HAL
open science

Mise en place et développement d'un outil de phénotypage assisté par vision artificielle pour la sélection du chou-fleur

Angéline Bellouard

To cite this version:

Angéline Bellouard. Mise en place et développement d'un outil de phénotypage assisté par vision artificielle pour la sélection du chou-fleur. Sciences agricoles. 2014. dumas-01082246

HAL Id: dumas-01082246

https://dumas.ccsd.cnrs.fr/dumas-01082246v1

Submitted on 26 Nov 2014

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

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

AGROCAMPUS OUEST

☑ **CFR Angers**

☐ **CFR Rennes**

Année universitaire : 2013 - 2014

Spécialité :

Horticulture

Spécialisation (et option éventuelle) :

Amélioration des plantes

Mémoire de Fin d'Études

☑ d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage

☐ de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage

☐ d'un autre établissement (étudiant arrivé en M2)

Mise en place et développement d'un outil de phénotypage assisté par vision artificielle pour la sélection du chou-fleur

Par : Angéline BELLOUARD

Soutenu à Rennes, le 10 Septembre 2014

Devant le jury composé de :

Président : Maria Manzanares-Dauleux

Maître de stage : Cindy Torres

Enseignant référent : Maria Manzanares-Dauleux

Autres membres du jury :

Hervé Nicolas (Rapporteur)

Marie Hélène Macherel (Enseignant référent)

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Fiche de confidentialité et de diffusion du mémoire

Confidentialité :

☐ Non ☑ Oui si oui : ☐ 1 an ☑ 5 ans ☐ 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible(1).
A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant).
Date et signature du maître de stage(2) :

Droits d'auteur :

L'auteur(3) autorise la diffusion de son travail

☑ Oui ☐ Non

Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☑ la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

Date et signature de l'auteur :

Autorisation de diffusion par le responsable de spécialisation ou son représentant :

L'enseignant juge le mémoire de qualité suffisante pour être diffusé

☐ Oui ☐ Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.
Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☐ la diffusion papier et électronique du mémoire

Date et signature de l'enseignant :

(1) L'administration, les enseignants et les différents services de documentation d'AGROCAMPUS OUEST s'engagent à respecter cette confidentialité.
(2) Signature et cachet de l'organisme
(3).Auteur = étudiant qui réalise son mémoire de fin d'études
(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé

Remerciements

Après deux belles années d'apprentissage passées au pôle Vision Artificielle de Vilmorin, je tiens à exprimer avec sincérité toute ma reconnaissance à mes encadrants, Pierre Ferraton et Cindy Torres, qui m'ont permis de vivre cette aventure professionnelle et humaine.

Je tiens tout particulièrement à remercier ma maître d'apprentissage, Cindy Torres, pour sa pédagogie, sa joie de vivre, sa franchise et sa gentillesse. Merci pour ta patience, tes encouragements, tes justes critiques et ta reconnaissance. Merci également de m'avoir fait découvrir et aimer ton travail, le traitement d'images. Les moments passés ensemble ont été riches et c'est avec beaucoup d'émotions et de souvenirs que je quitte ce service.

Je tiens également à remercier Ali Boudjedra, pour ces conseils avisés et sa présence masculine. Les moments passés au pôle Vision Articielle ont été de bons moments de plaisanteries et de complicité. Je remercie aussi également ceux et celles qui ont partagés ce grand bureau et qui ont dû me supporter, je pense biensur à Franck, Marine et Rokaya.

Je tiens également à remercier Armand, pour avoir cru en l'imagerie et sans qui, l'activité du pôle Vision Artificielle ne serait sans doute pas autant développée.

Merci à toute l'équipe R&D Technosemences, chef de projets et techniciens, pour ces bons moments partagés à la pause-café, souvent agréablement accompagnés de confiseries et de patisseries. Les moments partagés avec vous en entreprise et en dehors ont été très agréables. Vive la Yo Brasil Team ! Même si nous n'avons pas marqué d'essai, l'essentiel c'est d'avoir participé.

Je remercie également les équipes de la recherche avec lesquelles j'ai échangé avec intérêt sur la sélection et le phénotypage du chou-fleur.

Enfin, je remercie Marie Hélène Macherel d'avoir acceptée d'être ma tutrice et de m'avoir suivie durant ces deux années.

AGRO CAMPUS OUEST	Diplôme : Ingénieur de l'Institut Supérieur des Sciences Agronomiques, Agroalimentaires, horticoles et du Paysage Spécialité : Horticulture Spécialisation / option : Sciences des Productions Végétales/ Amélioration des Plantes Enseignant référent : Maria Manzanares-Dauleux

Auteur(s) : Angéline BELLOUARD Date de naissance : 07/04/1989	Organisme d'accueil : Vilmorin S.A. Adresse : Route du Manoir, 49250 La Ménitré
Nb pages : 36 Annexe(s) : 9	
Année de soutenance : 2014	Maître de stage : Cindy Torres

Titre français : Mise en place et développement d'un outil de phénotypage assisté par vision artificielle pour la sélection du chou-fleur.
Titre anglais : Building and development of a phenotyping tool assisted by artificial vision for the cauliflower breeding.

Résumé: Malgré un essor important depuis ces dernières années de machines industrielles visant à caractériser les fruits et légumes, il n'existe pas à ce jour de machines qui puissent phénotyper un grand nombre de caractères morphologiques et/ou colorimétriques sur le chou-fleur. Pour ces raisons et dans le but de développer des marqueurs moléculaires, un projet Vilmorin et Compagnie (VCO) de phénotypage assisté par imagerie a vu le jour en 2012 pour l'étude de 14 caractères sur le chou-fleur. A partir d'un système d'acquisition VCO développé par l'équipe Vision Artificielle chez Vilmorin, des bases de données et d'images ont d'abord été construites. Puis à partir de ces bases d'informations, des méthodes d'analyse d'images et des algorithmes ont été développées et validés statistiquement. La méthodologie pour la validation des méthodes d'analyse d'images et des algorithmes ainsi que les résultats sont illustrés pour quatre caractères morphologiques : le diamètre et le caractère pointu pour les caractères en deux dimensions, le volume et le caractère lisse pour les caractères en trois dimensions. Le diamètre et le volume sont prédits par des modèles de régressions linéaires dont les coefficients de détermination sont respectivement de 98% et 97%. Pour le caractère pointu et le caractère lisse, une méthode de classification par seuillage est utilisée pour classer respectivement en deux et trois catégories des pommes de chou-fleur, avec un pourcentage d'erreur respectif de 15% et 34%. Cet outil a un faible débit mais c'est un instrument dont les mesures sont objectives, précises et fiables (répétables et reproductibles).

Abstract: Despite an increasing number of industrial machines for the characterization of fruits and vegetables, there is no machine today which allows a characterization of an important number of morphological or colorimetric traits of the cauliflower. For these reasons and in order to develop molecular markers, a Vilmorin & Cie (VCO) project assisted by imagery was built in 2012 for the study of 14 traits of the cauliflower. Thanks to a VCO acquisition system, databases and images bases were firstly built. Then, with these bases, image analysis methods and algorithms were developed and statistically validated. Validation methods for the image analysis method and algorithm but also results are illustrated with four morphological traits: diameter and pointed trait for the two dimensional characters, volume and smoothness trait for the three-dimensional characters. Diameter and volume are predicted with regression linear model with a determination coefficient of respectively 98% and 97%. For the pointed trait and the smoothness trait, a classification method by threshold is used to class respectively in two and three categories cauliflower curds, with an error rate of 15% and 34%. This tool has a low throughput but measurement is objective, accurate and reliable (repeatable and reproducible).

Mots-clés : chou-fleur, analyse d'images, outil de phénotypage, caractères morphologiques et colorimétriques, diamètre, volume, caractère lisse, caractère pointu.
Key Words: cauliflower, image analysis, phenotyping tool, colorimetric and morphological traits, diameter, volume, smoothness trait, pointed trait.

Table des matières

Introduction 1

I. Matériel végétal 8

A. Caractéristiques générales 8

B. Les caractères étudiés, notés et mesurés sur le chou-fleur 8

C. Préparation du chou-fleur aux prises de vues 10

II. L'outil de phénotypage par imagerie 10

A. Le système d'acquisition 10

B. Guide d'utilisation 10

C. Les bases de données et d'images 11

D. Les logiciels 11

E. Analyses statistiques 12

III. Le contrôle des sources de variation du phénotypage 13

A. La variabilité inter-notateur 13

B. La relativité de l'évaluation phénotypique 13

IV. Une analyse fine des caractéristiques du chou-fleur 14

A. Exemples de caractères morphologiques en deux dimensions 14

B. Exemples de caractères morphologiques en trois dimensions 16

V. Les sources de variabilité 19

A. L'étude de la variabilité inter-experts 19

B. Le matériel végétal et la variabilité des caractères 20

C. Les réglages de la caméra et l'utilisation de chartes couleurs 21

VI. L'outil en tant qu'instrument de mesure 21

A. Le coût et le débit de l'outil 21

B. Les mesures par imagerie 22

C. La précision de l'évaluation 22

D. Quelques préconisations d'emploi de l'outil à des fins d'analyse génétique 22

VII. L'avenir de tels outils pour le phénotypage des espèces potagères 23

A. Utilisation de technologies encore plus innovante 23

B. Les outils de phénotypage à haut débit, des outils d'avenir 23

Conclusion 24

Bibliographie 25

Sitographie 27

Liste des annexes

Annexe I : Les notations sur images et les bases d'apprentissage pour la forme et les caractères pointu et lisse 28

Annexe II : Description du colorimètre utilisé dans le cadre du projet 29

Annexe III : Formules des descripteurs de forme sous ImageJ 30

Annexe IV : Ajustement de l'ouverture et du focus de la caméra 31

Annexe V : Vérification des conditions d'applications des modèles statistiques 32

Annexe VI : Description des critères d'évaluation de la qualité des modèles de prédiction 33

Annexe VII : Vérification des conditions d'applications des modèles statistiques pour les différents caractères traités dans la partie Résultats 34

Annexe VIII : Formules de volume testées 35

Annexe IX : Matrice de corrélation des variables testées pour l'évaluation du volume 36

Glossaire

Additivité : « en génétique quantitative, du point de vue des effets géniques à un locus, il y a additivité stricte lorsque la valeur de l'hétérozygote est égale à la moyenne de deux homozygotes ».

Allèles : « versions alternatives d'un même gène différent par leur séquence nucléotidique. Par extension désignent les variants de l'ADN non codant en un locus. »

Analyse d'images : « Travail qui consiste à mesurer des éléments sur les images qui sont le plus souvent déjà traitées. »

Dominant (récessif): « se dit d'un allèle ou d'une mutation qui à l'état hétérozygote, (ne) conditionne (pas) le phénotype. »

Exactitude : « qualifie un instrument qui donne des indications individuelles très voisines de la valeur vraie de la grandeur à mesurer. »

Fiabilité : « aptitude d'un instrument de mesure à accomplir uen fonction requise, dans des conditions données, pendant une durée donnée", ensemble de qualités que sont la fidélité, la sensibilité, la justesse et l'exactitude. »

Fidélité : « aptitude d'un instrument de mesure à donner, dans des conditions d'utilisation définies, des réponses très voisines lors de l'application répétée d'un même signal d'entrée. »

Fleurettes : parties supérieures de l'inflorescence hypertrophiée du chou-fleur.

Génétique d'association : « travail qui consiste à identifier des polymorphismes d'une variation qualitative / quantitative d'un caractère d'intérêts dans une population de large diversité. »

Héritabilité : « au sens large, part de la variance génétique dans la variance phénotypique, au sens étroit ou stricte, part de la variance d'additivité dans la variance phénotypique. »

Hétérosis : « phénomène de supériorité du produit d'un croisement par rapport à ses parents. »

Homozygote/Hétérozygote : « état d'un génotype diploïde avec un seul allèle à un locus (AA/aa) / état d'un génotype diploïde avec deux allèles à un locus (Aa). »

Hybride: « descendance d'un croisement contrôlé entre deux constituants (parents) qui peuvent être de nature différente (lignée, clone ou populations). L'hybride F1 ou hybride simple constitue la première génération (F1) d'un croisement de deux lignées homozygotes (hybride). »

Justesse : « qualifie un instrument pour lequel la moyenne des indications données est très voisine de la valeur vraie de la grandeur à mesurer. »

Locus : « position d'un gène sur le génome. «

Marqueur moléculaire (d'ADN) : « séquence codante ou non de l'ADN qui présente un polymorphisme. »

Niveau de gris: est d'une image RVB convertie en 8-bit (0-255) ou les valeurs des pixels sont comprises entre 0 (noir) et 255 (blanc).

Pixel : picture element en anglais, plus petit élément d'une image numérique.

Psychophysique : « étude des relations entre les stimuli et les réponses correspondantes »

QTL(Quantitative Trait locus): « locus déterminant un caractère quantitatif »

Répétabilité : « on parle de répétabilité lorsque les répétitions de la mesure sont faites dans les mêmes conditions. »

Reproductibilité : « on parle de reproductibilité lorsque lorsqu'on fait varier d'une répétition à l'autre certaines conditions telles que le lieu, l'expérimentateur, le moment de l'expérience, l'instrument de mesure lui-même. »

Segmentation: « Opération qui consiste à extraire d'une image des primitives géométriques. Les primitives les plus employées sont les segments (contours) ou les surfaces (régions). »

Sensibilité : « plus petit écart qu'un appareil de mesure est capable de détecter entre deux valeurs d'une grandeur »

SNP (Single Nucléotide Polymorphism): variation de séquence d'ADN due au changement d'un seul nucléotide.

Traitement d'images : « Processus qui permet de rendre apparent ou de cacher des éléments dans l'image. »

Variabilité interindividuelle : « due aux différences de réponses d'un sujet à l'autre. »

Variabilité intraindividuelle : « due aux différences d'une réponse à l'autre d'un même sujet. »

Variable ordinale : « variable dont les catégories se voient attribuer d'un rang. »

Abbréviations

2D: deux dimensions

3D: trois dimensions

ADN: Acide DésoxyriboNucléique

AIC: Akaike Information Criteria

ANOVA: ANalyse Of Variance

BIC: Bayesian Information Criteria

BU: Business Units

CA: Chiffre d'Affaires

CCD: Charge-Coupled Device

CIE: Commission Internationale de l'Eclairage

CMOS: Complementary Metal Oxide Semiconductor

CMS: Cytoplasmic Male Sterility

CTIFL: Centre Technique Interprofessionnel des Fruits et Légumes

etc.: et caetera

ICAB: Image Capture Box

INRA: Institut National de Recherche Agronomique

LEDs: diodes électroluminescentes

PRF: Product Request Form

Q-Q: Quantile-Quantile

QTL: Quantitative Trait Loci

R^2: Coefficient de détermination

R&D: Recherche et Développement

RGB: Red Green Blue (équivalent de RVB en français)

ROI: Region Of Interest

RVB: Rouge Vert Bleu

S.A.: Société Anonyme

SNP: Single Nucleotid Polymorphism

SSH: Segmentation de l'histogramme

TIFF: Tag Image File Format

VACS: Visar's Automatic Carrot Sorter

VCO: Vilmorin et COmpagnie

Liste des illustrations

❖ **Figures**

Figure 1 : Organisation du groupe Limagrain et place de l'entreprise Vilmorin SA dans la filiale Vilmorin et Compagnie (source Vilmorin)..... Avant-Propos

Figure 2 : Répartition de l'activité Vilmorin SA pour l'exercice 2010 – 2011 (source Vilmorin). ..... Avant-Propos

Figure 3 : Réglette étalon pour le calibre des gousses de haricot (source Vilmorin). .... Avant-Propos

Figure 4 : Evaluation du calibre, de la courbure et de la longueur d'une gousse de haricot (source Vilmorin). ..... Avant-Propos

Figure 5 : Composition moyenne d'une portion de 100 g de chou-fleur (Aprifel). 1

Figure 6 : Répartition des volumes produits par région en 2012 (Agreste)..... 1

Figure 7 : Volumes de choux-fleurs (frais et transformé) en France de 2007 à 2012 (Agreste). 1

Figure 8 : Répartition du chiffre d'affaires en millions d'euros des semences de chou-fleur par entreprise semencière pour l'année 2011-2012 (E. Bizais, 2013)..... 2

Figure 9 : Répartition des ventes 2011-2012 de semences potagères Vilmorin par espèce et par zone géographique (Vilmorin, 2013)..... 2

Figure 10 : Segmentation du marché du chou-fleur (CF) en trois catégories, les CF Tropicaux, les CF à Cycles courts et les CF à cycles longs, avec leurs surfaces de production respectives et leur valeur dans le marché (E. Bizais, 2013)..... 2

Figure 11 : Etapes de développement variétal depuis la définition d'un produit type jusqu'à la commercialisation de la semence créée. 3

Figure 12 : Résistances aux maladies recherchées par Vilmorin : la pourriture noire, le mildiou, la hernie des crucifères et la tâche annulaire (K. S. Kapoor, 1999; Koike et al., 2007; Pacific Northwest Plant Disease Handbook, 2014) 3

Figure 13 : Niveaux auxquels peut être évalué le phénotype d'une plante : biologique, structural, temporel et environnemental (C. Alonso-Blanco and B. Méndez-Vigo, 2014). 4

Figure 14 : Code couleur des inflorescences de chou-fleur (CTIFL, 1991). CTIFL : Centre Technique Interprofessionnel des Fruits et Légumes..... 5

Figure 15 : Différents niveaux de traitement d'images (Zhang et al., 2014). 5

Figure 16 : Principaux composants d'un système de vision assisté par ordinateur (R. B. Pandit et al., 2007) 5

Figure 17 : Chaine automatisée par vision artificielle du tri des oranges (N. Kondo, 2010).... 6

Figure 18 : Etat des choux fleurs, coupes réalisées et photographies associées (Vilmorin)..10

Figure 19 : Image Capture Box (Vilmorin).....10

Figure 20 : Rampes lumineuses de l'ICAB (Vilmorin) 10

Figure 21 : Présentation des bases de données et d'images (Vilmorin).....11

Figure 22 : Représentation des bases d'apprentissages et de test nécessaire au développement des programmes informatiques d'analyse d'images (Z. A. Desta and R. Ortiz, 2014). 11

Figure 23 : Les quatre images de pré-traitement (Vilmorin). 11

Figure 24 : Analyses statistiques pour les deux types de validations (méthodes d'analyse d'images, algorithmes) 12

Figure 25 : Taux de discordance (exprimé en pourcentage) des notes entre experts pris 2 à 2, entre les notes de tous les experts ou entre les notes d'experts et les notes au champ, pour la forme, le caractère lisse et le caractère pointu des pommes de chou-fleur. 13

Figure 26 : Répartition des notes pour l'évaluation de la lissitude par expert. 13

Figure 27 : Schéma illustrant les différents points de contrôles des sources de variation liées au phénotypage. 14

Figure 28 : Modèle statistique de régression linéaire pour la validation de la méthode d'analyse d'images. 15

Figure 29 : Modèle statistique de régression linéaire pour la validation de l'algorithme 15

Figure 30 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 150 pixels en fonction des valeurs du R^2(90-93) et de l'angle (131-136). 16

Figure 31 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 200 pixels en fonction des valeurs du R^2(90-95) et de l'angle (131-136). 16

Figure 32 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 250 pixels en fonction des valeurs du R^2(90-97) et de l'angle (131-136). 16

