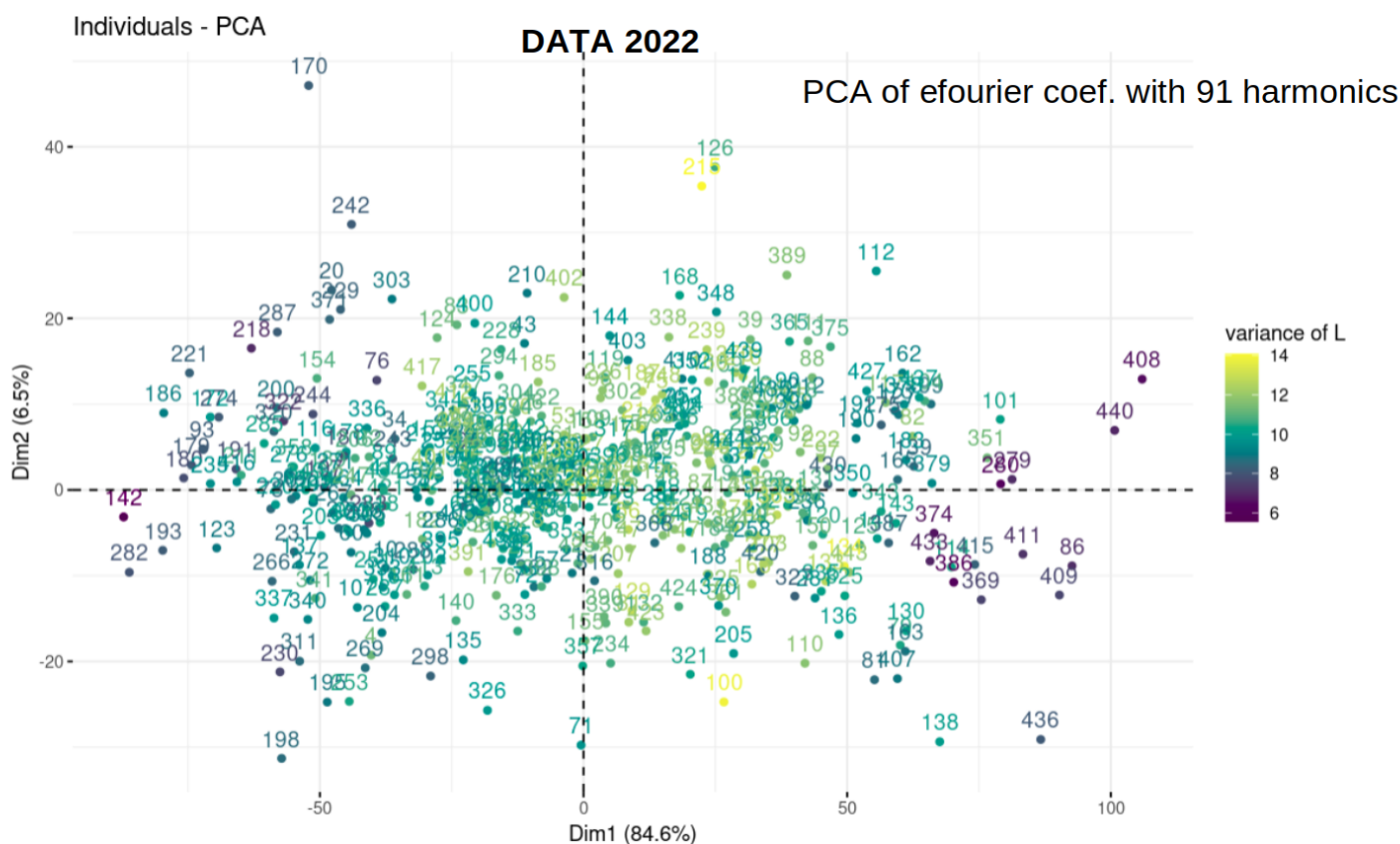


FIGURE 3.47 – Répartition des individus suivant les 2 premières dimensions obtenue après l’application de l’ACP sur les coefficients de fourrier extraits des données de 2022 coloriés par L_var.

Chapitre 4

Détection du brunissement au sein d'une pomme

Dans ce chapitre, nous abordons la problématique liée à la détection du brunissement au sein d'une pomme. Nous avons découvert un outil permettant d'extraire les fréquences caractéristiques, le poids et la fermeté d'un fruit. Par la suite, nous présenterons son fonctionnement, le protocole d'utilisation ainsi que quelques exemples de tests effectués sur deux familles de pommes. Étant donné que nous ne disposons pas d'un nombre suffisant de pommes brunies, nous illustrerons nos propos avec des exemples tirés de ces deux catégories.