Figure 33 : Comparaison des coefficients de détermination des modèles de régression linéaires simples correspondant aux différentes formules de volume. 17

Figure 34 : Comparaison des coefficients de détermination des différents modèles de régression linéaire multiple pour l'évaluation du volume. 17

Figure 35 : Modèle de régression linéaire pour l'évaluation du volume par l'algorithme. 17

Figure 36 : Sélection de la partie supérieure d'un chou-fleur en coupe longitudinale, duplication de cette sélection (A) et construction d'une forme ronde par l'assemblage des deux parties (rotation de 180° de l'une des deux parties, puis accolement des deux parties au niveau du segment de droite, B). 18

Figure 37 : Coefficient de détermination des modèles d'analyse de variance à un facteur avec en bleu les variables expliquées relatives aux mesures des images de pommes de chou-fleur observées vue du dessus, en jaune les variables expliquées relatives aux mesures des images de pommes de chou-fleur observées en coupe longitudinale (face externe) et en orange, la moyenne de la variable mesurée sur les deux vues. 18

Figure 38 : Taux d'erreur en fonction de la valeur de seuil pour la limite A-B (le taux d'erreur minimal est encadré en rouge) 18

Figure 39 : Taux d'erreur en fonction de la valeur de seuil pour la limite B-C (le taux d'erreur minimal est encadré en rouge) 18

Figure 40 : Images de chou-fleur utilisées pour la notation de la lissitude, la forme et le caractère pointu de la pomme de chou-fleur (source Vilmorin). 28

Figure 41 : Répartition des pommes prélevées en fonction des notations au champ de la forme et du caractère pointu. 28

Figure 42 : Répartition des pommes prélevées en fonction des notations au champ de la lissitude. 28

Figure 43 : Colorimétre « CR-400 » Konica Minolta Sensing, INC (Vilmorin).....29

Figure 44 : Espace couleur L*a*b* ou CIELAB 29

Figure 45 : Bagues de réglage de la camera : le focus et l'ouverture.....31

Figure 46 : Image de QPcard avec un bon ajustement de l'ouverture.....31

Figure 47: Ajustement correct et incorrect de la netteté de l'image (respectivement à droite et à gauche) 31

Figure 48 : Tests statistiques et analyses graphiques pour la vérification des conditions d'applications des modèles statistiques. 32

- **Tables**

Table 1 : Liste et descriptif des caractères mesurés par imagerie 8

Table 2 : Descriptif des mesures (physiques, visuelles, sur images) menées pour la validation terrain. 9

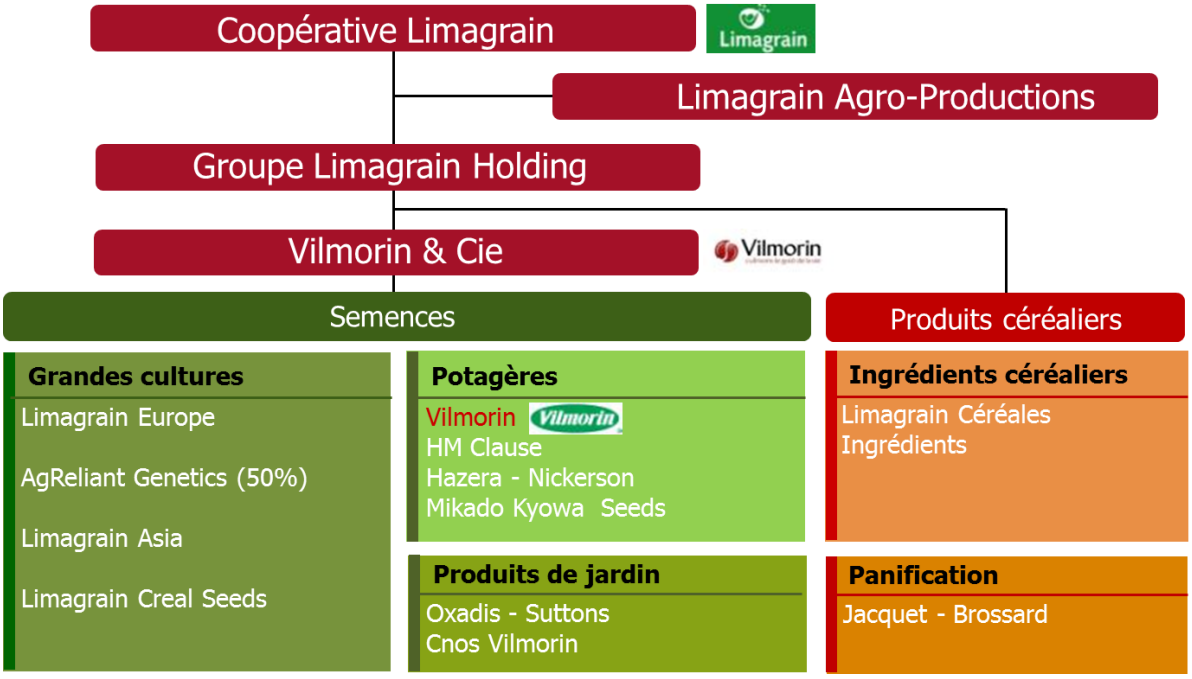


Figure 1 : Organisation du groupe Limagrain et place de l'entreprise Vilmorin SA dans la filiale Vilmorin et Compagnie (source Vilmorin).

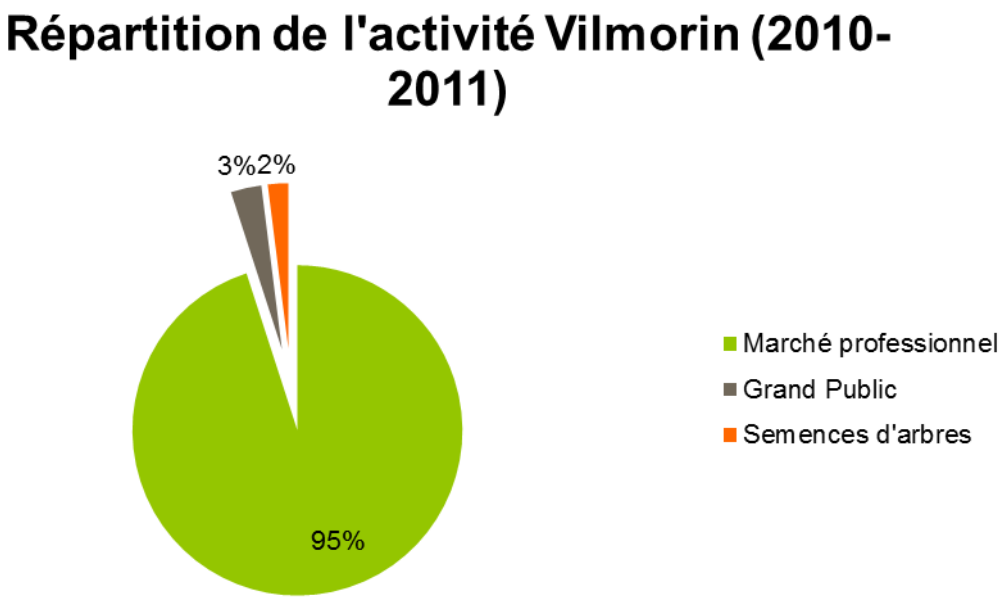


Figure 2 : Répartition de l'activité Vilmorin SA pour l'exercice 2010 – 2011 (source Vilmorin).

Avant-Propos

Mon stage de fin d'études en spécialisation Sciences et productions Végétales option Amélioration des Plantes s'intègre dans le cadre d'un cursus par apprentissage. Cette formation m'a permis d'intégrer l'équipe Vision artificielle de l'entreprise Vilmorin SA durant deux années, au cours desquelles j'ai eu l'opportunité de travailler sur des projets internationaux, portant sur le développement d'outils de phénotypage assisté par vision artificielle, essentiellement sur des organes de plantes potagères.

Vilmorin SA est une entreprise française aujourd'hui spécialisée dans la sélection, la production et la commercialisation de semences potagères et d'arbres. Les semences sont destinées aux professionnels, tels que les maraîchers, les producteurs de jeunes plants, les industriels (conserve, surgelé, 4ème gamme), les distributeurs et les pépiniéristes. L'activité de Vilmorin SA est développée à l'international grâce aux filiales commerciales (Vilmorin Iberica pour l'Espagne, Vilmorin INC pour les Etats Unis, Vilmorin Italia pour l'Italie, Vilmorin Do Brasil pour le Brésil…) aux stations de sélection, aux centres de recherche et aux distributeurs locaux. Depuis 1975, Vilmorin SA appartient au groupe Limagrain et s'allie à ces principaux concurrents regroupés en Business Units (BU) que sont Harris Moran - Clause, Hazera - Nickerson, Mikado Kyowa Seeds. L'ensemble de ces BU sont regroupés dans la branche des semences potagères sous l'identité « Vilmorin et Compagnie » (VCO) comme illustré Figure 1.

Vilmorin SA est une entreprise très concurrentielle et se place en tant que leader sur plusieurs espèces potagères, telles que le haricot d'industrie, la carotte, la laitue et la tomate allongée. Une trentaine de variétés sortent sur le marché tous les ans grâce à un investissement très important du chiffre d'affaires dans la recherche, près de 15.6% en 201-2012. Moins connue, l'activité de production de semences d'arbres représente 2% de l'activité en 2010/2011. Près de 300 espèces sont produites et près d'un arbre sur deux planté en France provient de Vilmorin. La production et la commercialisation de semences pour le marché amateur sous la marque « Vilmorin, depuis 1743 » est assurée par une filiale du groupe Limagrain, Oxadis. La part respective des différentes activités Vilmorin SA est représentée Figure 2.

Son siège social est implanté dans le bassin de production horticole de l'Anjou, à La Ménitré où est situé un centre de recherche. Différents services structurent l'entreprise après le comité de direction présidé par Rodolphe Millet : la Recherche, la Production qui comprend le service R&D Techno-semences et Vision Artificielle, le Contrôle Qualité, la direction des opérations (usine), le marketing et la communication, l'administration et la finance, les ressources humaines. Les activités de sélection sont coordonnées par la division Recherche qui est structurée par 8 équipes de sélections et qui disposent de laboratoires de pathologie, de transgénèse, de culture in vitro et de marquage moléculaire. La Production assure l'approvisionnement des semences (quantité, qualité, prix et délai) grâce à un réseau de multiplicateurs. Toutes les semences sont acheminées, triées, nettoyées, conditionnées et expédiées dans l'usine. Le service Contrôle qualité dispose de trois laboratoires répartis par domaine d'activité, un laboratoire pour les tests de germination, un laboratoire pour les tests de pathologie et un laboratoire de contrôle génétique pour les tests de pureté variétale. Le service Techno-semence et Vision Artificielle vient en support principalement à l'activité de l'usine, du Contrôle Qualité et à l'activité de la Recherche. Le deuxième site français de Vilmorin est situé à la Costière, entre Avignon et Nîmes, où sont implantés un centre de recherche et une partie de l'activité de la production de semences. L'entreprise compte globalement près de 570 personnes permanentes et une centaine de personnes en CDD et saisonnières.

Figure 3 : Réglette étalon pour le calibre des gousses de haricot (source Vilmorin).

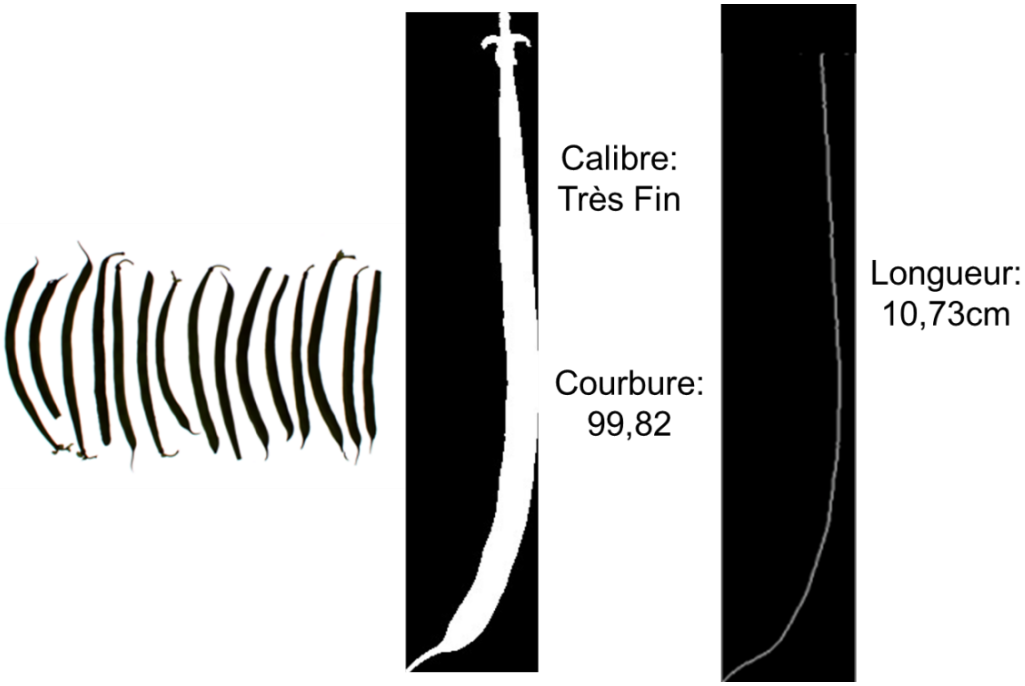


Figure 4 : Evaluation du calibre, de la courbure et de la longueur d'une gousse de haricot (source Vilmorin).

J'ai intégré l'équipe du service Technologie des Semences et Vision artificielle en apprentissage suite à un stage de Master 1 encadré par Cindy TORRES, chef de projet et docteur en traitement d'images. Sous la responsabilité de Pierre Damien Gouache (directeur de la Production), Pierre Ferraton dirige ce service, à la fois orienté sur la qualité des semences pour la partie Techno-semence et sur le développement de systèmes de vision assistés par ordinateur pour la partie Vision Artificielle.

Vilmorin SA s'est intéressée à l'imagerie dans les années 1990, par le biais d'Armand Feutry, responsable recherches, qui percevait déjà les avantages de cette technologie. Le pôle Vision artificielle chez Vilmorin SA se créé en 2011, suite à la thèse de Cindy Torres sur des problématiques d'imagerie. Ce pôle comprend deux chefs de projets R&D permanents, Cindy Torres pour la partie traitement d'images et Ali Boudjedra pour la partie automatisation des systèmes de vision artificielle. Les premiers besoins en imagerie ont d'abord été identifiés en sélection puis en Technologie des Semences. La première application fut l'évaluation des symptômes foliaires de Bremia sur jeunes plantes. La subjectivité de l'évaluation des symptômes (d'intensité variable pour le cas d'une résistance horizontale) a poussé à développer un outil qui puisse par imagerie, évaluer les surfaces foliaires atteintes par cette maladie. Un second projet, a ensuite été construit en collaboration avec le GEVES dans le cadre d'un contrat Etat-Région, sur la qualité germinative de graines. Des systèmes de vision par ordinateur se sont ensuite développés pour d'autres besoins comme par exemple le calibrage de gousses, la caractérisation de semences et le phénotypage de racines. Grâce à la filiale Vilmorin et Compagnie, des projets internationaux de vision artificielle se créent entre les différentes entreprises sur des problématiques communes dans un objectif de compétitivité du groupe. Des projets se sont développés pour les sélectionneurs comme par exemple, un projet sur le calibrage des gousses de haricot, un projet de caractérisation des racines de carottes et un projet de caractérisation du chou-fleur.

La machine destinée au haricot consiste à identifier de manière automatique les différents calibres des gousses de haricot dans un lot (manuellement réalisé avec une réglette auparavant, Figure 3), de compter le nombre de gousses par calibre, de les mesurer (longueur, courbure, surface) et d'évaluer le poids grâce au traitement des images (Figure 4). Le poids est estimé à partir de la surface des gousses.

Le projet sur la caractérisation des racines de carottes comprend deux parties, une partie ou une machine sera développée pour réaliser une première évaluation au champ des carottes et une deuxième partie, ou les carottes travaillées en sélection seront évaluées par analyse d'images grâce au développement d'un programme informatique et grâce au système d'acquisition développé par l'équipe de vision artificielle : l'Image Capture Box (ICAB).

La caractérisation des pommes de chou-fleur est construite de la même manière que pour l'évaluation des carottes travaillées en sélection, avec l'utilisation de l'ICAB pour l'acquisition des images et le développement d'un programme informatique pour le traitement des images. Mes principales missions ont été essentiellement portées sur ce projet, à savoir la création des bases d'images et de données, le test de différentes méthodes d'analyse d'images et la validation statistique de ces méthodes ainsi que les algorithmes développés. Le mémoire de fin d'étude portera sur ce projet, c'est-à-dire sur le développement d'un outil de phénotypage assisté par vision artificielle pour la sélection du chou-fleur.

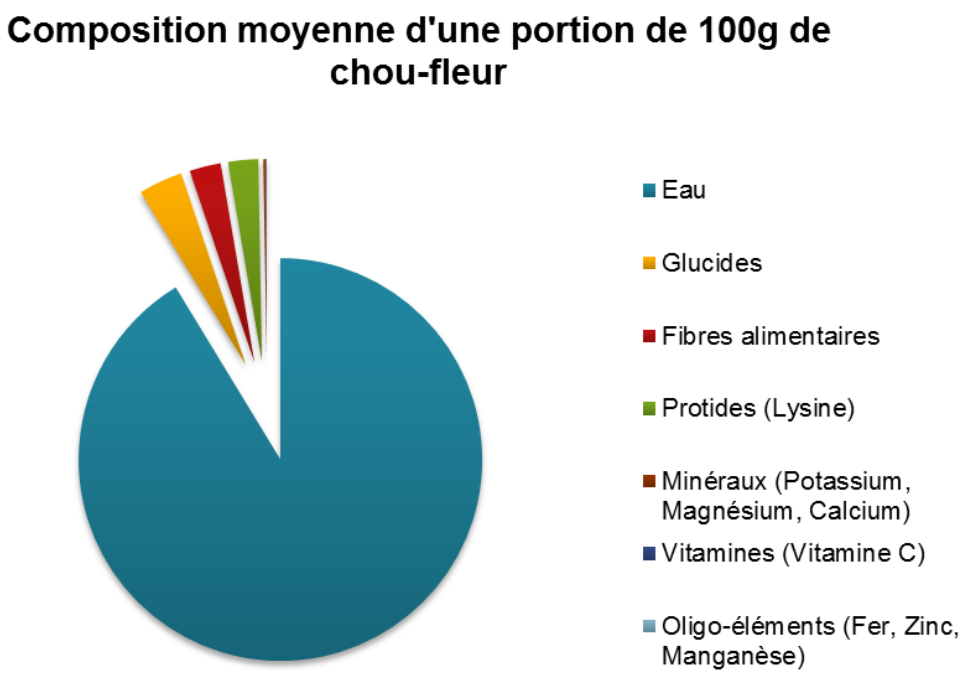


Figure 5 : Composition moyenne d'une portion de 100 g de chou-fleur (Aprifel).

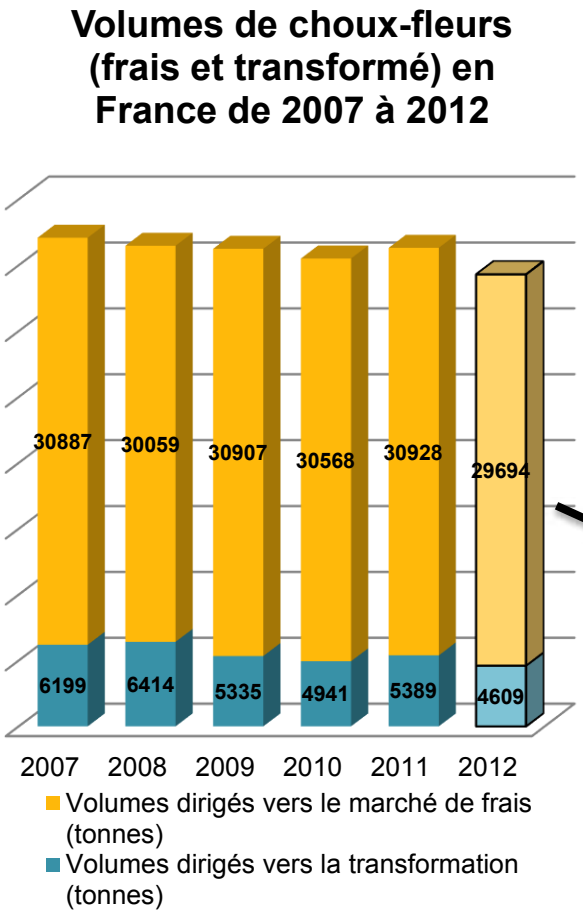


Figure 7 : Volumes de choux-fleurs (frais et transformé) en France de 2007 à 2012 (Agreste).

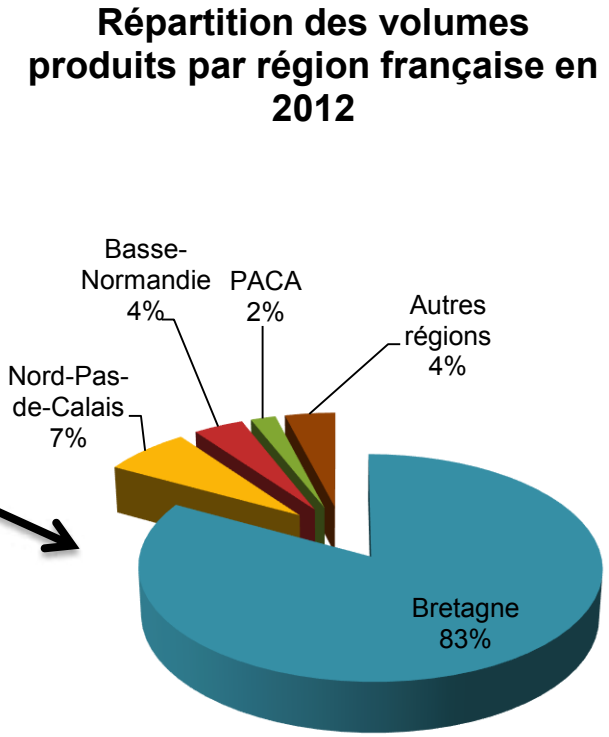


Figure 6 : Répartition des volumes produits par région en 2012 (Agreste).

Introduction

Le chou-fleur est une espèce végétale identifiée par Linnæus dont le nom taxonomique est *Brassica oleracea* L. var. *botrytis* convar. *botrytis*. Implantée dans le bassin méditerranéen comme le sont la plupart des espèces sauvages apparentées à *Brassica oleracea*, ce végétal a pour centre d'origine l'Italie, comme le brocoli (*Brassica oleracea* L. var. *italica*). Le chou-fleur a été domestiqué il y a 2000 ans et près de 500 ans de sélection et de mutations sur cette plante ont été nécessaires pour arriver à l'obtention d'un méristème floral hypertrophié de taille importante (R. Schmidt and I. Bancroft, 2011).

La « pomme de chou-fleur » ou la pré-inflorescence hypertrophiée récoltée au stade natif est consommée sous forme cuite et parfois crue (P. Tjeertes, 2004a). D'après la composition moyenne décrite par Aprifel et illustrée Figure 5, ce légume est très peu calorique, il représente 24 kcals soit 100 kJ pour 100g. Il contient de l'eau à hauteur de 90%, des fibres (2.5%), des protides (dont la lysine avec 2.4 g pour 100g, un acide aminé essentiel uniquement puisé dans l'alimentation), des vitamines dont la Vitamine C en forte proportion avec 60 mg pour 100g, des minéraux dont le potassium et des oligoéléments tels que le fer. Le chou-fleur contient aussi un glucosinolate, la sinigrine, responsable de son gout si spécifique (P. Tjeertes, 2004a). Ce composé bioactif présente des propriétés anticancérigènes (C. W. Beecher, 1994).

Malgré ces propriétés nutritionnelles intéressantes, la consommation française moyenne de chou-fleur a baissé depuis 30 ans et la consommation de chou-fleur et de brocoli stagne au cours des 10 dernières années (Agreste, 2009). Le chou-fleur est destiné soit au marché de frais où la pomme entière est utilisée, soit au marché de la transformation, appelé aussi marché de l'industrie (congélation, conserverie) où les fleurettes de la pomme sont détachées de la tige principale pour être transformées. Les informations qui suivent sont tirées d'une étude sur le chou-fleur de FranceAgriMer pour la campagne de 2012-2013 par (Y. Jacob and M. Lefort, 2013).

Les volumes de choux-fleurs en frais et en transformé depuis 2007 en France métropolitaine sont en diminution comme indiqué Figure 6 et les deux principales régions françaises productrices de chou-fleur en 2012 sont la Bretagne avec 83% et le Nord Pas de Calais avec 7% (voir Figure 7).

En France, les principaux acheteurs de chou-fleur en frais sont les magasins de la grande distribution (hyper, super et hard discount), les marchés et les primeurs qui représentent respectivement 60%, 17% et 12% en volume dans les achats de chou-fleur. Le chou-fleur est un légume consommé en frais traditionnellement et principalement en Bretagne par des retraités de classe moyenne. Le volume français de chou-fleur destiné au marché de frais a diminué de 4% entre 2011-2012. La consommation française en transformé est aussi en baisse, soit une diminution de 8% des quantités achetées par rapport à 2011. Pour le chou-fleur transformé, il s'agit principalement du chou-fleur surgelé et des plats cuisinés surgelés à base de chou-fleur. En restauration, le chou-fleur est le troisième légume surgelé utilisé, soit 45% des établissements utilisent en moyenne 238 Kg de chou-fleur par an. Les volumes français de chou-fleur destinés à la transformation ont diminués de 14% entre 2011-2012.

En Europe, les principaux pays producteurs de choux-fleurs/brocolis sont l'Espagne, l'Italie, la France, la Pologne et l'Allemagne. L'Allemagne et la Pologne ont des volumes de production en augmentation avec respectivement 176.7 et 306.8 milliers de tonnes en 2012. Ces deux pays ainsi que la France consacrent essentiellement leurs surfaces à la production de chou-fleur contrairement à l'Espagne et l'Italie qui eux se consacrent à la production de brocoli. L'Espagne est le premier producteur de chou-fleur/brocoli avec 505.6 milliers de tonnes en 2012. La France est en troisième position en 2011 derrière l'Italie avec plus de 380 milliers de tonnes. En 2010, la Pologne produit 25% des volumes européens de chou-fleur destiné à la surgélation. En ce qui concerne les échanges en frais dans l'union européenne, l'Espagne domine l'exportation du chou-fleur/brocoli et le Royaume Uni est le premier importateur de chou-fleur depuis des années.

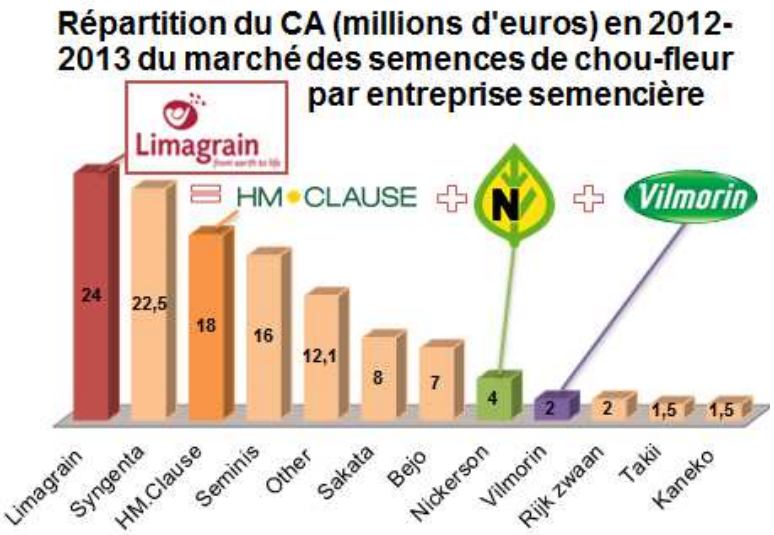


Figure 8 : Répartition du chiffre d'affaires en millions d'euros des semences de chou-fleur par entreprise semencière pour l'année 2011-2012 (E. Bizais, 2013)

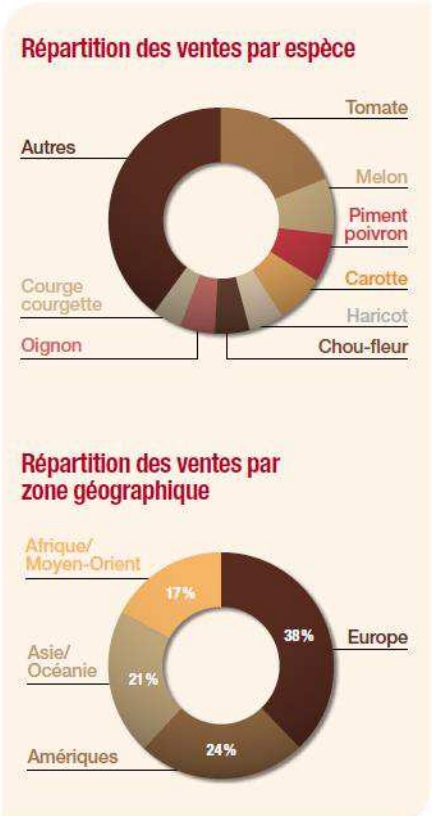


Figure 9 : Répartition des ventes 2011-2012 de semences potagères Vilmorin par espèce et par zone géographique (Vilmorin, 2013)

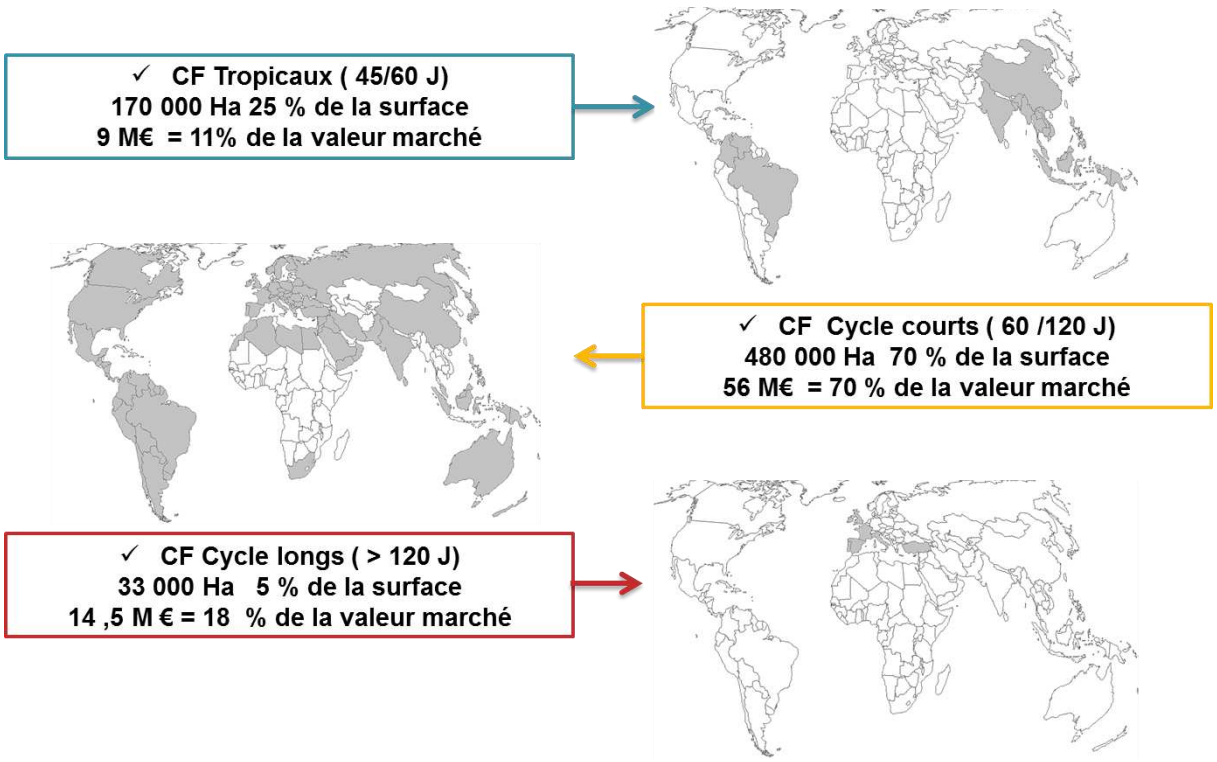


Figure 10 : Segmentation du marché du chou-fleur (CF) en trois catégories, les CF Tropicaux, les CF à Cycles courts et les CF à cycles longs, avec leurs surfaces de production respectives et leur valeur dans le marché (E. Bizais, 2013).

Le positionnement de ces deux pays est conservé à l'échelle mondiale. La baisse de production du chou-fleur s'observe aussi à l'échelle européenne (Agreste, 2009). La production asiatique et européenne représente respectivement 81.3% et 11% de la production mondiale de chou-fleur et de brocoli en 2012. La Chine et l'Inde sont les deux principaux producteurs asiatiques de chou-fleur et de brocoli (Food and Agriculture Organization of the United Nations, 2014).

Contrairement à la consommation et à la production du chou-fleur, le marché des semences de chou-fleur en valeur est stable voir en augmentation et représente près de 84 millions d'euros en 2013 (E. Bizais, 2013). Les semences potagères sont des semences à forte valeur ajoutée et le marché des semences de chou-fleur bénéficie encore du progrès génétique. Le passage de variétés populations aux variétés hybrides dans certaines régions s'opère encore (D. Drouet, 2014).

Pour Limagrain avec sa filiale Vilmorin et Compagnie, le chou-fleur est une espèce stratégique qui représente plus de 33% des parts de marché devant Syngenta et Seminis (Monsanto) sur l'exercice fiscal de 2012-2013 (J. Kolopp et al., 2013). HM-Clause, Hazera-Nickerson et Vilmorin S.A. ont respectivement 30%, 4% et 2.5% des parts de marché et se place respectivement en 2ème, 7ème et 8ème positions (Vilmorin, 2013), comme l'illustre la Figure 8. Ces entreprises sont à la fois concurrentes et complémentaires car ensemble, elles sont présentes sur la plupart des marchés. La collaboration entre ces entreprises ont permis le développement de co-hybrides comme par exemple Shakaris, un hybride positionné sur le créneau 80/120 jours (cycle de production décrit plus loin) et développé sur la campagne 2012-2013 entre Nickerson et Vilmorin (E. Bizais, 2013). Les lignées males sont produites par une des deux entreprises et les lignées femelles à stérilité mâle cytoplasmique CMS (dont la signification est donnée plus loin), sont produites par l'autre entreprise.

Harris Moran – Clause est représentée par le sélectionneur chou-fleur Rémi Leveil. Pour Hazera - Nickerson, Dick Zandt de l'entreprise Nickerson, représente la Business Unit en tant que sélectionneur chou-fleur. Enfin, chez Vilmorin, le sélectionneur chou-fleur Damien Drouet est aussi sélectionneur betterave avec 80% de son activité sur le chou-fleur et 20% sur la betterave. A présent, nous nous positionnerons selon le point de vue de l'entreprise Vilmorin S.A., pour introduire la problématique et le projet qui se sont construits dans le cadre de mon apprentissage.

Le chou-fleur est une espèce mineure de l'entreprise Vilmorin puisqu'elle représente seulement 4% du CA (soit 5 millions d'euros) en juin 2013 contre 28% pour la carotte, 14 % pour la tomate et 7% pour la laitue. Le chou-fleur se place en 6ème position comme illustré Figure 9 (Vilmorin, 2013). Les semences de chou-fleur Vilmorin sont commercialisées principalement aux Etats-Unis, en France, au Royaume-Uni et en Italie (E. Bizais, 2014). Les zones de production de chou-fleur pour Vilmorin sont situées en Maine et Loire où la culture est faite principalement sous abris et en Bretagne, près de Saint Malo. Le chou-fleur est aussi produit dans le sud de l'Espagne, en Amérique du Sud au Chili et en Amérique du Nord, aux Etats Unis (D. Drouet, 2014).

Le marché du chou-fleur est segmenté par la géographie, le climat et la longueur du cycle de production dont l'unité est le nombre de jours après plantation. Le chou-fleur en frais est aussi adapté pour la transformation. Toutefois, des critères plus spécifiques sont requis pour le marché de l'industrie. Par exemple, les choux-fleurs asiatiques sont des choux-fleurs à cycle court (45-60 jours après plantation) qui représente un marché de 9 millions d'euros en 2013 alors que les choux-fleurs d'hiver de l'Europe de l'ouest sont des choux-fleurs à cycle long (plus de 120 jours) et représente aujourd'hui un marché de près de 14 millions d'euros en 2013. Les choux-fleurs à cycles courts entre 60 et 120 jours après plantation représente 56 millions d'euros et constitue le plus gros marché, soit 70% en valeur. La segmentation du marché du chou-fleur est illustrée Figure 10 (E. Bizais, 2013). Vilmorin possède une gamme d'une vingtaine de variétés, dont les variétés Boris, Bodilis et Cendis qui ont fait son succès avec plus de 400 000 euros de chiffre d'affaires.

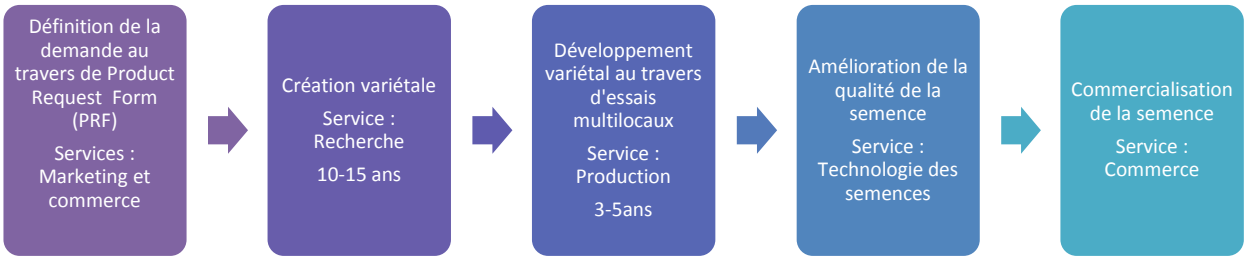


Figure 11 : Etapes de développement variétal depuis la définition d'un produit type jusqu'à la commercialisation de la semence créée.