4.1 Prise en main du capteur acoustique afs d'aweta

Un capteur acoustique de fermeté est un dispositif conçu pour mesurer la fermeté d'un matériau ou d'un objet en utilisant des principes acoustiques. Il utilise des ondes sonores pour évaluer la résistance ou la rigidité d'un matériau en fonction de la réponse acoustique qu'il génère.

Le capteur émet des signaux acoustiques (sons) vers l'échantillon ou l'objet à tester, puis mesure la manière dont ces signaux sont modifiés lorsqu'ils interagissent avec le matériau. Les caractéristiques de l'onde sonore réfléchie, absorbée ou transmise par le matériau fournissent des informations sur sa fermeté.



1- AFS

2- Le câble de raccordement pour l'alimentation électrique

3- 2 câbles de raccordement pour la carte son

4- 2 godets de mesure différents

5- disque d'installation du logiciel pilote

6- 2 poids de référence (100 et 200 grammes)

7- Un manuel d'utilisation

FIGURE 4.1 – Matériels de l'afs, d'aweta, Nootdorp, Pays-Bas.

Nous utilisons dans notre cas le capteur AFS (Acoustic Firmness Sensor) (voire le matériel associé à la figure 4.1) qui est conçu pour déterminer la fermeté des fruits. Il est produit par l'industrie Aweta, un leader en matière de fourniture et de fabrication de solutions clés pour le tri des fruits et légumes. Les mesures à l'aide de ce capteur commencent lorsque l'on pose le produit (fruit) dessus, il commence d'abord par déterminer le **poids** du produit ; ensuite, il crée à l'aide d'une petite masse un léger choc sur le produit et analyse le son. Le poids et le son sont combinés pour produire une valeur numérique appelée **l'indice de fermeté**. Il est à noter qu'avant toute manipulation qu'il faut retirer le dispositif de verrouillage au bas de l'AFS, puis calibrer sa balance à l'aide des masses de 100 et 200 grammes.

Dispositif expérimental

En effet dans la littérature, il est montré que les propriétés élastiques des tissus biologiques présentent un comportement vibratoire qui est utilisé comme indicateur de leurs caractéristiques internes [14].

Différentes techniques non destructives ont été développées pour observer ce comportement : le produit peut être frappé avec un petit marteau [15], tombant sur une plaque d'impact [16], soumis à un générateur ultrasonore [17] ou à un vibromètre laser Doppler [18]. La rigidité a été corrélée avec des paramètres texturaux des fruits tels que le module d'Young [19] et la fermeté des fruits [20] mesurés par des mesures pénétrométriques de référence, ou avec l'état hydrique [19].

Dans notre cas, la première technique mentionnée ci-dessus, où le fruit est positionné pour recevoir un impact avec un petit marteau, générant ainsi un signal acoustique détecté par un microphone positionné près du fruit (Figure 4.2 à l'extrême droite), a été utilisée. Le marteau est très léger et son extrémité est sphérique, de sorte que le fruit ne soit pas endommagé après l'impact.

Une analyse modale sur des produits sphériques tels que les pommes a montré que le meilleur signal est produit lorsque la réponse est enregistrée à 0 degré ou à 180° de l'endroit de l'impact [21] [22] [23].

Ce signal est enregistré, traité et transformé par le dispositif expérimental **AFS**. La transformée de Fourier rapide (FFT) du signal est effectuée sur la signature acoustique pour obtenir un spectre de puissance et identifier les **fréquences caractéristiques** [24] [25].

L'AFS dispose à son actif d'un logiciel connexe qui mesure et affiche bien les données cependant les fréquences caractéristiques, les signaux mesurés ne sont pas enregistrés, en effet un autre programme a été implémenté en Matlab pour résoudre ses insuffisances remarquées sur l'AFS.

A la fin des différentes mesures effectuées, grâce au nouveau programme on a pu générer un fichier avec les mesures de poids (gramme), de fréquence (hertz) et d'indice de fermeté des pommes recueillis lors des expérimentations ;



FIGURE 4.2 – Illustration des accessoires de l'AFS, et du positionnement du produit à mesurer.

Protocole d'utilisation de l'AFS

A/ Connexion

La 1ère chose est de **retirer le boulon qui sert de verrouillage au bas du dispositif**.

L'AFS doit être raccordé à l'ordinateur par 3 câbles (un blanc et 2 noirs). La connexion de l'AFS se fait en branchant le câble blanc sur un des ports USB (rectangulaire numéroté 1) du Pc et les 2 autres bouts aux 2 ports de l'AFS nomme **mouse keyboard**.

Il faut aussi connecter la sortie principale de la carte son du pc (prise verte (autocollant No1)) à l'entrée **In** de l'AFS avec un câble noir et connecter l'entrée **line in** de la carte son du pc (prise bleue (autocollant NO 2)) à la sortie **Out** de l'AFS avec l'autre câble noir (**SVP veuillez respecter l'ordre de branchement**).