Figure 12 : Résistances aux maladies recherchées par Vilmorin : la pourriture noire, le mildiou, la hernie des crucifères et la tâche annulaire (K. S. Kapoor, 1999; Koike et al., 2007; Pacific Northwest Plant Disease Handbook, 2014)

Boris et Bodilis sont des variétés de chou-fleur à cycle court tandis que Cendis est à cycle long. Cette dernière connait un succès depuis 15 ans (E. Bizais, 2014). Le succès d'une variété est expliqué par une bonne évaluation de la demande, dans des régions où la concurrence est peu implantée. Le succès est aussi dû au circuit bien rodé de l'entreprise (Figure 11). Les critères de l'idéotype variétal sont définis par les services marketing et commerce sous la forme de Product Request Form (PRF). Ensuite, pour répondre à ces exigences, les sélectionneurs mettent en place des plans de croisements pour sélectionner le matériel végétal sur la base des critères de sélection prédéfinis. Une fois le matériel végétal sélectionné, ces prototypes sont testés durant trois années de développement par la production, où ont lieu des essais multilocaux. Suite à ces essais, la décision est prise ou non de lancer commercialement la variété (qualité germinative ou nombre de graines par exemple). Intervient ensuite la division Technologie des semences, dont l'objectif premier est d'améliorer la qualité de la semence par exemple en retirant les graines plates, les graines avec des téguments abimés ou les graines contenant un taux élevé de chlorophylle (L. Gautier, 2014). Bien que tout ait été mis en place pour que la variété soit performante, la durée de vie des variétés tend à se raccourcir même si le marché est plutôt conservateur. Ceci est en partie expliqué par l'augmentation du nombre de concurrents et du nombre d'innovations variétales. De ce fait, l'objectif est de remplacer l'ensemble des variétés de la gamme tous les 10 ans (D. Drouet, 2014).

La production de chou-fleur est adaptée à tous les climats, ce qui permet de l'étaler sur toute l'année. Vilmorin, au travers de sa gamme variétale (principalement à cycle long) ne peut pas répondre aux besoins de production sur toute l'année, puisqu'elle possède une gamme dont la production recouvre une période allant du mois d'octobre jusqu'au mois de juin. Les objectifs aujourd'hui de Vilmorin sont donc de couvrir ces créneaux, d'être présent sur le marché des choux-fleurs à cycle court (80-120 jours) mais aussi de se placer sur d'autres segments comme le marché de l'industrie, majoritairement dominé aujourd'hui dans la filiale Vilmorin et Compagnie par Nickerson (D. Drouet, 2014). Les objectifs Vilmorin pour 2020 sont de doubler le CA de juin 2012 qui s'élevait à près de 4,1 millions d'euros (E. Bizais, 2013).

Pour répondre au marché de frais et dans un second temps au marché de la transformation, une variété performante pour Vilmorin est une variété sélectionnée sur la vigueur de la plante, la couverture de pomme par les feuilles, la blancheur, le poids et la forme de la pomme. Par exemple, les feuilles couvrantes permettent de protéger la pomme de son environnement (la lumière, les ravageurs) et joue donc un rôle dans sa conservation. La blancheur est un critère de qualité qui renvoie à la fraicheur du produit (C. Doré and F. Varoquaux, 2006). Enfin, la pomme doit être lourde et plutôt plate que ronde pour qu'elle puisse être facilement placée dans les cagettes. Dans un second temps, le chou-fleur doit être résistant à certaines maladies selon le marché considéré. La résistance à *Mycosphaerella brassicicola* est requise pour le marché d'hiver et le chou-fleur à cycle long. Pour le cycle court, le chou-fleur doit être résistant à *Peronospora parasitica*, à *Plasmodiophora brassicae* et surtout à *Xanthomonas campestris*. Les symptômes des principales résistances recherchées sont illustrés Figure 12.

Dans le cadre du marché de l'industrie, pour Vilmorin la pomme doit être très lourde, avec des fleurettes réparties de manière homogène dans la pomme se détachant facilement de la tige principale. La couverture de pomme est un critère moins important et le cycle de production du chou-fleur d'été ou d'automne doit être de 80-95 jours.

Vilmorin crée des variétés hybrides F1 (croisements de lignées fixées homozygotes) à partir d'une population améliorée. La sélection est aujourd'hui assistée par de nouveaux outils que sont les marqueurs moléculaires, des locus polymorphes de l'ADN. Une des pratiques utilisée en sélection chez Vilmorin est le rétrocroisement assisté par marqueur, en particulier pour l'introduction de la stérilité mâle cytoplasmique dans les lignées. Cette castration « génétique » de la lignée male devenue femelle permet un meilleur contrôle de l'hybridation (croisement femelle-male) et assure ainsi une production fiable d'hybrides F1.

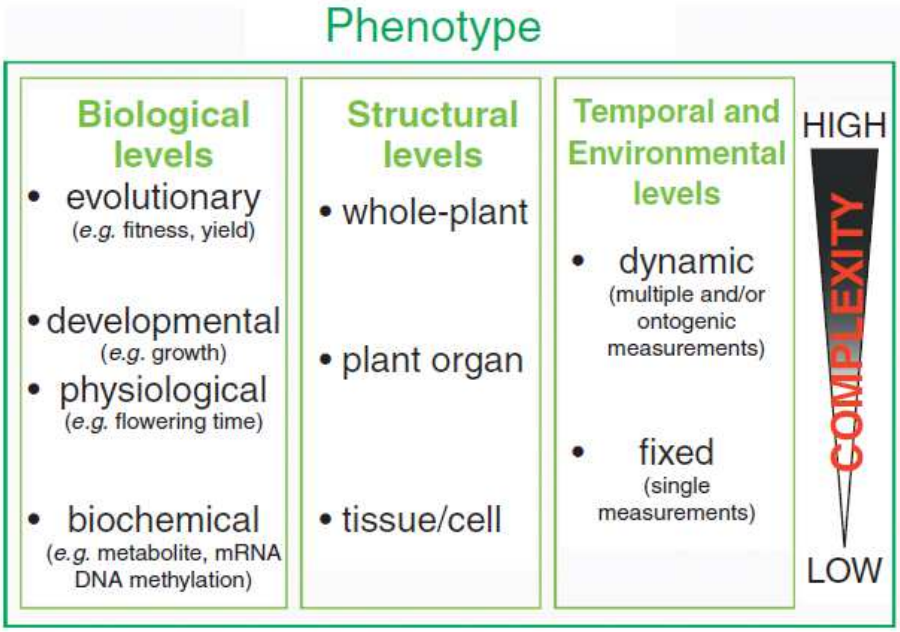


Figure 13 : Niveaux auxquels peut être évalué le phénotype d'une plante : biologique, structural, temporel et environnemental (C. Alonso-Blanco and B. Méndez-Vigo, 2014).

La stérilité est cytoplasmique et héritée maternellement. Les marqueurs permettent de suivre le fond génétique du génome récurrent, de façon rapide et précoce, abaissant ainsi les couts de production. Les marqueurs moléculaires présentent un réel avantage et la sélection précoce des plantes en n'est qu'une illustration. Vilmorin espère pouvoir développer des marqueurs pour sélectionner des têtes de lignées. Avant d'initier les schémas de sélection, les têtes de lignées sont sélectionnées sur la base d'une bonne Aptitude Générale à la Combinaison, c'est-à-dire sur la base de la valeur moyenne des descendants. Aujourd'hui, chez Vilmorin, l'évaluation de la descendance des lignées est faite grâce à deux ou trois plantes appelées testeurs et le développement des têtes de lignées dure 4 ans. La descendance issue des croisements lignées-testeurs est très hétérogène en termes de précocité et cet étalement de la production biaise son évaluation. Le développement de marqueurs sélectionnés sur des critères d'aptitude à la combinaison leur permettrait de sélectionner les têtes de lignées de manière plus efficace et plus rapide.

Bien que le marquage moléculaire soit encore couteux par rapport au phénotypage traditionnel, l'intérêt majeur de développer des marqueurs est de réduire la durée d'un programme de sélection estimé actuellement entre 7 et 15 ans selon les méthodes de sélection et les outils biotechnologiques utilisés (J. Kolopp et al., 2013; D. Drouet, 2014). Ce qui permettrait à Vilmorin de renouveler sa gamme variétale plus rapidement et d'accélérer le développement de nouvelles variétés sur le marché du chou-fleur à cycle court et le marché de l'industrie.

L'intérêt de développer des marqueurs est partagé à l'échelle de la filiale Vilmorin et Compagnie. Un projet de développement de marqueurs moléculaires liés à des caractères d'intérêt sur le chou-fleur a donc été lancé en 2012. Après le choix commun des caractères d'intérêt, l'étude est menée en génétique d'association (Linkage Desequilibrium Mapping) sur une population de 447 hybrides obtenus à partir du croisement de 149 lignées males et de trois lignées femelles à stérilité male cytoplasmique qui sont étudiées également en valeur propre et avec près de 20000 marqueurs Single Nucleotid Polymorphism. Compte tenu du caractère quantitatif des traits phénotypiques, l'analyse vise à détecter plusieurs régions du génome appelées Quantitative Trait Loci (QTL), contrôlant le caractère. La méthodologie mis en œuvre pour réaliser l'analyse génétique ne sera pas présentée dans ce rapport mais seulement discutée succinctement dans une dernière partie.

Afin de lier avec précision les marqueurs aux caractères, l'accent est mis dans ce projet sur le phénotypage et en particulier sur le développement d'un outil de phénotypage assisté par vision articifielle. En effet, le phénotypage est une étape déterminante puisqu'elle conditionne les résultats de l'analyse de détection de QTL. De plus, il existe un réel besoin d'augmenter la vitesse, la justesse et la précision de l'évaluation phénotypique (J. Cobb et al., 2013). Alors que le génotypage s'est fortement automatisé depuis les années 1990, le phénotypage est aujourd'hui considéré comme le principal frein aux analyses génétiques (R. T. Furbank and M. Tester, 2011; J. Cobb et al., 2013).

En sélection, les plantes sont évaluées sur la base de leur phénotype, c'est-à-dire sur la base de caractères observables. Le phénotype est le résultat d'un ensemble de composantes complexes que sont l'environnement, le génotype et l'interaction génotype-environnement. Le phénotype peut être évalué à différents niveaux biologiques, structuraux, temporels et environnementaux (C. Alonso-Blanco and B. Méndez-Vigo, 2014) ; Figure 13). Dans le cadre du projet, l'évaluation est faite ponctuellement sur les organes d'une plante arrivée à maturité. Traditionnellement, la caractérisation phénotypique est réalisée au champ, à différents instants de la journée et dans des conditions lumineuses fluctuantes. De plus, l'évaluation phénotypique est le résultat d'une réponse sensorielle, déterminée par des facteurs complexes que sont les facteurs psychophysiques et psychochimiques (Watada, 1980). Cette réponse sensorielle est variable entre expérimentateurs et au sein même d'un expérimentateur. Tous ces facteurs de variations rendent fragiles les notations d'autant plus que le travail de phénotypage est long, couteux et laborieux.

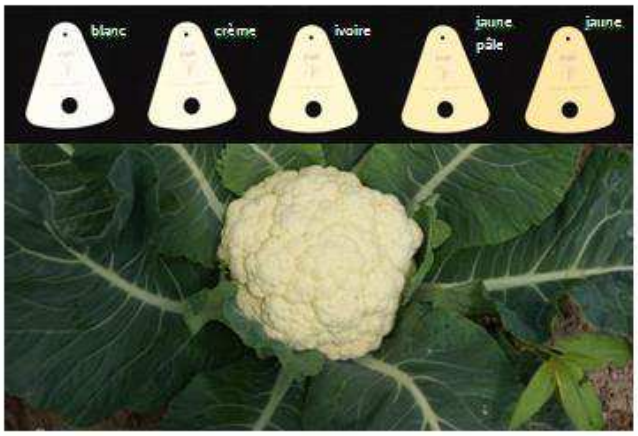


Figure 14 : Code couleur des inflorescences de chou-fleur (CTIFL, 1991). ***CTIFL : Centre Technique Interprofessionnel des Fruits et Légumes.***

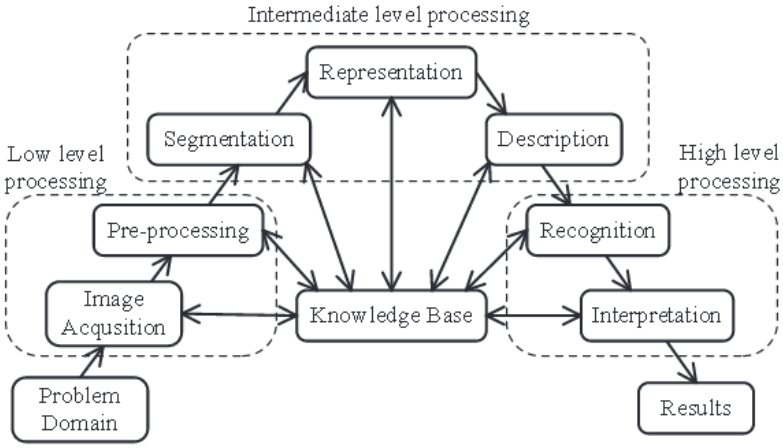


Figure 15: Différents niveaux de traitement d'images (Zhang et al., 2014).

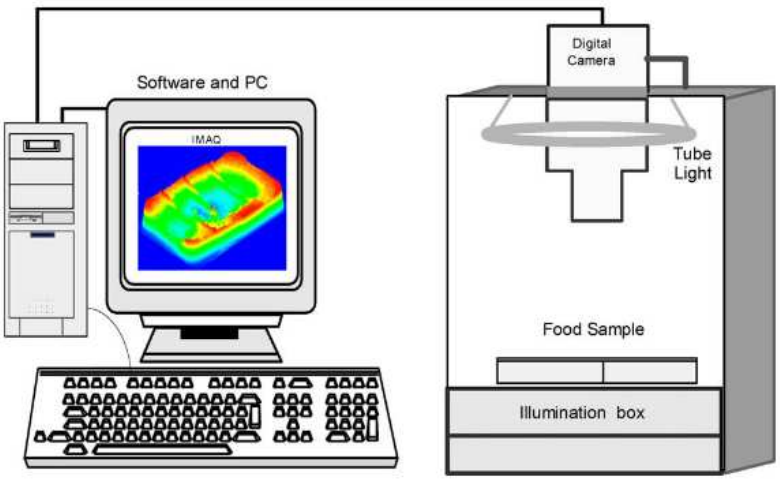


Figure 16 : Principaux composants d'un système de vision assisté par ordinateur (R. B. Pandit et al., 2007)

Pour diminuer la variabilité inter et intra individuelle, un sélectionneur junior est formé durant plusieurs années. Les notations deviennent alors répétables et précises mais valables dans des conditions expérimentales données, c'est-à-dire sur la base d'une référence de notations. Des références peuvent être introduites dans le dispositif expérimental (variété témoin) ou utilisées au cours du phénotypage comme par exemple l'utilisation d'une charte couleur développée pour le chou-fleur (CTIFL, 1991), illustrée Figure 14. Ces notations sont relatives et donc plus reproductibles. Toutes ces méthodes visent à entraîner ou aider l'expérimentateur à ce que la réponse sensorielle soit la plus reproductible, répétable, juste et exacte possible (P. Mac Leod and F. Strigler, 2009). Cependant, il n'existe pas d'observateur standard et il existera toujours des différences interindividuelles.

L'imagerie initialement destinée au domaine du médical et à l'industrie automobile et qui a vu le jour dans les années 1960 (G. A. Baxes, 1994), se développe dans le domaine du végétal à partir de la fin des années 1970 (C. Torres, 2010). L'agriculture est un domaine d'activités fortement dépendant de l'expertise humaine qui a su tirer profit de ces avancées technologiques devenues accessibles financièrement. Des instruments de mesures et des outils d'aide à la décision basée sur l'imagerie se sont ainsi construits. Les systèmes de vision assistée par ordinateur (ou artificielle, numérique, cognitive) sont des machines qui combinent différentes disciplines comme la mécanique, l'instrumentation optique, la détection électromagnétique, la vidéo numérique et le traitement d'images (K. K. Patel et al., 2012). Ces systèmes décrivent des objets physiques de manière explicite et significative sur des images (D.A. Ballard and C.M. Brown, 1982) et intègrent l'acquisition, le traitement et l'analyse d'images en deux dimensions (A.J.M. Timmermans, 1998) grâce à un système d'acquisition, du matériel et des logiciels informatiques (H.-H. Wang and D.-W. Sun, 2002). Les différents niveaux de traitement des images sont illustrés sur la Figure 15 (Zhang et al., 2014). Dans un premier temps, Les images sont acquises, numérisées puis pré-traitées. Le pré-traitement consiste à améliorer la qualité de l'image par exemple en augmentant le contraste ou en supprimant le bruit d'une image (dû à la température des capteurs de la caméra par exemple). La segmentation des images consiste à décomposer l'image en plusieurs zones d'intérêt selon différents type de méthodes, soit par seuillage, par contour, par région ou par classification. Les différentes régions segmentées sont ensuite représentées et décrites en termes de couleur, forme, texture ou de taille. L'analyse d'images consiste à mesurer et analyser statistiquement la valeur et la position des pixels des régions d'intérêt de l'image. Il existe principalement deux méthodes d'analyse quantitative : une méthode basée sur une simple analyse quantitative des caractéristiques de l'image (régions segmentées ou pixels) et une méthode de reconnaissance où les images sont classées à partir d'une base d'apprentissage (Zhang et al., 2014).

Un système de vision assisté par ordinateur se compose généralement d'une boite de prise d'images contenant un système d'éclairage, d'une caméra, d'un ordinateur et d'une interface informatique pour l'acquisition des images comme illustré sur Figure 16 (R. B. Pandit et al., 2007 ; H.-H. Wang and D.-W. Sun, 2002). En fonction des caractéristiques recherchées et donc de la question scientifique initiale, le système d'acquisition ainsi que le matériel informatique sont adaptés. La position, le type et la qualité colorimétrique des lumières sont des paramètres ajustables. En éclairage direct ou par rétroéclairage, diffus ou localisé, à incandescence ou à diodes électroluminescentes, dans le spectre du visible ou dans l'infrarouge, les sources lumineuses influent de manière importante sur la qualité et la performance de l'analyse d'images (B. G. Bachelor, 1985). Une image peut être obtenue par différents capteurs d'images, à partir d'un spectroscope ou un scanner mais le plus souvent par une caméra couleur ou monochrome (T. Brosnan and D.-W. Sun, 2004). Deux types de caméras sont principalement utilisées en vision artificielle et se distinguent par le type de capteur utilisé, des caméras CCD (Charge-Coupled Device) et des caméras CMOS (Complementary Metal-Oxide-Semiconductor), disposant d'un ou trois capteurs (mono ou tri-CCD, mono ou tri-CMOS). Le capteur transforme la lumière reçue en signaux électriques après avoir récupérer l'information soit de toutes les photodiodes (CCD), soit de chaque photodiode (CMOS),(C. Torres, 2010).