B/ Lancement

Démarrer l'ordinateur, l'écran de l'ordinateur est sombre car le matériel pour régler la luminosité est en panne. Veuillez connecter un deuxième écran au Pc puis **fn + F5** pour changer d'écran.

Lancer MATLAB en double-cliquant sur l'icône de Matlab placé sur le bureau. Dans la zone "Current Directory" de MATLAB (en haut de la fenêtre), changer le répertoire de travail de Matlab par C:\Documents and Settings\AFS\Desktop\dossier_travail_28062011 ensuite, tapez en ligne de commande **interface_fruit** puis sur entrée. L'interface de l'AFS s'ouvre alors comme ci-dessous (fig. 4.3) avec produit : pour le numéro de mesure, le poids en grammes, f0 : la fréquence fondamentale, Fi : la fermeté du fruit ou produit.

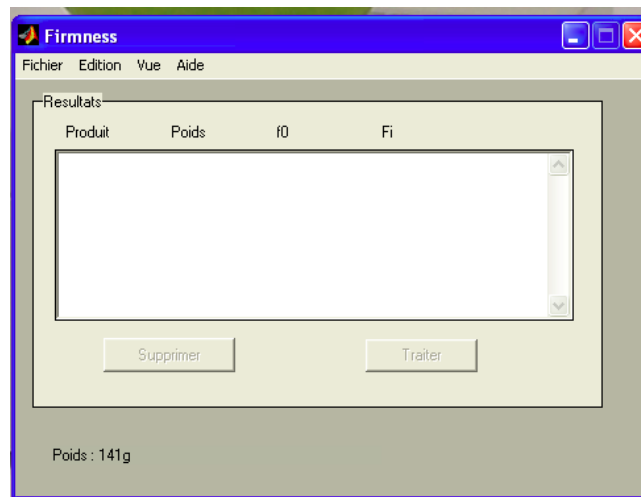


FIGURE 4.3 – Interface de l'AFS.

C/ Calibration

Avant de commencer les mesures, il est conseillé de calibrer l'appareil. Pour cela, cliquez sur le menu **Edition**, puis **Calibration**. Suivez ensuite les différentes étapes indiquées à l'écran (placer 0g puis 100g, ensuite 200g, enfin 300g = 100g + 200g).

Ensuite, vous verrez à l'écran une fenêtre titrée **Droite de calibration** comportant une droite rouge et quatre points bleus. La droite rouge passe par les points bleus lorsque la calibration est parfaite. Le programme demande de placer 200g pour tester la calibration.

Si celle-ci est réussie, alors le programme vous en informe et vous n'avez qu'à cliquer sur **OK**. Sinon, vous avez le choix entre recommencer une calibration ou annuler.

D/ Réglages

Vous devez aussi régler les paramètres de votre lot. Pour cela, cliquez sur le menu **Edition**, puis sur **Préférences**. La fenêtre suivante s'affiche (voir fig. 4.4)

L'on peut ainsi régler la force de frappe du marteau sur le fruit, le nombre de faces à mesurer sur chaque fruit, le nombre de mesures à faire sur chaque face. Cliquez sur Valider pour enregistrer vos choix.

E/ Mesures du lot

Vous pouvez maintenant commencer les mesures. Les mesures sont numérotées et s'ajoutent automatiquement dans le cadre **Résultats** lorsque le programme détecte que le poids est stabilisé à une valeur non nulle.

Il vous suffit donc de soulever puis de reposer le fruit pour déclencher la mesure suivante. Si, une mesure semble fautive, alors sélectionnez-la puis cliquez sur le bouton **Supprimer** en dessous du cadre de l'interface.

Recommencer la procédure pour ajouter une autre mesure. Si vous cliquez sur une mesure déjà faite avant de faire une nouvelle mesure et qu'il y a un trou entre la numérotation de la mesure sélectionnée et la suivante,

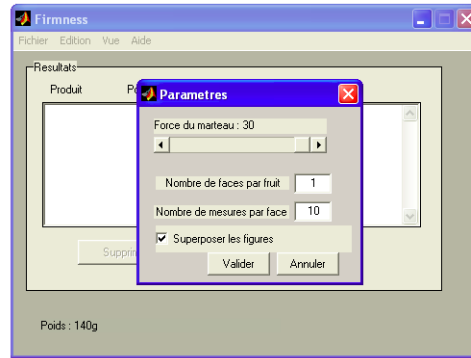


FIGURE 4.4 – Configuration des paramètres de l'AFS.

alors la nouvelle mesure est insérée juste après la mesure sélectionnée, sinon cette mesure est ajoutée à la fin de la liste.