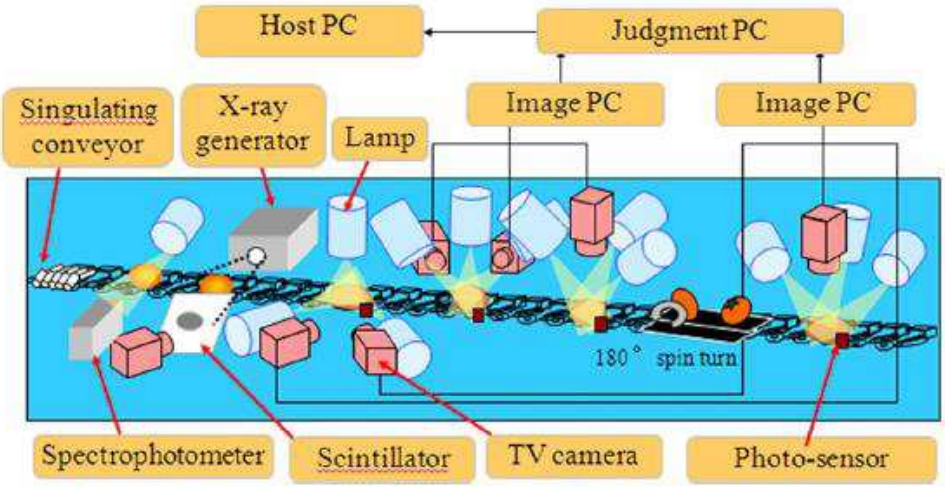


Figure 17 : Chaine automatisée par vision artificielle du tri des oranges (N. Kondo, 2010)

Le débit de ces machines automatisées varie en fonction du temps nécessaire depuis l'acquisition jusqu'au traitement des images mais aussi en fonction du nombre et de la durée des interventions humaines. Par exemple, la vitesse peut être de 2.2s par échantillon pour l'évaluation de la taille d'une pomme de chou pommé (soit 1600 objets par minutes) jusqu'à 3000 pommes triées par minute sur la base de défauts externes (S. Hayashi et al., 1998; Y. Tao and Z. Wen, 1999).

Afin d'illustrer le choix des différents capteurs d'images par rapport aux caractéristiques recherchées, une machine de tri assistée par vision artificielle développée par une équipe japonaise sera présentée. Les oranges sont triées en plusieurs catégories de qualité puis en fonction de leur taille à une cadence de 5-6 oranges par seconde. Une illustration de cette chaine est donnée Figure 17 (N. Kondo, 2010). Dans un premier temps, les oranges d'un container sont pris en charge par un convoyeur pour être alignées sur une chaine. Puis, les oranges sont acheminés au niveau d'un système d'inspection infra-rouge pour mesurer de manière non destructive leur teneur en sucre et leur acidité ainsi que leur granulation en surface. La granulation en surface est un indicateur de la teneur en eau du fruit. Un système à rayons X permet dans un second temps de détecter de manière non destructive la présence ou non d'un défaut de la structure interne de l'orange appelé « rind-puffing » en anglais. Ensuite, les oranges sont dirigées vers six caméras couleurs pour une évaluation des caractéristiques externes. Quatre caméras vont visionner les côtés de l'orange alors que deux autres vont prendre en images le dessus et le dessous de l'orange. Les caractéristiques mesurées sont la taille, la forme, la couleur et plusieurs défauts externes. Ces informations sont prises en compte pour réaliser le tri (Kondo, 2010).

Appliqué dans le domaine du végétal, de tels outils biométriques ont prioritairement été développés pour l'évaluation de la qualité des produits agricoles de la filière des fruits et légumes, pour des considérations économiques et politiques. En effet, la valeur commerciale des fruits et légumes destinés au marché de frais est définit à l'échelle européenne par des critères de qualité externe et interne (INTERFEL). Des machines de tri se sont donc développées et des entreprises se sont spécialisées dans leur production comme par exemple la société Ellips ou encore l'entreprise Fachaux (Fachaux, 2013; Ellips B. V., 2014). Cette dernière fabrique des calibreuses notamment adaptables pour le chou-fleur avec un choix possible de la méthode de classification (poids, couleur, défauts, diamètre). Ces machines sont adaptées pour un besoin particulier qu'est le tri et sur la base d'un faible nombre de critères et d'une variabilité restreinte.

D'autres applications des systèmes de vision artificielle existent comme celui d'étudier le déterminisme génétique de caractères complexes. De nombreuses plateformes de phénotypage se sont développées pour une étude dynamique de certains caractères tel que la résistance à la sécheresse par exemple. Des entreprises se sont également spécialisées dans la création de matériel adapté à ce type d'analyse comme par exemple l'entreprise Lemnatech. Aucune étude à ce jour est menée dans ces plateformes de phénotypage sur la caractérisation des pommes de chou-fleur.

De nombreux programmes informatiques de traitement et d'analyse d'images se sont développés et sont utilisables grâce à des logiciels libres tels que ImageJ. Par exemple, un plugin (extension d'un programme informatique) a été développé pour évaluer morphologiquement des racines de carottes, des gousses de pois et de haricot, des cotylédons de betterave à sucre, des graines et des capsules de lin (G. Polder et al., 2012). Ces données morphologiques obtenues de façon objective sont utilisées pour la caractérisation et l'identification variétale (S. R. Drapper and P. D. Keefe, 1989; G. Polder et al., 2012), pour répondre aux critères d'obtentions végétales définis par l'Union internationale pour la Protection des Obtentions Végétales (UPOV) que sont la Distinction, l'Homogénéité et la Stabilité. Or, les critères utilisés précisément pour l'identification variétale ne sont pas toujours les mêmes critères que ceux utilisés par les sélectionneurs pour identifier et choisir leur matériel végétal.

D'après ces recherches bibliographiques, il n'existe pas de machines développées spécifiquement pour les sélectionneurs sur le chou-fleur qui puissent mesurer un grand nombre de caractères, ni de programmes informatiques qui puissent mesurer tous les caractéristiques recherchées.

C'est dans ce contexte que s'introduit le développement d'un outil de phénotypage assisté par vision artificielle par l'équipe Vision Artificielle de Vilmorin à savoir, qu'est-ce que peut apporter la vision artificielle par rapport aux techniques traditionnelles de phénotypage pour l'étude du contrôle génétique de caractères complexes et in fine pour la sélection du chou-fleur.

Différentes questions se posent :

- Quel est le système d'acquisition et le matériel informatique le plus adapté ? Quelles sont les conditions d'utilisation de ce système ?

- Comment mesurer les caractères du chou-fleur sur images, quelles méthodes d'analyse d'images? Comment préparer le chou-fleur aux prises de vues ?

- Comment organiser les données pour développer les programmes informatiques ? Comment valider les algorithmes développés ?

La méthodologie mise en œuvre pour développer le système d'acquisition ne sera pas présentée, mais le matériel utilisé sera décrit. De plus, la méthodologie utilisée pour adapter le système d'acquisition au phénotypage du chou-fleur sera brièvement présentée. Ensuite, nous présenterons une partie du travail de préparation du matériel végétal pour les prises de vues ainsi que le développement des bases de données et d'images. Le développement des programmes informatiques ne sera pas présenté, en revanche, la méthodologie utilisée pour choisir et valider statistiquement les méthodes d'analyse d'images sera donnée ainsi que celle utilisée pour la validation statistique des algorithmes.

Une partie des résultats statistiques seront présentés afin d'illustrer la validation de certaines méthodes d'analyse d'images sur quelques caractères. De plus, un bilan sera fait sur l'apport de l'outil de phénotypage, à savoir ses qualités en tant qu'instrument de mesure.

Nous discuterons dans un dernier axe, des atouts et des limites d'utilisation de cet outil. Nous ouvrirons sur les possibilités d'utilisation de systèmes de vision artificielle pour le phénotypage des espèces cultivées.

Table 1 : Liste et descriptif des caractères mesurés par imagerie.

Organe(s) concerné(s)		Descriptif du caractère et de la mesure (2D ou 3D)	Prise(s) de vue(s)
Pomme de chou-fleur sans les feuilles	Diamètre	Le diamètre de la pomme de chou-fleur correspond au segment de droite de distance maximale passant par le centre de la pomme et délimité par les contours de la pomme, 2D.	Vue du dessus
	Hauteur	La hauteur de pomme est un segment de droite mesuré depuis la base de la pomme (dernière couronne de fleurettes) jusqu'au sommet de la pomme, 2D.	Vue en coupe longitudinale
	Surface	La surface de la pomme de chou-fleur représente la surface occupée par les fleurettes vue du dessus, 2D.	Vue du dessus
	Volume	Le volume ici est mesuré sur la pomme entière en prenant en compte les espaces vides internes, 3D.	Vue du dessus, vue en coupe longitudinale
	Caractère pointu	Le caractère pointu est évalué au sommet de la pomme (pointe conique observée au sommet de la pomme), 3D.	Vue en coupe longitudinale
	Forme	La forme correspond au caractère rond, ovale ou plat des pommes observées sur le côté (ovale horizontalement = plate, ronde, ovale verticalement=ovale), 2D.	Vue en coupe longitudinale
	Caractère lisse	Le caractère lisse s'observe sur l'ensemble de la surface de la pomme depuis la dernière couronne jusqu'au sommet et correspond à l'intensité des irrégularités observées en surface, 3D.	Vue du dessus, vue en coupe longitudinale
	Couleur	La couleur de la pomme s'évalue à la base des fleurettes (partie cachée par les feuilles), 2D.	Vue en coupe longitudinale face externe
	Rose et vert interne	Le rose et le vert interne sont des colorations qui s'observent à l'intérieur de la pomme, en coupe longitudinale, 2D.	Vue en coupe longitudinale
	Rose externe	Le rose est une coloration externe qui apparait en surface de la pomme, 2D.	Vue du dessus
	Nombre et surface des bractées	Les bractées sont des petites feuilles vertes mesurées en surface de la pomme, 2D.	Vue du dessus
	Fermeture	La fermeture de pomme correspond à l'espacement entre la tige principale et les fleurettes de la dernière couronne. Plus la pomme est dite « fermée », plus l'espacement est petit, 2D.	Vue du dessous
Pomme de chou-fleur avec les feuilles	Couverture	La couverture de pomme est mesurée grâce au recouvrement des feuilles sur la pomme, 2D.	Vue du dessus
Pomme de chou-fleur avec ou sans les feuilles	Tige creuse	La tige est dite creuse lorsqu'un trou y est observé, 2D.	Vue du dessous (avec ou sans feuilles), vue en coupe longitudinale (sans feuilles)

I. Matériel végétal

A. Caractéristiques générales

Les pommes de chou-fleur ont été prélevées par l'équipe Vision Artificielle et les équipes de sélection au cours de plusieurs essais menés en plein champ en France et aux Pays-Bas. L'échantillonnage a été effectué sur différentes périodes de production du chou-fleur (principalement l'été et l'automne) et sur du matériel végétal génétiquement variable. Les choux-fleurs prélevés sont en majorité des lignées, cependant quelques hybrides ont été aussi photographiés. La plante est étudiée à un stade intermédiaire où l'inflorescence est hypertrophiée. Les échantillons ont été prélevés sur un ensemble de pommes de chou-fleur représentatives de la variabilité phénotypique des caractères, soit un prélèvement minimal de 30 pommes de chou-fleur par caractère. Pour les caractères où une échelle de notation est utilisée, un prélèvement est mené pour chaque note. Au total, près de 900 pommes de chou-fleur ont été photographiées (en comptant le nombre d'images de pommes de chou-fleur vues du dessus).

B. Les caractères étudiés, notés et mesurés sur le chou-fleur

Les caractères étudiés par analyse d'images ont été définis par les différentes équipes de sélection chou-fleur en fonction de leurs importances pour chaque équipe. La liste des caractères a évolué au cours du projet. Cette liste est déterminée par la faisabilité des mesures par analyse d'images, soit 14 caractères ont été retenus sur 17 donnés initialement. Les différents traits évalués phénotypiquement au vue d'une étude par analyse d'images sont présentés et décrits dans la Table 1. Les critères uniquement évalués visuellement pour l'étude génétique ne seront pas présentés dans ce rapport.

Les caractères qui attraient à la morphologie de la pomme du chou-fleur sont : le diamètre, la hauteur, la surface, le caractère lisse de la surface de la pomme de chou-fleur, le caractère pointu du haut de la pomme de chou-fleur, le volume, la forme et l'espacement entre les fleurettes de la dernière couronne et le tronc appelé « fermeture » de pomme. Les caractères liés à la couleur sont : la blancheur de la pomme de chou-fleur, le rose externe, les vert et rose internes de la pomme de chou-fleur, le nombre et la surface des bractées, la couverture de la pomme par les feuilles et la tige creuse.

Pour le diamètre, l'organe concerné est la pomme de chou-fleur sans les feuilles et le chou-fleur est observé vue du dessus. La mesure est en deux dimensions et correspond au segment de droite de distance maximale passant par le centre de la pomme de chou-fleur et délimité par les contours de la pomme de chou-fleur. Pour tous les autres caractères, se référer à Table 1.

Une fois les pommes récoltées, le phénotype est évalué par les experts au champ ou dans des locaux d'expérimentation pour la validation terrain. Les caractères sont évalués physiquement par des instruments de mesure, visuellement par des experts mais aussi informatiquement sur images. La validation terrain consiste à confronter les résultats obtenus par le traitement des images aux données terrain que sont les notations, les mesures physiques et les mesures supervisées sur images (pour la validation des algorithmes). Si un parallèle est fait avec la sélection génomique, l'objectif commun est de prédire le phénotype. Mais dans le cas de notre étude, la prédiction est faite grâce aux résultats d'analyse d'images.

Table 2 : Descriptif des mesures (physiques, visuelles, sur images) menées pour la validation terrain.

Caractères	Mesures pour la validation terrain (physique, visuelle)	Unités de mesures (mesures physiques ou évaluation visuelle)	Mesures sur images sous ImageJ
Diamètre	Mesure du diamètre maximal (pied à coulisse, mètre ruban)	cm	- surfaces détourées (SSH), - diamètre tracé à main levée sur images (Straight Line Selection Tool), - diamètre maximum de Feret (Feret).
Hauteur	Mesure de la hauteur (pied à coulisse, mètre ruban)	cm	- surfaces détourées (SSH), - hauteur tracée à main levée sur images (Straight Line Selection Tool), - diamètre minimum de Feret (MinFeret).
Surface		cm²	- surfaces détourées à main levée sur images (Freehand Selection Tool), - surfaces détourées (SSH).
Volume	Mesure du volume par la méthode du déplacement d'eau (récipient doseur)	mL	- surfaces détourées à main levée sur images (Freehand Selection Tool), - surfaces détourées (SSH), - hauteur et diamètre tracés à main levée sur images (Straight Line Selection Tool), - diamètre minimum de Feret (MinFeret), - diamètre maximum de Feret (Feret), - mesures testées : AspectRatio, Roundness, Circularity.
Caractère pointu	Notations	Présence/Absence	- surfaces détourées (SSH), - coordonnées du diamètre maximum de Feret, - coordonnées du contour des surfaces et distances par rapport au diamètre de Feret, - point culminant, régressions linéaires sur une sélection de plusieurs pixels de part et d'autres de ce point, calcul d'un angle.
Forme	Echelle de notations	Oval, Plat, Rond	- surfaces détourées (SSH), - coordonnées du diamètre maximum de Feret, - coordonnées du contour de la surface, - sélection des coordonnées de la partie supérieure (délimitée par le diamètre de Feret) puis projection symétrique centrale des coordonnées, - profil rayons, - mesures testées : AspectRatio, Roundness, Circularity
Caractère lisse	Echelle de notations	1 à 9 (bossu à lisse)	- surfaces détourées (SSH), - coordonnées du diamètre maximum de Feret, - coordonnées du contour de la surface, - profil rayons, - mesures testées : Circularity, Solidity.
Couleur	Echelle de notations, charte couleur	1 à 9 (jaune à blanc) Charte couleur : C2 (blanc) à C10 (jaune)	- surfaces détourées (SSH), - sélection de plusieurs zones (Freehand Selection Tool)
Rose et vert interne	Notations	Présence/Absence	- surfaces détourées (SSH), - sélection de plusieurs zones (Freehand Selection Tool)
Rose externe	Notations	Présence/Absence	- surfaces détourées (SSH), - sélection de plusieurs zones (Freehand Selection Tool)
Nombre et surface des bractées	Notations	Présence/Absence, nombre de bractées	- surfaces détourées (SSH), - surfaces détourées à main levée sur images (Freehand Selection Tool)
Fermeture	Echelle de notations	1 à 9 (Espacement important à peu important)	- surfaces détourées (SSH), - surfaces détourées à main levée sur images (Freehand Selection Tool)
Couverture de pomme	Echelle de notations	1 à 9 (pomme non couverte à couverte totalement)	- surfaces détourées (SSH), - surfaces détourées à main levée sur images (Freehand Selection Tool)
Tige creuse	Notations	Présence/Absence	- surfaces détourées (SSH), - surfaces détourées à main levée sur images (Freehand Selection Tool)

Les notations

Les notations ont été harmonisées entre les équipes de sélection, au cours de réunions et une échelle commune a été définie pour chacun des caractères. Toutefois, l'échelle de notations peut être différente de celle utilisée quotidiennement par les sélectionneurs. Les notations au champ sont réalisées soit par les sélectionneurs, soit par les techniciens de sélection et il n'y pas de répétitions faites au champ. Deux types de notations sont utilisés, soit une échelle de catégories soit une notation de type présence/absence (Table 2).

Par exemple, pour la couleur, l'échelle de notations utilisée varie de 1 à 9 et correspond à des pommes dont la couleur varie du jaune au blanc. Un code couleur pour le chou-fleur développé par le CTIFL est également utilisé. Une illustration est donnée en introduction (Figure 14). Pour les autres caractères, se référer à la Table 2.

Pour la forme, le caractère pointu et le caractère lisse de la pomme de chou-fleur, deux séries de notation ont été menées en plus sur images. La méthode utilisée pour mener à bien la première série de notations est précisée en Annexe I. Face à la subjectivité de cette première série de notations (menées sur images de façon indépendante par les experts), une deuxième série de notations est réalisée sur images par les experts, de façon concertée. Pour le caractère lisse, la notation est faite par comparaison à des images de référence. Cette deuxième série de notations a permis de définir une nouvelle base de référence, une base d'apprentissage pour la forme, le caractère pointu et le caractère lisse.

Les mesures physiques

Les mesures physiques des caractères sont menées par divers appareils de mesures : un pied à coulisse et un mètre ruban pour les mesures de hauteur et de diamètre, un récipient doseur pour la mesure du volume par la méthode du déplacement d'eau et un colorimètre « CR-400 » de marque Konica Minolta Sensing INC, pour la mesure de la couleur (tons de blanc, rose, vert). Cet appareil de mesure colorimétrique est présenté en Annexe II.

Les mesures supervisées sur images

Les mesures sur images sont réalisées sous ImageJ par des experts. Les mesures sont dites supervisées, lorsque les mesures sont réalisées de manière non automatisée. Les mesures sur images peuvent aller jusqu'au tracé manuel de régions d'intérêt. Des méthodes d'analyse d'images sont testées, comme par exemple la mesure automatisée du diamètre maximum de Feret pour la mesure du diamètre. Pour calculer ce diamètre, il a fallu au préalable sélectionner la zone d'intérêt, soit manuellement, soit par une segmentation par seuillage de l'histogramme (SSH). Pour chaque caractère, les mesures réalisées sur images sont listées dans la Table 2. Pour les mesures sur images sous ImageJ, les outils de sélection sont détaillés dans le guide d'utilisation du logiciel (W. Rasband, 2012).