Lorsque le lot de mesures est terminé, cliquez sur le bouton **Traiter** afin de traiter l'ensemble du lot. Une boîte de dialogue indique la progression du traitement. Le programme extrait alors les fréquences, les amplitudes, les amortissements et les déphasages de toutes les courbes en utilisant la méthode de Prony.

Vous pouvez maintenant sauvegarder le lot en cliquant sur **Sauver** dans le menu **Fichier** ; s'il n'est pas traité, il sera demandé de le faire. Toutes les valeurs seront enregistrées dans un fichier Excel lors de la sauvegarde.

Pour commencer un nouveau lot, cliquez sur **Nouveau** dans le menu **Fichier**. Le cadre **Résultats** est alors vidé. Pour quitter le programme, il suffit de faire **Quitter** dans le menu **Fichier** puis de cliquer sur **Oui**.

Il est à noter qu'à chaque mesure les courbes correspondant aux réponses impulsionnelles et fréquentielles sont affichées. La mise à jour de l'affichage des courbes se fait automatiquement lorsqu'une nouvelle mesure est faite ou lorsque l'utilisateur clique dans le cadre **Résultats** sur une mesure effectuée auparavant.

Dans le menu **Vue**, vous pouvez choisir d'afficher ou non les courbes de réponse temporelle ou fréquentielle en cliquant sur le sous-menu correspondant. Si vous voulez que toutes les courbes affichées auparavant ne soient pas effacées lorsque l'affichage d'une nouvelle courbe se fait, dans le menu **Edition** puis **Préférences**, cochez la case **Superposer les figures** (figure 4.5). Sinon, l'affichage d'une nouvelle courbe effacera les courbes précédentes.

Pour ouvrir un lot enregistré auparavant, cliquez sur **Ouvrir** dans le menu **Fichier**. Le cadre **Résultats** contient alors les mesures du lot choisi.

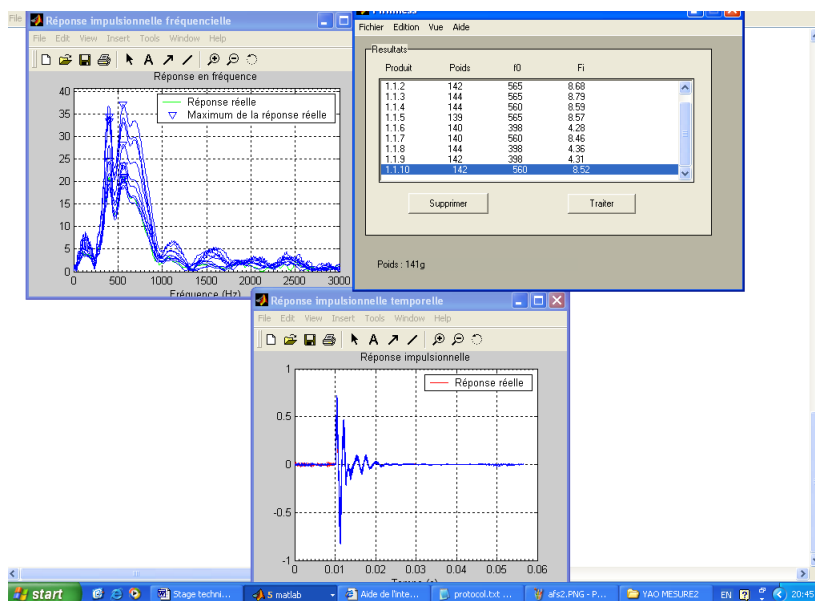


FIGURE 4.5 – Mesure prise par l'AFS.

F/ Données enregistrées

Le fichier Excel enregistré est constitué de 4 feuillets dont le premier contient les informations telles que le poids, la fréquence fondamentale F0 (hz) et la fermeté (Fi) pour chaque coup enregistré sur un fruit, ainsi que les fréquences, les amplitudes, les amortissements et les déphasages des réponses impulsionnelles fréquentielles par la méthode de Prony. Le feuillet No2 contient les moyennes des mesures : poids(g), F0(Hz) et Fi. Fi est calculé suivant la relation suivante :

$$Fi = f0^2 \times poids^{(2/3)} / 10^6$$

Les feuillets 3 et 4 quant à eux contiennent respectivement les réponses temporelles et fréquentielles du signal enregistré lors d'un coup sur le produit.

4.2 Analyse et présentation des résultats obtenus grâce à l'utilisation du matériel



FIGURE 4.6 – Pomme rouge de même famille normale.



FIGURE 4.7 – Pomme rouge de même famille pourris.

Fruit et face	Poids (g)	Fréq. (hz)	Fi (fermeté)
Pomme normale			
p1f1	103,63	478,22	5,045
p1f2	104,1	478,52	5,1
p1f3	103,83	475,23	4,99
p1f4	111,3	482,25	5,38
Mean	105.715	478.555	5.13
Pomme pourris			
p2f1	97,84	455,34	4 ,40
p2f2	98,01	454,6	4,39
p2f3	98,17	379,67	3,1
p2f4	98,32	417,36	3,72
Mean	98.085	426.7425	3.90

FIGURE 4.8 – Deux pommes rouge de même famille (1 ère mesure).