Dans un premier temps, l'objectif est de trouver des méthodes d'analyse d'images, qui traduisent au mieux ce qui est mesuré ou noté sur le chou-fleur. Les résultats des méthodes d'analyse d'images sont confrontés aux mesures physiques et aux évaluations visuelles pour valider les méthodes. Dans un second temps, ces méthodes d'analyse d'images une fois validées, sont implémentées dans des programmes informatiques par des experts en traitement d'images. Ces programmes informatiques sont ensuite évalués et confrontés aux mesures physiques, aux évaluations visuelles mais aussi aux mesures supervisées sur images. En résumé, deux validations sont menées, une première pour les méthodes d'analyse d'images et une deuxième pour les programmes informatiques qui intègrent ces méthodes. La méthode d'analyse d'images est parfois directement implémentée dans le programme informatique pour être validée en même temps que l'algorithme.

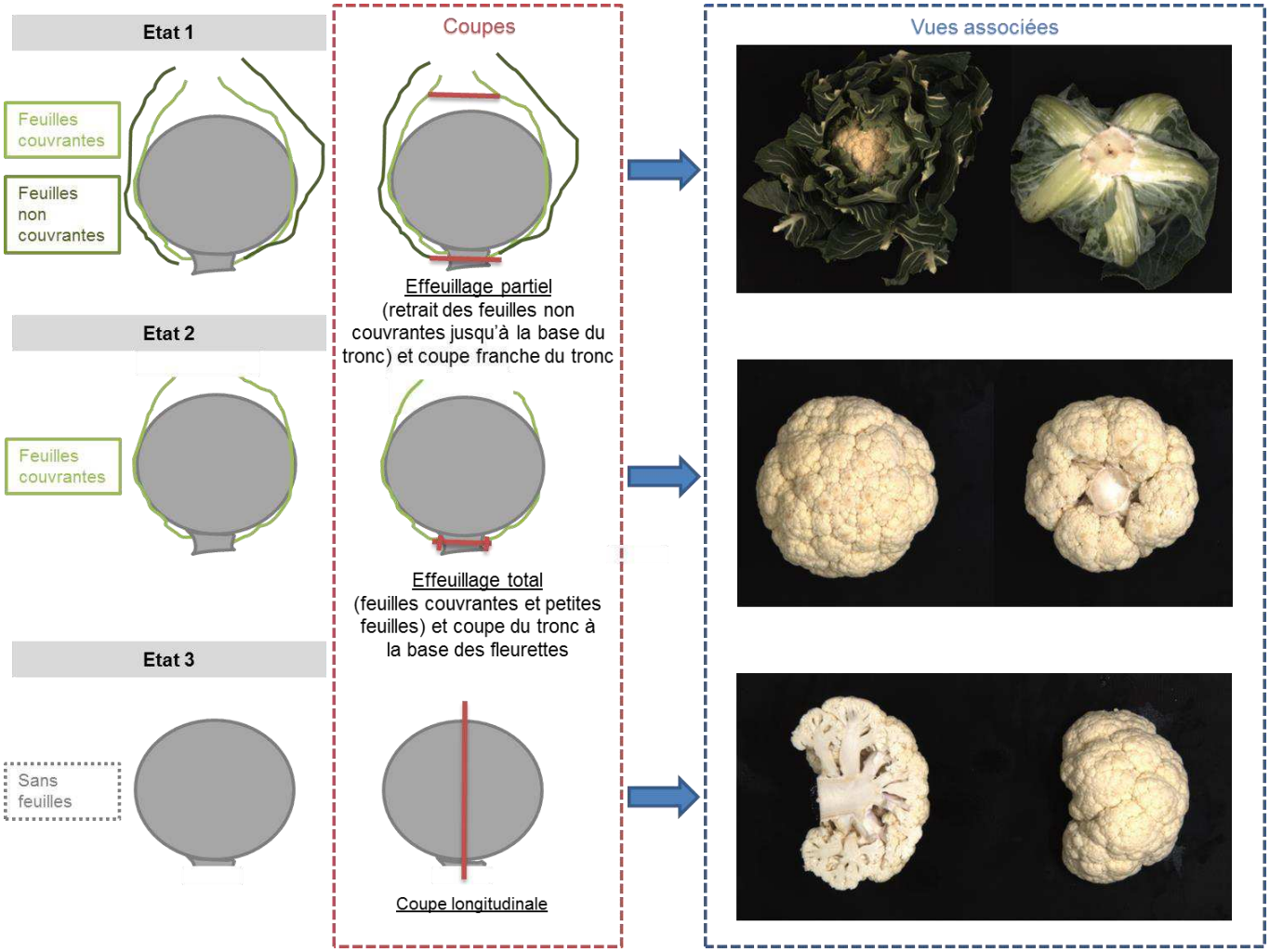


Figure 18 : Etat des choux fleurs, coupes réalisées et photographies associées (Vilmorin).

Figure 19 : Image Capture Box (Vilmorin)

Figure 20 : Rampes lumineuses de l'ICAB (Vilmorin)

C. Préparation du chou-fleur aux prises de vues

Le chou-fleur, une fois coupé au champ, est effeuillé partiellement. Les feuilles non couvrantes de la pomme sont entièrement ôtées et coupées à la base du tronc. Une deuxième coupe est réalisée en haut de la pomme, à 2 cm du sommet de la pomme de chou-fleur, telle une coupe commerciale. Le chou-fleur est ainsi pris en photo vue du dessus et vue du dessous avec les feuilles. Ensuite, toutes les feuilles couvrantes sont également coupées à la base du tronc. La pomme de chou-fleur nue est ainsi photographiée vue du dessus et vue du dessous. Enfin, la pomme est coupée longitudinalement, axé par rapport au tronc. Une moitié est tournée face interne et l'autre moitié est tournée face externe pour la dernière prise de vue. L'ensemble de ces étapes sont décrites Figure 18. La préparation du chou-fleur pour les prises de vues conditionne les résultats par analyse d'images. Cette préparation doit donc être reproductible, d'où la mise en place d'un protocole de préparation du chou-fleur. Par exemple, les experts ont dû harmoniser la coupe pour le critère de recouvrement des feuilles. Différentes coupes du feuillage existent et celle retenue est la coupe commerciale.

II. L'outil de phénotypage par imagerie

A. Le système d'acquisition

Les images sont prises en environnement contrôlé, dans une boite à images développée par l'équipe Vision Artificielle Vilmorin pour Vilmorin et Compagnie sous le nom d'ICAB pour Image CApture Box (Figure 19). Cette boite générique existe en format transportable ou fixe. Elle est composée de deux lampes Phillips à diodes électroluminescentes dont la température de couleur est de 5000 K (Figure 20). Un tapis noir adapté (pas d'effet d'ombres de l'objet volumineux, bon contraste entre le fond noir et le chou-fleur) est disposé au fond de la boite. La démarche entreprise pour choisir les matériaux de la boite ainsi que les lumières ne sera pas présentée dans ce rapport.

Les images sont prises grâce à une caméra de 10 millions de pixels (de marque IDS uEye), de technologie CMOS, avec une définition maximale de 3840 x 2748 pixels. L'image est décomposée en trois canaux : le Rouge, le Vert et le Bleu (RVB). Pour chaque pixel de l'image, l'intensité du Rouge, du Vert et du Bleu est donnée en 8 bits, soit des valeurs comprises entre 0 et 255. La valeur maximale 255 signifie que l'intensité est maximale pour le Rouge, le Vert ou le Bleu sur l'image dite en niveau de gris. La distribution des intensités de l'image est donnée par l'histogramme. Les images RVB sont enregistrées au format TIFF. La caméra est reliée à un ordinateur qui dispose d'une mémoire vive et d'une capacité de stockage suffisante pour l'acquisition et la sauvegarde des images non compressées et non réduites. Une trentaine de choux-fleurs étudiés par jour représente environs 5 Gigaoctet d'images par jour.

B. Guide d'utilisation

Un guide d'utilisation est destiné aux opérateurs. Il décrit pas à pas l'utilisation de l'ICAB, les réglages de la caméra et l'acquisition des images de pré-traitement et de chou-fleur. Par exemple, pour le réglage de la caméra, la netteté et la luminosité des images sont ajustées respectivement grâce aux mollettes du focus et de l'ouverture de la caméra. La luminosité des images peut également être ajustée grâce au logiciel de la caméra, en jouant sur le temps d'exposition. Une charte couleur (QPcard 201) est utilisée pour réaliser ces ajustements. En vérifiant les valeurs du rouge, du vert et du bleu des patchs de référence d'une photographie de QPcard, la luminosité est ajustée pas à pas avec la bague de réglage de l'ouverture ou

avec le logiciel grâce au temps d'exposition. Des informations réglages de la netteté et de la luminosité des images sont données en Annexe IV.

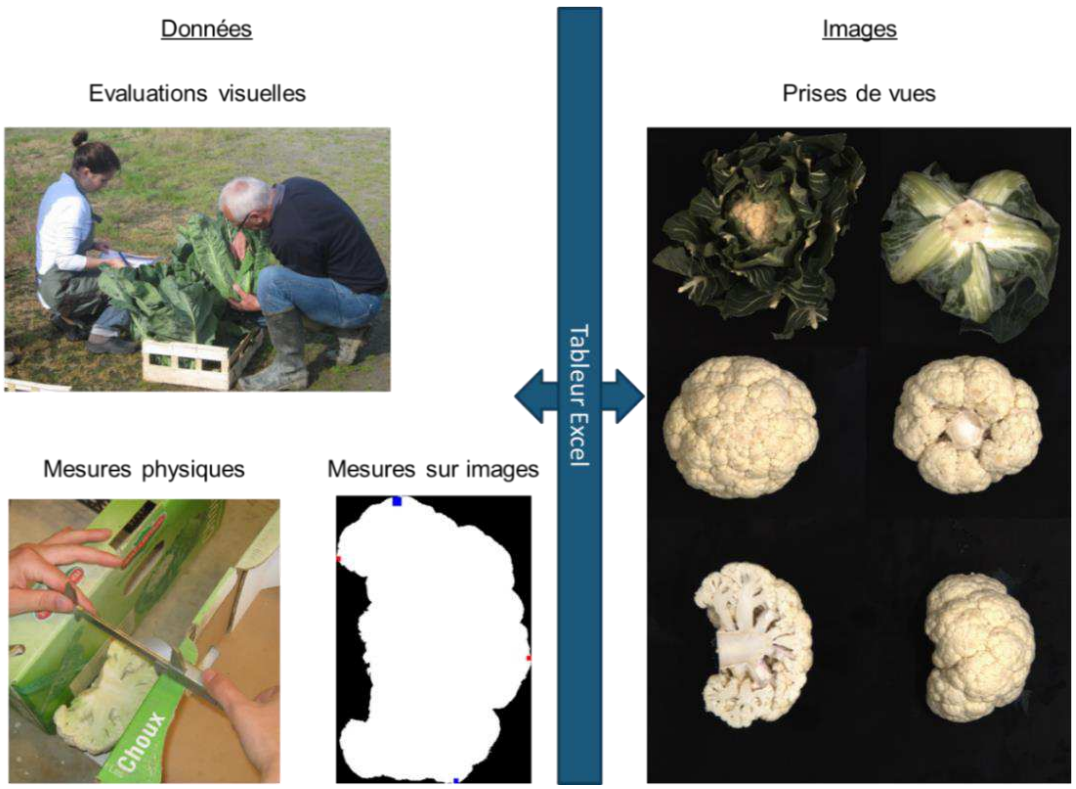


Figure 21 : Présentation des bases de données et d'images (Vilmorin).

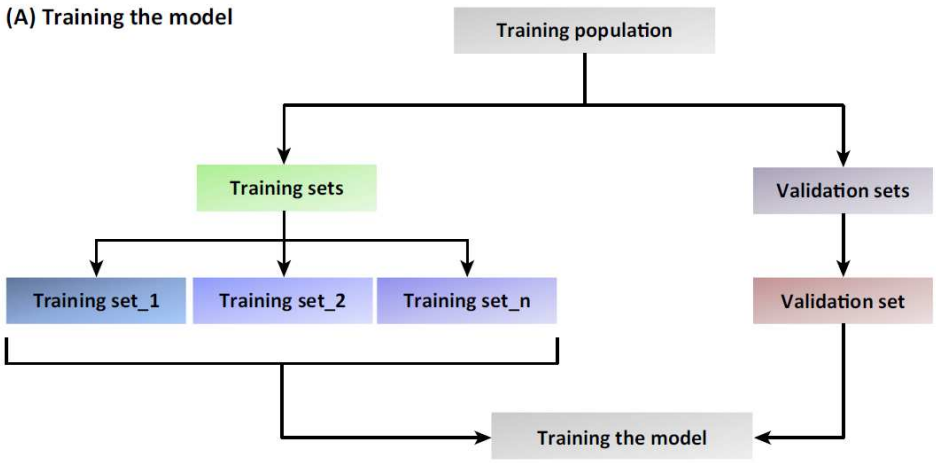


Figure 22 : Représentation des bases d'apprentissages et de test nécessaire au développement des programmes informatiques d'analyse d'images (Z. A. Desta and R. Ortiz, 2014).

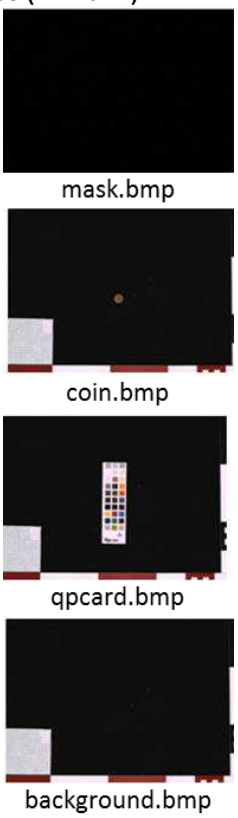


Figure 23 : Les quatre images de pré-traitement (Vilmorin).

C. Les bases de données et d'images

La base de données a été construite sous un tableur Excel et les données mentionnées correspondent aux notations, aux mesures physiques du chou-fleur mais aussi aux mesures sur images. La base de données est organisée par caractère puis par chou-fleur. Chaque chou-fleur photographié est identifié et les données relatives aux notations et aux mesures sont traçables depuis les fichiers Excel (Figure 21).

La base d'images est organisée par date et par série de prises de vues. Pour chaque série de prise de vue, quatre images de pré-traitement sont enregistrées comme illustrée sur la Figure 23. La phase de pré-traitement consiste à préparer l'image de chou-fleur à son traitement informatique (détection des régions d'intérêt). La première image correspond à une image prise lorsque l'objectif de la caméra est protégé par son capuchon. Cette image noire est utilisée pour retirer le bruit de la caméra. Une pièce puis une charte couleur (QPcard) sont prises en photographie pour réaliser respectivement une calibration spatiale et colorimétrique des images. La netteté et la luminosité de l'image sont ajustées avec la charte couleur (comme précisé plus haut). Enfin, une photographie du fond de la boite avec les objets fixés (support d'étiquettes par exemple) permet d'enlever les régions inutiles au traitement et à l'analyse des images de chou-fleur.

Des bases d'apprentissage et de test sont construites par caractère. Ces bases se composent d'images et de données représentatives de la variabilité phénotypique du caractère. Les choux-fleurs sont sélectionnés en fonction des données de vérité terrain (notes données au champ ou mesures physiques) et de la qualité des images. Les bases d'apprentissage utilisées pour la forme, le caractère pointu et le caractère lisse sont décrites en Annexe I. Les bases d'apprentissage sont utilisées pour la construction des méthodes d'analyse d'images et des algorithmes alors que les bases de test sont utilisées pour valider les méthodes d'analyse d'images et les algorithmes et pour évaluer la performance de l'outil. Les bases d'apprentissage et de test sont le plus souvent de constructions identiques (même base d'images utilisée pour le développement et la validation des méthodes d'analyse d'images et des algorithmes). Dans ce cas précis, la méthode de calcul sur images est indépendante de l'image traitée. Ce système d'apprentissage (bases d'apprentissage et de test) est assez répandu. Il est utilisé par exemple en sélection génomique pour la recherche et la validation des modèles statistiques, comme illustré Figure 22.

D. Les logiciels

Les analyses statistiques sont menées avec le logiciel libre R version 3.1.1 (R Development Core Team, 2008) et différents packages dont le package Rcmdr pour R Commander (J. Fox, 2005). Les mesures sur images sont réalisées avec le logiciel d'analyse d'images ImageJ (W. Rasband, 2012). Les outils du logiciel sont décrits dans le document de référence du logiciel (W. Rasband, 2012). Les programmes informatiques sont développés par les spécialistes en traitement d'images sous Matllab (The MathWorks Inc., 2012).

Deux logiciels d'édition et de reconnaissance d'étiquettes développés dans le cadre d'un projet VCO sont utilisés pour la renomination des images.

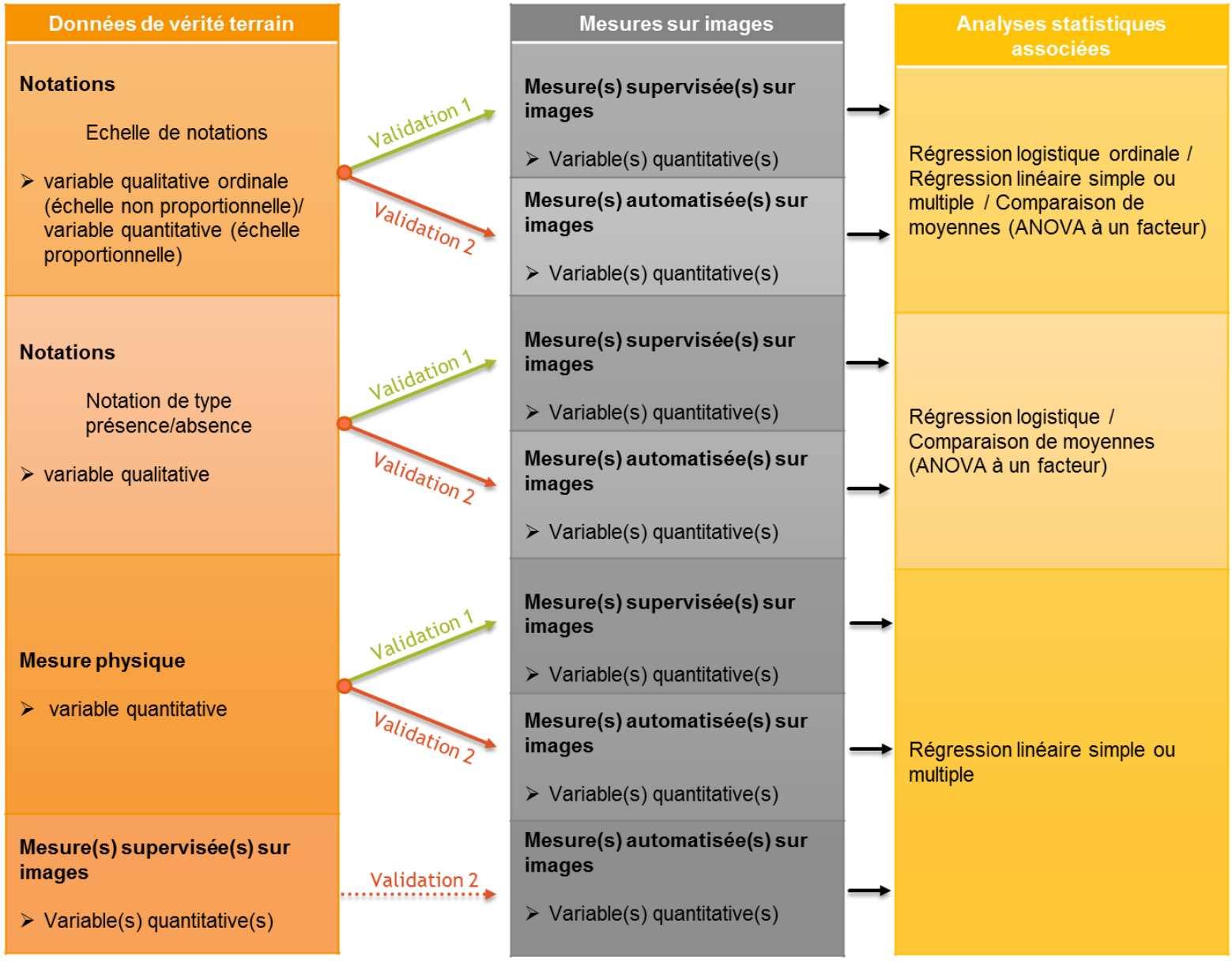


Figure 24 : Analyses statistiques pour les deux types de validations (méthodes d'analyse d'images, algorithmes).