Fruit et face	Poids (g)	Fréq. (hz)	Fi (fermeté)
Pomme normale			
p1f1	86.14	310.62	1.88
p1f2	86.3	314.38	1.93
p1f3	86.39	311.15	1.89
p1f4	87.32	312	1.91
Mean	86.5	312.0375	1.9
Pomme pourris			
p2f1	62	----	-----

FIGURE 4.9 – Deux pommes rouge de même famille (2 ème mesure)

Nous réalisons une série de mesures sur deux groupes de pommes : le premier groupe est constitué de deux pommes de même famille (voir figure 4.6 et 4.7) qui viennent de la cueillette faite sur le site de l'Inrae, le second groupe constitué de 4 pommes est surnommée "pink lady" (voir figure 4.10), produit par l'association "pink lady", vient du commerce. Nous mesurons à l'aide de l'AFS la fermeté (Fi), la fréquence fondamentale et le poids.

La fermeté d'une pomme fait référence à la résistance ou à la dureté du fruit lorsqu'on le presse ou le touche (dans notre cas nous utilisons un petit marteau inclus dans l'AFS). Il s'agit d'un indicateur important de la maturité et de la qualité du fruit. Les pommes qui sont trop molles peuvent être trop mûres ou même décomposées, tandis que les pommes trop fermes peuvent être encore vertes et pas assez sucrées.

Les signaux extraits se servent de la formule de Prony. La formule de Prony est une méthode couramment utilisée pour analyser les signaux sonores et extraire des informations comme la fermeté d'un fruit ; c'est une technique mathématique utilisée pour décomposer un signal en une somme de fonctions exponentielles amorties [1] , ce qui permet de quantifier les amortissements et d'obtenir des informations sur les caractéristiques du signal, la fermeté est calculée à l'aide des coefficients amortissements.

Le coefficient d'amortissement est une mesure qui indique à quelle vitesse les oscillations d'un signal s'amortissent avec le temps. Lorsque le coefficient d'amortissement est grand, cela signifie que les oscillations du signal diminuent rapidement et le signal s'éteint rapidement. Cela peut être interprété comme un indicateur de fermeté du fruit. D'un autre côté, lorsque le coefficient d'amortissement est faible, les oscillations du signal diminuent lentement et le signal peut persister plus longtemps. Cela peut être interprété comme une fermeté plus faible du fruit. En d'autres termes, une faible valeur du coefficient d'amortissement suggère que le fruit est mou.

Cependant dans le logiciel connexe de l'AFS, la fermeté a été évaluée par la relation suivante :

$$F_i = f_0^2 \times \text{poids}^{(2/3)} / 10^6$$

Nous avons pris des mesures de toutes ces pommes sur leurs 4 faces, d'abord le premier groupe des 2 pommes, dont les mesures sont regroupées dans les tableaux mis en images aux figures 4.8 et 4.9. On effectue des prises sur deux périodes séparées de 2 semaines : on peut déjà remarquer que les résultats sur chacune des faces diffèrent légèrement : cela peut être dû à une différence de texture de la chair sur la face mesurée. En moyenne, la première pomme (pomme normale/saine) est légèrement plus lourde (environ 105.7 g) et les fréquences fondamentales avoisinent 478.5 Hz et la fermeté 5 ; et la seconde pomme (partiellement pourrie) est moins lourde et moins ferme avec un poids de 98 g, une fréquence de 426 Hz et 3.9 de fermeté. Plus le diamètre est grand plus la masse fraîche est élevée.

Deux semaines passées, nous effectuons une 2ème série de mesures, la deuxième pomme pourrie est totalement pourrie avec des pertes d'eau, et une mesure de masse fraîche de 62 g contre 98 g à la première mesure. Tandis que la première pomme est restée pratiquement saine, mais a légèrement ramoli : de 105.7 g à 86.5 g, de 478.5 Hz à 312.0375 Hz et de 3.9 à 1.9 de fermeté.

Les pommes nous les laissons à l'air ambiant de notre local de travail, ainsi plus on la laisse sans mesure de conservation la chair ainsi que la peau vont se ramollir, la quantité d'eau dans la chair diminue ;

on observe une modification de la texture avec la dégradation des structures cellulaires de la peau et une modification de la porosité résultant des processus de maturation.

cela se remarque sur les mesures avec une diminution de la fermeté, poids, et du son émis lors des mesures qui devient de moins en moins sec et la fréquence de résonance qui diminue également. On peut facilement identifier celle qui est pourrie car assez ramollie de sorte que la fermeté soit très faible de l'ordre de l'unité.

L'analyse des résultats au vu des mesures effectuées (voir figures 4.11, 4.12 et 4.13) sont les mêmes dans le deuxième lots de pomme surnommée "pink lady", cette fois nos premières mesures sont effectuées sur les pommes No 1 et 3 comme le montre la figure 4.11, (celles du No2 et 4 ayant été égarées) puis en deuxième mesure nous avons les mesures des 4 pommes.

"Pink Lady" est une variété de pomme spécifique, et bien que son profil de maturation puisse être différent de celui des autres variétés, les mêmes conclusions s'appliquent concernant les changements de texture, de fermeté et d'acoustique avec le temps :

- perte de turgescence : diminution de la pression interne des cellules végétales, ce qui provoque la diminution de sa rigidité et de sa fermeté. En fait, la fermeté des pommes diminue progressivement à mesure qu'elles mûrissent. La dégradation de la structure cellulaire entraîne une perte de rigidité dans la chair, ce qui se traduit par une baisse de la fermeté.

- les fruits mûrs ont tendance à accumuler plus d'eau, qu'ils perdent ensuite, ce qui peut agir sur certains composants et influencer la texture globale de la pomme puis la diminution du poids.

- une modification des propriétés acoustiques : la diminution de la fermeté peut affecter les propriétés acoustiques de la pomme lorsqu'elle est frappée. Cela se traduit par un son moins sec et une fréquence de résonance plus basse.

En résumé l'utilisation de l'AFS pour extraire les fréquences, le poids et la fermeté est une méthode appropriée pour détecter des changements au sein de la chair de la pomme ; les informations recueillies peuvent être de bons paramètres pour pouvoir classer en deux catégories les pommes : celles qui sont pourries et celles qui ne le sont pas. Dans notre thème il est question de détecter le brunissement, faute de pommes brunies nous avons effectué des tests sur l'AFS pour s'assurer de son bon fonctionnement dans la différenciation des pommes saines et pourries ; le temps de trouver assez de pommes pour des mesures quantitatives concernant la détection du brunissement.



FIGURE 4.10 – Quatre pommes d'une autre famille (pink lady).

Fruit et face	Poids (g)	Fréq. (hz)	Fi (fermeté)
p1f1	159,35	925,39	25,17
p1f2	159,15	918,93	24,8
p1f3	159,11	925,93	25,17
p1f4	161,52	938,85	26,14
Mean	159.7825	927.275	25.32
p3f1	150,69	923,24	24,14
p3f2	151,03	925,39	24,28
p3f3	151,03	925,39	24,28
p3f4	150,62	928,62	24,41
Mean	150.8425	925.66	24.3

FIGURE 4.11 – 1 ère mesure des pommes No1 et No3 des 4 pommes (pink lady).

Fruit et face	Poids (g)	Fréq. (hz)	Fi (fermeté)
Pomme1 face1 (p1f1)	150,69	558,79	8,84
p1f2	151,21	794,57	17,92
p1f3	151,55	557,17	8,82
p1f4	151,43	590,55	9,91
Mean	151.22	781.59	11.37
p2f1	145,04	622,85	10,71
P2f2	145,51	626,1	10,84
p2f3	145,4	629,85	10,97
p2f4	145,97	629,85	10,99
Mean	145.48	627.16	10.87

p3f1	142,13	618,54	10,42
p3f2	142,36	424,20	4,91
p3f3	131,33	409,13	4,32
p3f4	132,03	440,89	5,04
Mean	136.96	473.19	6.17
p4f1	138,06	566,32	8,57
p4f2	138,04	458,12	5,61
p4f3	138,53	387,6	4,02
p4f4	139,14	507,11	6,9
Mean	138.44	479.78	6.25

FIGURE 4.13 – No3 et No 4 des "pink lady".

FIGURE 4.12 – 2 ème mesure (2 semaines après) des quatre pommes "pink lady" (No1 et No2).

Chapitre 5

Conclusion

Le thème de ce stage s'articule sur deux problématiques, dans un premier temps l'étude de la morphologie des motifs de coloration nous a permis de développer des descripteurs dont les résultats ont été analysés à l'aide de l'ACP, un outil puissant permettant de s'affranchir du problème de multidimensionnalité, et d'observer qu'on pouvait séparer les motifs en des groupes différents suivant les données de coloration. Nous nous sommes ensuite focalisés sur les motifs de la classe 1, nous avons pu les redécrire via des landmarks ; ce qui nous permettait de s'affranchir de la masse de pixels qui décrivent nos contours puis de les caractériser par des coefficients de Fourier elliptique. Nous avons appliqué des ACP sur ces coefficients : les résultats nous ont montré que les contours sont bien liés aux données de colorations des pommes, et se regroupent bien en des groupes tels qu'espérés : des motifs issus de pommes à chair rouge ou rouge clair ou blanc. Tous les codes en python et en R sont regroupés sous github : https://github.com/letyae/Stage-descripteur-morphometrique_2023.