E. Analyses statistiques

Les analyses statistiques ont pour objectif de valider dans un premier temps les méthodes d'analyse d'images (Validation 1, Figure 24) et dans un second temps les algorithmes développés (Validation 2, Figure 24). Les données « vérités terrains » correspondent aux notations, aux mesures physiques et aux mesures supervisées sur images. Pour la Validation 1, les notations, les mesures physiques sont comparées aux mesures supervisées sur images soit au travers de modèles linéaires simples soit au travers de modèles linéaires multiples lorsque plusieurs variables explicatives sont étudiées en même temps. Pour la Validation 2, les notations, les mesures physiques et les mesures supervisées sur images sont comparées aux mesures automatisées sur images, dans des modèles linéaires simples ou multiples. La comparaison des mesures supervisées et des mesures automatisées n'est pas systématique (tiret en pointillé, Figure 24). De plus, les deux validations peuvent être faites simultanément, lorsque la méthode d'analyse d'images est déjà implémentée dans un programme informatique.

En fonction de la nature des variables étudiées, c'est-à-dire s'il s'agit de variables quantitative(s) ou qualitative(s), le type d'analyses statistiques diffère. L'échelle de notations peut être considérée soit comme une variable qualitative ordinale si on admet que la proportionnalité de l'échelle de notations n'est pas vérifiée, soit comme une variable quantitative. Les mesures physiques, les mesures supervisées et les mesures automatisées sur images sont des données quantitatives. Pour des variables quantitatives, la régression linéaire simple ou multiple est utilisée. Lorsque la variable à expliquer est une variable qualitative (les variables explicatives sont quantitatives), la régression logistique est utilisée. La régression logistique est dite ordinale lorsque les catégories de la variable qualitative sont ordonnées. Une analyse de variance (ANOVA pour Analyse Of Variance) à un facteur peut également être réalisée afin de comparer les moyennes des différentes modalités du facteur.

Certaines conditions d'applications doivent être vérifiées pour exploiter les résultats des analyses statistiques. Pour la régression linéaire et l'analyse de variance à un facteur, les conditions à respecter sont la normalité des résidus, l'indépendance des résidus et des observations et l'absence de valeurs extrêmes (écart maximum de 3 entre la plus petite et la plus grande valeur des résidus normalisés). Pour l'analyse de variance à un facteur, l'égalité des variances (appelée aussi homoscédasticité) est également vérifiée. Plusieurs tests statistiques et analyses graphiques sont associés à ces vérifications. Par exemple, pour la normalité des résidus, les tests de Shapiro-Wilk et de Kolmogorov-Smirnov sont employés. Le diagramme Q-Q ainsi que l'histogramme des résidus sont également utilisés. La distribution des résidus doit suivre la courbe de Gauss (densité de la loi Normale) et les quantiles des résidus estimés doivent être proche des quantiles de la distribution normale (droite de Henry). Pour les autres postulats à vérifier, se référer à l'Annexe V.

Enfin, pour choisir la meilleure méthode d'analyse d'images, quelques critères sont utilisés pour juger de la qualité des modèles linéaires : les critères d'information bayésien ou d'Akaïké (BIC, AIC), le coefficient de détermination (R^2) et les intervalles de confiance et de prédiction pour une valeur prédite. Un pourcentage d'erreur est également calculé pour des caractères ou des seuils sur les variables brutes sont utilisés. Ces critères sont décrits en Annexe VI. Pour valider les algorithmes (mesures automatisées sur images), de la même façon, la qualité des modèles est jugée via les mêmes critères que ceux utilisés pour valider les méthodes d'analyse d'images. En plus, pour tester le caractère prédictif des modèles, des cross-validations sont réalisées. L'erreur moyenne de prédiction qui résulte de cette méthode est aussi utilisée comme critère de validation (Annexe VI). Le meilleur modèle est élu lorsque les critères BIC et AIC, les intervalles de confiance et de prédiction pour une valeur prédite, le pourcentage d'erreur, l'erreur moyenne de prédiction après une cross-validation sont minimaux et lorsque le coefficient de détermination est maximal.

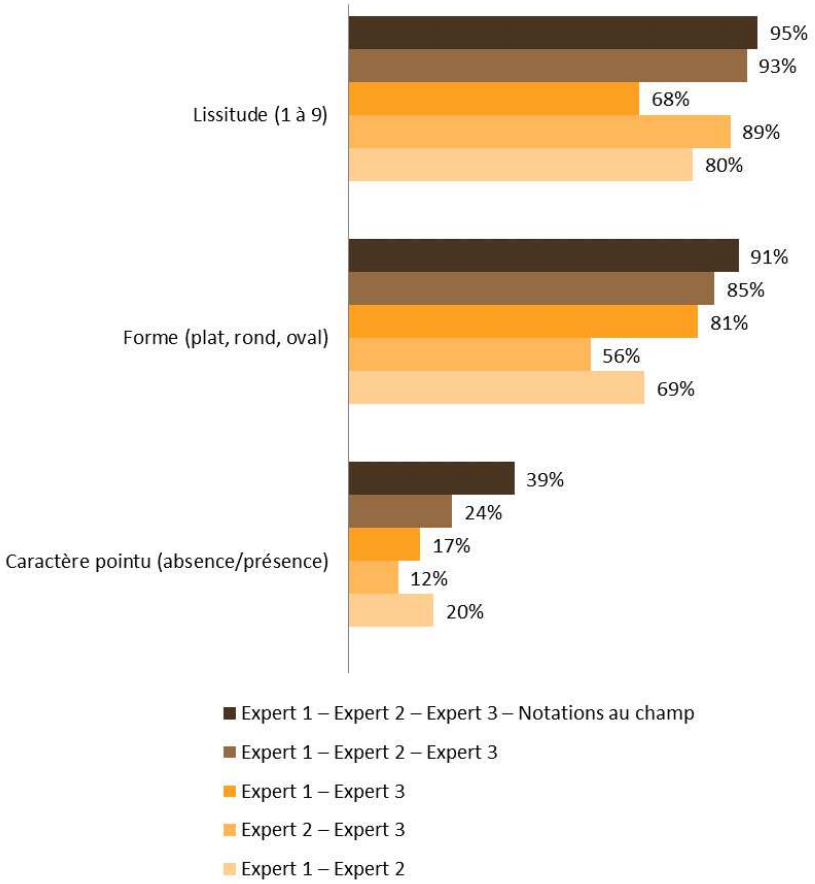


Figure 25 : Taux de discordance (exprimé en pourcentage) des notes entre experts pris 2 à 2, entre les notes de tous les experts ou entre les notes d'experts et les notes au champ, pour la forme, le caractère lisse et le caractère pointu des pommes de chou-fleur.

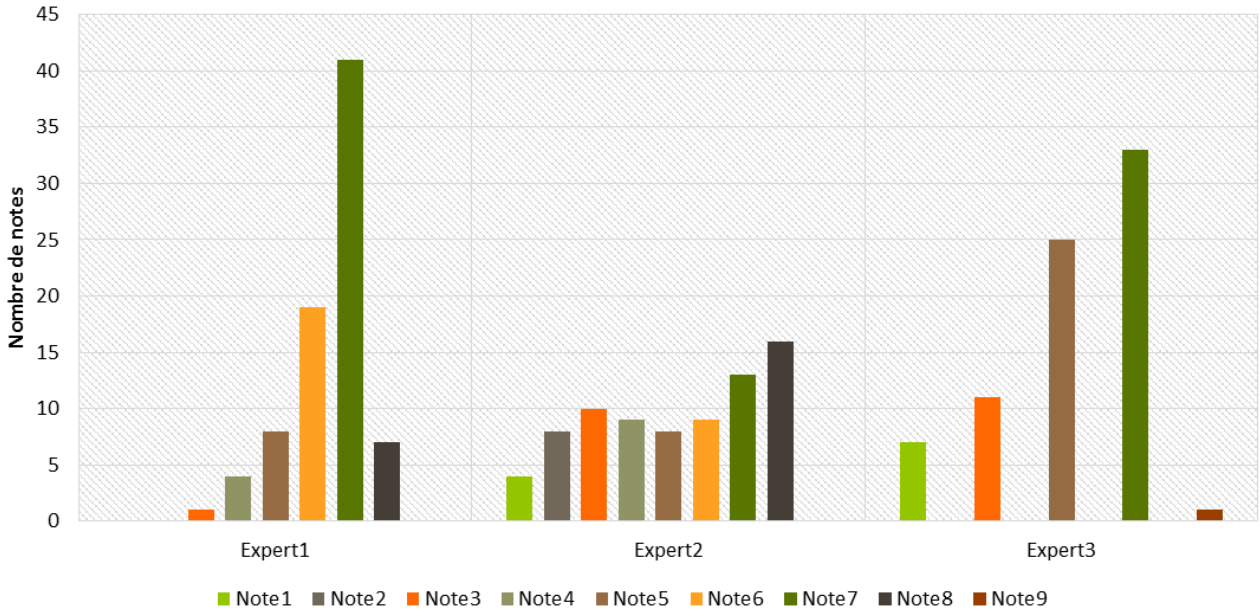


Figure 26 : Répartition des notes pour l'évaluation de la lissitude par expert.

III. Le contrôle des sources de variation du phénotypage

Dans l'évaluation phénotypique visuelle, il existe plusieurs sources de variation qui affectent le jugement de l'expert et par conséquent les notations. Certains facteurs de variation sont d'ordre physiologique, comme par exemple la perception de la lumière qui diffère d'un expert à l'autre. Il y a également le référentiel de notation utilisé qui est variable entre expert. Ce référentiel dépend des outils utilisés (abaque de couleur par exemple) mais également du matériel végétal étudié. La variabilité inter-notateur rend ainsi les notations subjectives, comme illustré ci-après.

A. La variabilité inter-notateur

Initialement destinée à harmoniser les notations des sélectionneurs, une étude de la variabilité inter-notateur pour l'évaluation phénotypique du chou-fleur n'a pas été menée au champ mais sur images pour trois caractères : la forme, le caractère lisse et le caractère pointu. Les notations sont comparées entre sélectionneurs (Expert 1, Expert 2, Expert 3) et un taux de correspondance exprimé en pourcentage est calculé pour chaque caractère entre notateurs pris deux à deux puis entre tous les notateurs (Figure 25).

Plus il y a de choix dans les notations, plus les notations divergent entre experts. Pour le caractère lisse où il y a 8 choix possibles, il y a près de 95% de dissemblance entre les notes des trois experts alors que pour la forme évaluée selon trois catégories et le caractère pointu évalué avec deux modalités, il y a respectivement 85% et 24% de réponses différentes entre les trois experts. Même si en réduisant le nombre de choix possible pour la notation, la probabilité que la réponse soit liée au hasard augmente, le pourcentage de notes différentes entre experts reste élevé.

Pour le caractère pointu, au regard des différences de notes entre experts pris 2 à 2, il y a entre 12 à 20% de pommes pour lesquelles il y a un des experts en désaccord avec les deux autres, en fonction du couple d'expert considéré. Cette erreur reste tout de même non négligeable pour une notation à deux modalités. Malgré trois possibilités pour la notation de la forme et une définition à priori commune du caractère, le pourcentage de discordance de notations entre experts pris 2 à 2 oscille entre 56% et 81%, ce qui correspond à un pourcentage d'erreur très important. Pour le caractère lisse, malgré des consignes identiques données pour mener à bien cette notation, les pourcentages de discordance de notations entre experts pris 2 à 2 sont très importants et compris entre 68% et 89%.

La variabilité inter-notateur est ainsi démontrée au travers des notations sur images. Cette étude a amené à une deuxième série de notations par les sélectionneurs et cette fois-ci de façon concertée. La subjectivité des notations au champ n'a pas pu être testée car les notations au champ n'ont pas été menées par plusieurs experts sur la même série de choux-fleurs.

B. La relativité de l'évaluation phénotypique

Les équipes de sélection du chou-fleur étudient avec plus ou moins d'importance les caractères de ce projet. De manière générale, l'échelle de notations d'un caractère est propre à chaque équipe et elle est relative à la variabilité phénotypique observée quotidiennement par l'équipe. La relativité des notations peut être illustrée par la première série de notations menées sur images de chou-fleur pour le caractère lisse. Malgré les mêmes consignes de notations données aux experts et des images de choux-fleurs dont la variabilité phénotypique du caractère observée sur images a été exploitée au maximum, la répartition des notes est très différente d'un expert à l'autre comme l'illustre la Figure 26.

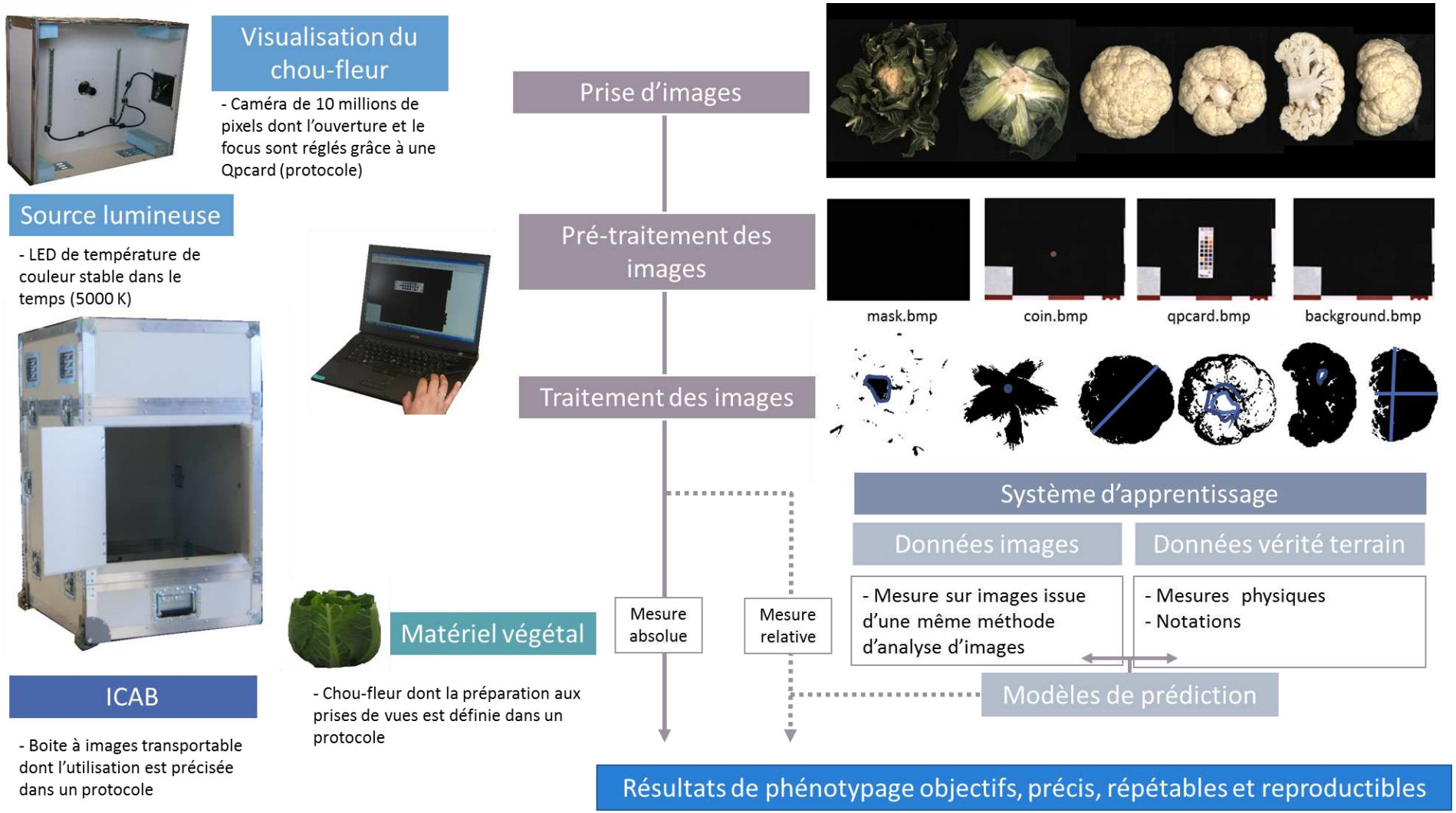


Figure 27 : Schéma illustrant les différents points de contrôles des sources de variation liées au phénotypage.

Pour le premier expert (Expert1), les notes sont comprises entre 3 et 8. Selon cet expert, il existe des pommes de chou-fleur encore plus lisses et plus bossues. Pour les deux autres experts (Expert2 et Expert3), les notes sont plus étalées. En revanche, pour le troisième expert, cinq notes sont seulement exploitées. La variation du caractère n'est pas notée de la même façon par les experts. Les notations sont relatives à une référence qui diffère d'un expert à l'autre. Ce facteur de variation participe également à la variabilité inter-expert.

Nous avons pu mettre en avant la variabilité inter-expert mais d'autres sources de variation existent et sont également dues à l'expert, comme par exemple l'humeur et l'état de fatigue qui peuvent affecter son jugement au cours de la journée. Cette source de variation rend également les notations subjectives et entre dans le cadre de la variabilité intra-notateur. Les variabilités intra-expert et inter-expert n'ont pas pu être démontrées au champ. Pourtant, au champ, les conditions lumineuses sont variables et constitue également un facteur de variation dans les notations.

L'intérêt majeur d'utiliser des systèmes de vision artificielle pour le phénotypage est le gain en fiabilité dans l'évaluation des caractères et en particulier, le gain dans la répétabilité et la reproductibilité des mesures. Les sources de variation telles que le changement de luminosité et la subjectivité des notations (variabilité intra et inter-expert) sont contrôlées par ces systèmes. La luminosité est gérée par la boite à images (ICAB), grâce à un système lumineux stable dans le temps et un réglage fin et reproductible de la caméra. Les notations sont traduites par des mesures sur images dont les méthodes d'analyse d'images sont les mêmes pour chaque image traitée. Par conséquent, la mesure automatique et répétée d'une même image renvoie des résultats identiques. La mesure est également reproductible car les manipulations du système d'acquisition et la préparation du matériel végétal sont décrits dans des protocoles. L'information est résumée Figure 27. L'apport du système de vision artificielle est également illustré au travers d'exemples donnés pour des caractères morphologiques.