Tous ces regroupements suggèrent, en plus de la répétabilité que nous avons observée en comparant les données de morphométrie des pommes récoltées en 2021 et 2022, que les motifs sont bien liés à des gènes. Ces résultats bien qu'intéressants, ne nous permettent pas d'extraire les informations nécessaires pour trouver les gènes responsables des motifs de coloration ; d'autres analyses continuent d'être effectuées en laboratoire.

Dans un second temps, nous avons travaillé sur la détection du brunissement. Faute de pommes brunies, nous avons établi un protocole d'utilisation de l'AFS (Acoustic Firmness Sensor) puis nous avons analysé quelques pommes qui permet de récupérer pour un produit donné (fruits) : la fermeté, la fréquence fondamentale, et le poids du fruit en question. Avec quelques pommes de type différent (variété et stade de sénescence) nous avons vu qu'il était possible de différencier avec ces mesures une pomme pourrie d'une pomme saine ; les pommes pourries subissant une dégradation des structures cellulaires contenant l'eau du fruit, ce qui agit (en la diminuant) sur la fréquence de résonance obtenue en générant du son sur le fruit, avec comme conséquence une diminution de fermeté. La masse fraîche du fruit est également diminuée au cours de la conservation à température ambiante. Des travaux futurs permettront de confirmer ces constatations sur des pommes brunies.

5.1 Annexe

Protocole d'acquisition

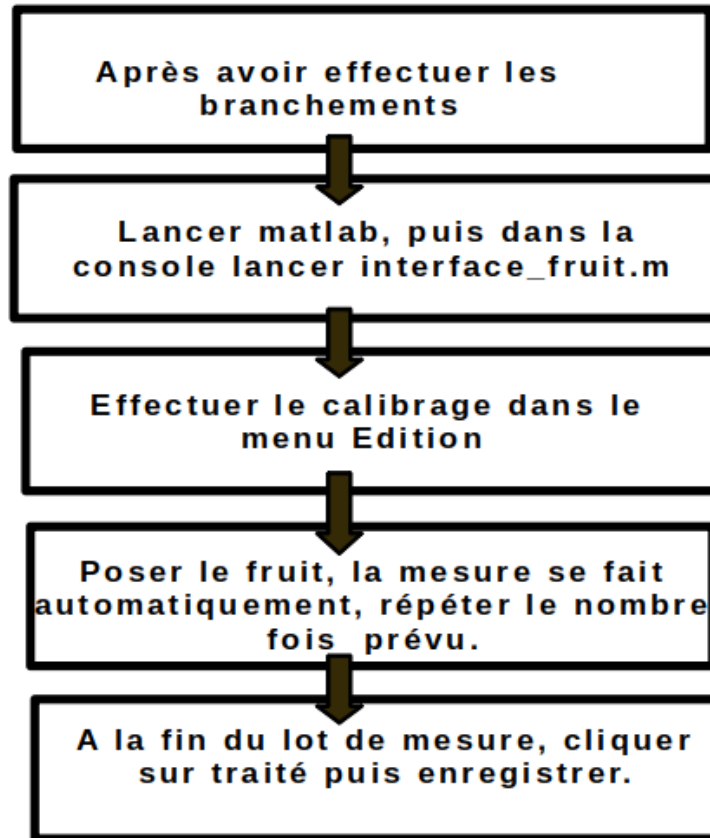


FIGURE 5.1 – Protocole d'acquisition avec l'AFS, aweta

Bibliographie

- [1] Diagnostic et Signal. Détection de chocs en tête de suspente d'un siège de télésiège par la méthode de Prony temps-fréquence. Journal sur l'enseignement des sciences et technologies de l'information et des systèmes, Nadine Martin., 2005, 18 (2), 18 p. hal
- [2] Morphometric analysis of Passiflora leaves : the relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade, Chitwood, Daniel H and Otoni, Wagner C, GigaScience 2017, Oxford University Press
- [3] Analyse de la variabilité morphologique chez huit populations spontanées de Pistacia atlantica en Algérie, Belhadj, Safia and Derridj, Arezki and Auda, Yves and Gers, Charles and Gauquelin, Thierry , Botany , 2008
- [4] Michael A. Wirth, Ph.D. University of Guelph Computing and Information Science Image Processing Group, 2004, page 9
- [5] Shape Descriptor/Feature Extraction Techniques Fred Park UCI iCAMP 2011, Circularity Ratio page 18.
- [6] D. Zhang and G. Lu, "A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval," Visual Communication and Image Representation, vol. 14(1), 2003.
- [7] Le nombre d'Euler Christian RONSE (02/11/2007) LSIIT UMR 7005 CNRS-ULP, Département d'Informatique de l'ULP
- [8] Principal Component Analysis, web article : <http://www.sthda.com/english/wiki/fviz-pca-quick-principal-component-analysis-data-visualization-r-software-and-data-mining>
- [9] Principal Component Analysis, web article : <https://datascientest.com/acp>
- [10] FICHE MÉTHODOLOGIQUE, Série de Fiches Méthodologiques en Recherche et en Rédaction Scientifique Fiche Méthodologique n : 3/2021, Analyse en Composantes Principales (ACP) Principal Component Analysis (PCA), Kamel Ben Salem1, Ahmed Ben Abdelaziz2, Réseau Maghrébin PRP2S*, * Réseau- Maghrébin : Pédagogie- Recherche- Publication en Sciences de Santé (PRP2S).
- [11] Automatic fruit morphology phenome and genetic analysis : An application in the octoploid strawberry, Zingaretti, Laura M and Monfort, Amparo and Pérez-Enciso, Miguel, Plant Phenomics, 2021, AAAS
- [12] Cierra Andaur, using kmeans clustering for image segmentation, web article : <https://cierra-andaur.medium.com/using-k-means-clustering-for-image-segmentation-fe86c3b39bf4>
- [13] Firmness Measurement of Stored Delicious' Apples by Sensory Methods, Magness-Taylor, and Sonic Transmission , Abbott, Judith A and Affeldt, Henry A and Liljedahl, Louis A, Journal of the American Society for Horticultural Science, American Society for Horticultural Science, pages 590-595, 1992
- [14] Clark, H.L. and W. Mikelson. 1942. Fruit ripeness tester. U.S. Patent 2277037.
- [15] Duprat, F., Grotte, M., Piétri, E., Loonis, D. and Studman, C. J., The Acoustic Impulse Response Method for Measuring the Overall Firmness of Fruit, Journal of Agricultural Engineering Research, 66(4), 251-259, 1997
- [16] Pearson, T. C., Cetin, A. E., Tewfik, A. H. and Haff, R. P., Feasibility of impact-acoustic emissions for detection of damaged wheat kernels, Digital Signal Processing, 17(3), 617-633, 2007.
- [17] Bechar, A., Mizrach, A., Barreiro, P. and Landahl, S., Determination of Mealiness in Apples using Ultrasonic Measurements, Biosystems Engineering, 91(3), 329-334, 2005
- [18] Taniwaki, M., Takahashi, M. and Sakurai, N., Determination of optimum ripeness for edibility of postharvest melons using nondestructive vibration, Food Research International, 42(1), 137-141, 2009.
- [19] Herppich, W. B., Landahl, S., Herold, B. and De Baerdemaeker, J., Interactive effects of water status and produce texture - An evaluation of non-destructive methods, In Proceedings of the International Conference : Postharvest Unlimited, Leuven, Belgium, 2003.
- [20] De Belie, N., Schotte, S., Lammertyn, J., Nicolai, B. and De Baerdemaeker, J., Firmness Changes of Pear Fruit before and after Harvest with the Acoustic Impulse Response Technique, Journal of Agricultural Engineering Research, 77(2), 183-191, 2000

- [21] Chen, H., Analysis on the acoustic impulse resonance of apples for non-destructive estimation of fruit quality, PhD diss. Leuven, Belgium : Katholieke University, 1993
- [22] Huarng, L., Chen, P. and Upadhyaya, S., Determination of acoustic vibration modes in apples, Transactions of the American Society of Agricultural Engineers, 36(5), 1423-1429, 1993.
- [23] Shmulevich, I., Galili, N. and Howarth, M. S., Nondestructive dynamic testing of apples for firmness evaluation, Postharvest Biology and Technology, 29(3), 287-299, 2003.
- [24] Tesch, R., Normand, M. D. and Peleg, M., Comparison of the acoustic and mechanical signatures of two cellular crunchy cereal foods at various water activity levels, Journal of the Science of Food and Agriculture, 70(3), 347-354, 1996.
- [25] Valente, M., Leardi, R., Self, G., Luciano, G. and Pain, J. P., Multivariate calibration of mango firmness using visible/NIR spectroscopy and acoustic impulse method, Journal of Food Engineering, 94(1), 7-13, 2009.
- [26] Momocs : outline analysis using R, Bonhomme, Vincent and Picq, Sandrine and Gaucherel, Cédric and Claude, Julien, Journal of Statistical Software, volume=56, pages=1-24, 2014
- [27] Linking Genes to Shape in Plants Using Morphometrics Hao Xu and George W. Bassel School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom ; email : george.bassel@warwick.ac.uk
- [28] <https://fr.wikipedia.org/wiki/Morphometrie>.