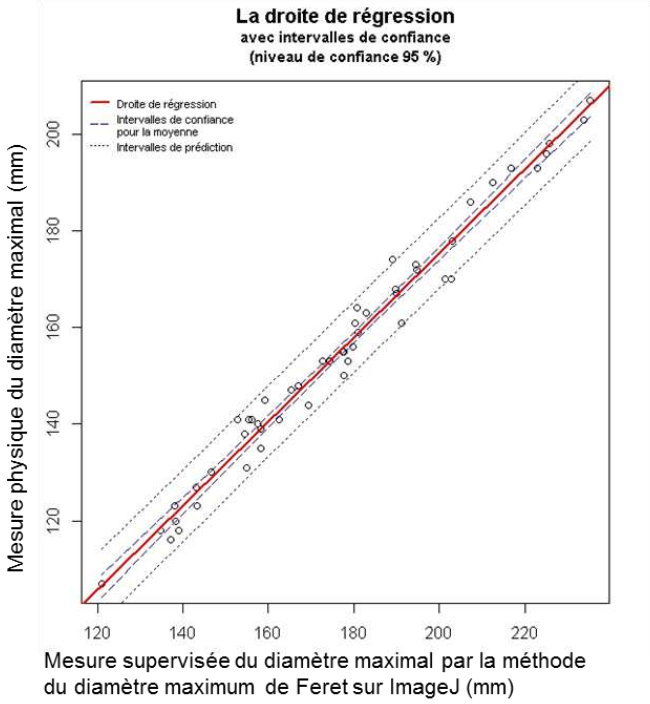
IV. Une analyse fine des caractéristiques du chou-fleur

A. Exemples de caractères morphologiques en deux dimensions

Prédiction de données quantitatives, exemple du diamètre

Le diamètre est une dimension importante pour les sélectionneurs puisque ce critère est utilisé pour définir la maturité des pommes de chou-fleur. Ce caractère, moins héritable que la masse du chou-fleur, est soumis à un effet hétérosis. Des choux-fleurs issus du croisement de deux lignées homozygotes ont un diamètre plus important que les lignées parentales. Cet effet hétérosis s'est observé notamment pour des hybrides F1 issus du croisement de choux-fleurs de typologie différente mais aussi pour des hybrides F1 issus du croisement de choux-fleurs d'hiver précoces (P. Crisp and C. R. Tapsell, 1993). L'héritabilité de ce caractère ainsi que l'effet hétérosis observé sur ce caractère justifient son étude en génétique d'association. La mesure physique fastidieuse du diamètre justifie sa mesure par imagerie.

La méthode d'analyse d'images testée est le calcul automatique du diamètre de Feret sous ImageJ. Le diamètre de Feret correspond au diamètre maximum d'une surface. Afin de valider la méthode d'analyse d'image et l'algorithme associé, les résultats sont comparés aux mesures physiques du diamètre (mètre ruban ou pied à coulisse). Pour ces deux validations, nous nous plaçons dans le cas de régressions linéaires simples. Les conditions d'applications de la régression linéaire simple sont vérifiées et validées statistiquement et graphiquement pour les deux validations. Les résultats de ces tests sont donnés en Annexe VII.



Caractéristiques du modèle

Droite de régression linéaire

Equation de la droite des moindres carrés :
Y= 0.865 X + 2.219

Qualité explicative

Coefficient de Détermination :
R^2 = 98%

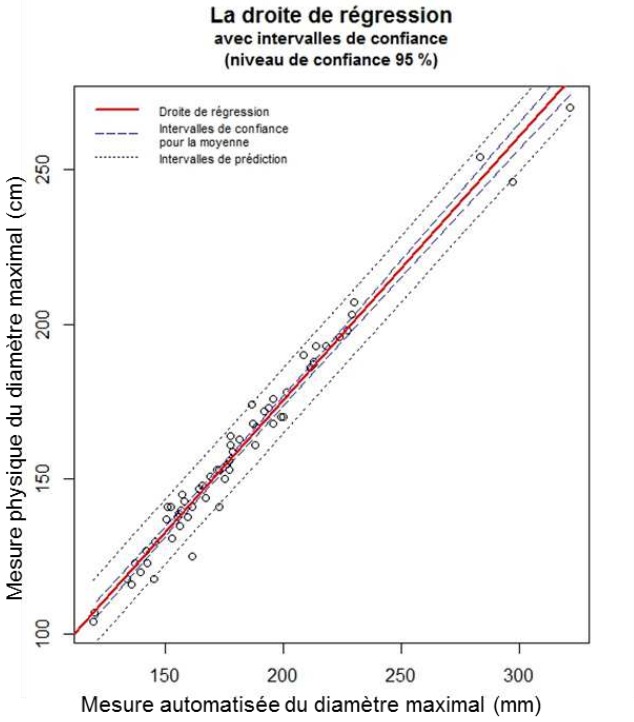
Qualité prédictive

Erreur de prédiction, pour une prédiction d'un diamètre mesuré sur images de 18 cm :

± 0,7 cm (intervalle de prédiction) soit ± 4% (par rapport à la valeur du diamètre moyen),

± 0,1 cm (intervalle de confiance)

Figure 28 : Modèle statistique de régression linéaire pour la validation de la méthode d'analyse d'images



Caractéristiques du modèle

Droite de régression linéaire

Equation de la droite des moindres carrés :
Y= 0.851 X + 5.350

Qualité explicative

Coefficient de Détermination :
R^2 = 98%

Qualité prédictive

Erreur de prédiction, pour une prédiction d'un diamètre mesuré sur images de 18 cm :

± 1 cm (intervalle de prédiction) soit ± 6% (par rapport à la valeur prédite),

± 0.1 cm (intervalle de confiance)

Erreur de prédiction après une cross-validation (10 folds) : 0.3 cm

Figure 29 : Modèle statistique de régression linéaire pour la validation de l'algorithme

Les images et les données utilisées pour l'apprentissage correspondent à 57 pommes de choux-fleurs dont le diamètre mesuré physiquement varie de 11 à 21 cm environs.

La première validation est celle de la méthode d'analyse d'images. Le coefficient de détermination met en avant le fort caractère explicatif de la variable qui correspond aux mesures sur images (R^2 = 98%). De plus, les intervalles de confiance et de prédiction d'un diamètre prédit de 18 cm (diamètre moyen) sur images illustrent également la bonne qualité prédictive du modèle, soit au maximum une erreur de prédiction de ± 0.7 cm, soit ± 4% de la valeur prédite (Figure 28). Au vue de la qualité prédictive et explicative du modèle, la méthode du diamètre maximum de Feret est validée.

A présent, la deuxième validation est la validation de l'algorithme. Le modèle de prédiction est construit sur un échantillon de 63 pommes et contient les 57 pommes étudiées pour la validation des méthodes d'analyse d'images. Les pommes ont un diamètre mesuré physiquement qui varie de 10 à 27 cm et d'un diamètre mesuré sur images de 12 à 32 cm. Ce modèle est de très bonne qualité explicative et prédictive (R^2 = 98%, erreur de prédiction de ± 1 cm, Figure 29). Les mesures sur images sont surestimées par rapport aux mesures physiques, un ajustement des valeurs est donc nécessaire pour être au plus proche des mesures terrains. Cet ajustement est mené grâce à l'équation de la droite de régression linéaire du modèle. Une validation croisée est réalisée pour étudier la qualité prédictive du modèle. L'erreur de prédiction moyenne calculée est environs de 0.3 cm. Les mêmes conclusions sont données pour la qualité du modèle de régression linéaire. L'algorithme est ainsi validé.

La mesure physique, fastidieuse et imprécise du diamètre maximum au champ est remplacée par une mesure automatisée et précise du diamètre maximum sur images.

Prédiction de données qualitatives, exemple du caractère pointu

Une pomme de chou-fleur est dite pointue lorsqu'une forme de cône est observée au sommet du chou-fleur. En coupe longitudinale, le sommet de la pomme de chou-fleur prend une forme triangulaire. La forme pointue a été étudiée sur des choux pommés (Séguin-Swartz G. et al., 1997). Il a été montré que ce caractère est dominant par rapport au caractère rond de la pomme. Son étude en génétique d'association est donc justifiée car il existe un contrôle génétique de ce caractère. La mesure subjective au champ justifie son étude par imagerie.

Les mesures supervisées en deux dimensions sur images sont le calcul de l'angle formé par les segments de droite de part et d'autre du point culminant de le pomme de chou-fleur en coupe longitudinale et la linéarité des segments de droite (régressions linéaires) de part et d'autre du point culminant. L'angle est calculé à partir des pentes des segments de droite. Différente tailles de segments sont testés, soit 50 pixels (environs 1 cm), 150 pixels (environs 4 cm), 200 pixels (environs 5 cm) et 250 pixels (environs 6 cm). Différents seuils de valeurs sont également testés en fonction de la taille des segments. Les critères de décision utilisés (valeur d'angle et valeur du coefficient de détermination) sont définis grâce à une base d'apprentissage (identique à la base de test) de 86 pommes notées sur images de façon concertée par plusieurs experts (23 images notées pointues et 63 images notées non pointues). Les pommes de chou-fleur choisies pour la notation sur images sont également variables sur la forme. La répartition du nombre de pommes en fonction de la forme notée au champ est donnée Annexe I. L'étude est menée à la fois sur la face interne et sur la face externe, puisque en fonction de la coupe, la pointe apparaît sur au moins une des deux faces. Afin de valider la méthode d'analyse d'images et l'algorithme associé, les résultats sont comparés aux notations sur images. Pour ces deux validations, nous considérerons un taux de correspondance ou un pourcentage d'erreur, avec un nombre de faux positifs et de faux négatifs. Le seuil entre les pommes pointues et non pointues est défini de façon empirique, en minimisant le pourcentage d'erreur global.

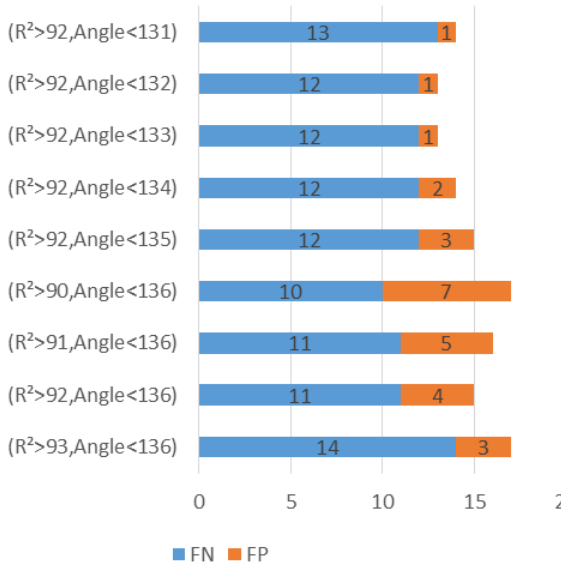


Figure 30 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 150 pixels en fonction des valeurs du R^2(90-93) et de l'angle (131-136).

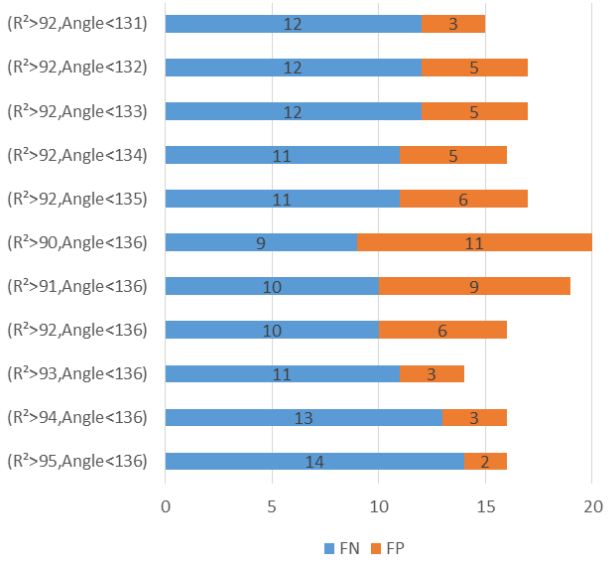


Figure 31 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 200 pixels en fonction des valeurs du R^2(90-95) et de l'angle (131-136).

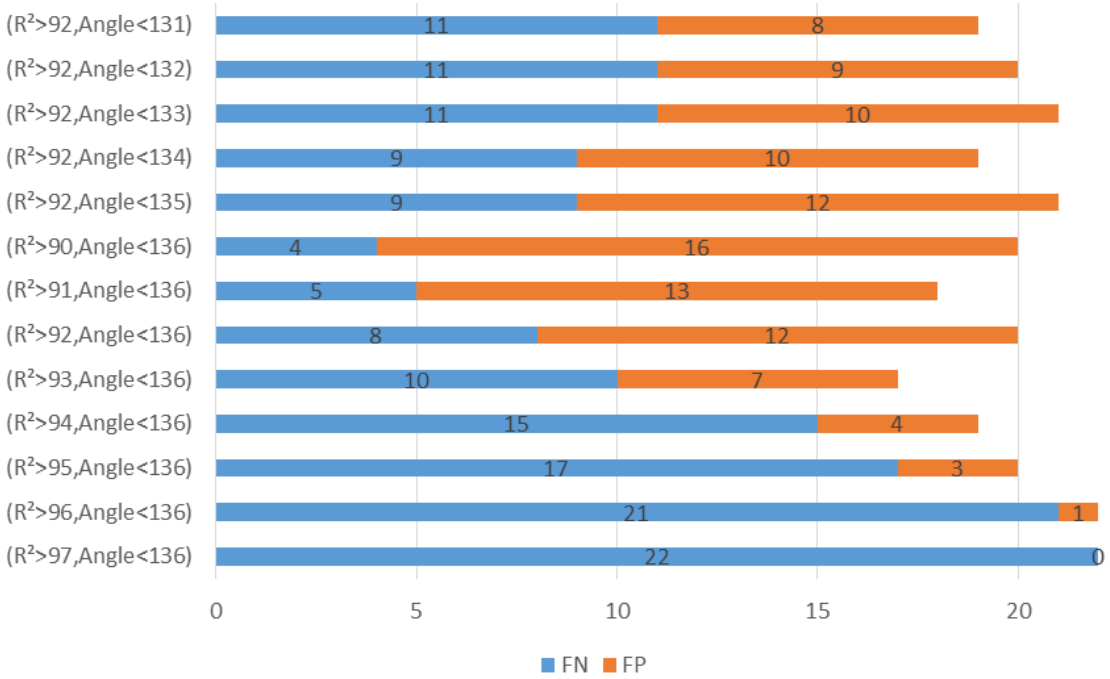


Figure 32 : Taux d'erreur (nombre de faux négatifs en bleu, nombre de faux positifs en orange) pour des segments de droite de 250 pixels en fonction des valeurs du R^2(90-97) et de l'angle (131-136)

Les valeurs d'angle et les coefficients de détermination des régressions linéaires des segments de droite sont déterminés grâce à la sélection de la partie supérieure des faces internes et externes du chou-fleur et au diamètre de Feret (voir illustration pour l'étude du caractère lisse, Figure 36). La distance maximale entre le contour de la pomme et la coupe (diamètre de Feret) identifie le point culminant. Cette méthode a été directement implémentée dans un programme informatique qui renvoie également les caractéristiques de la pointe : l'angle et la linéarité de segments de part et d'autre du point culminant. La moyenne des coefficients de détermination des régressions linéaires des deux segments de droite est utilisée dans l'analyse. La valeur minimale des quatre R^2 des segments sur les faces internes et externes, permet d'éliminer la face correspondante. Les résultats sont donnés pour chaque longueur de segment respectivement Figure 30, Figure 31 et Figure 32.

Le nombre d'erreur est minimisée en choisissant un segment de 150 pixels et lorsque les pommes sont notées pointues quand le R^2 moyen est supérieur à 92 et l'angle est inférieur à 132 (avec un taux d'erreur de 13 sur 86, soit 15% d'erreur). Il y a 12 faux négatifs et 1 faux positif. Ce qui fait 12 pommes sur 23 non reconnues comme étant pointues (52%) et 1 pomme sur 63 non reconnue comme étant non pointu (2%). Toutefois, parmi les 23 pommes notées pointues lors de la deuxième série de notations, 15 pommes de chou-fleur n'ont pas été évaluées de la même façon par les trois experts lors de la première série de notations des images. L'évaluation est donc assez restrictive et permet d'évaluer les pommes de chou-fleur ayant une pointe fortement marquée. Par comparaison au pourcentage d'erreur donnée pour l'étude de la variabilité inter-expert compris entre 12% et 20%, le pourcentage d'erreur (par référence à la deuxième série de notations sur images où les experts se sont concertés) est de 15%.

Le caractère pointu évalué visuellement est désormais évalué par imagerie avec un apport d'information sur l'angle de la pointe et la linéarité des segments de droite qui définissent la pointe. Grâce à la base d'apprentissage, l'information est également donnée sur la présence ou non d'une pointe au sommet de la pomme.

B. Exemples de caractères morphologiques en trois dimensions

Prédiction de données quantitatives, exemple du volume

Le volume de la pomme de chou-fleur est un critère étudié par les sélectionneurs car ce critère est lié à la masse volumique et donc à la densité. Une étude a été menée sur la densité de la pomme de chou-fleur, à partir du croisement de pommes denses et moins denses (loosed-curd). Il a été démontré que la densité de la pomme est sous contrôle polygénique (P. Crisp and C. R. Tapsell, 1993). Si le diamètre ainsi que la densité est contrôlé génétiquement, on suppose que le volume de la pomme est également sous contrôle génétique, ce qui justifie son étude en génétique d'association. La mesure fastidieuse du volume justifie également l'utilisation de l'imagerie pour phénotyper ce caractère.

Les mesures supervisées de volume sur images sont réalisées en testant différentes formules de volumes et en testant différentes variables telles que des dimensions ou des surfaces. Dans la littérature, de nombreux volumes de fruits et légumes sont prédits à partir de mesures en deux dimensions telles que la surface (Moreda et al., 2012), d'où la justification de les tester pour tenter de prédire le volume. La base d'apprentissage, de construction identique à la base de test, est constituée de 53 pommes dont le volume a été évalué par la méthode du déplacement d'eau. Le volume est compris entre 500 et 4370 mL. Les formules de volume sont décrites en . Il y a deux façons d'évaluer un volume, soit en utilisant les rayons, soit en utilisant les surfaces et les rayons. Pour les variables prises indépendamment, il y a les surfaces (SSH), les hauteurs (MinFeret), les diamètres ou largeurs maximales (Feret), la circularité (Circularity), la rondeur (Roundness) et l'élongation (Aspect Ratio). Les formules sont décrites en Annexe III.

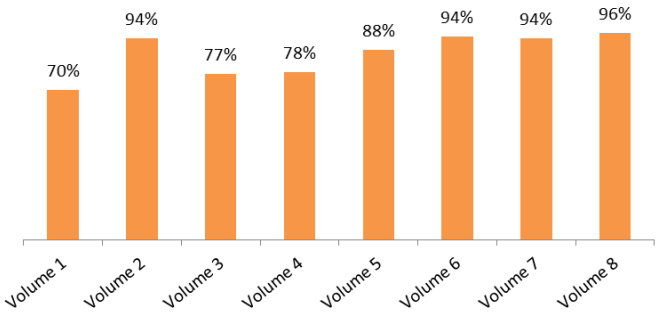


Figure 33 : Comparaison des coefficients de détermination des modèles de régression linéaires simples correspondant aux différentes formules de volume.

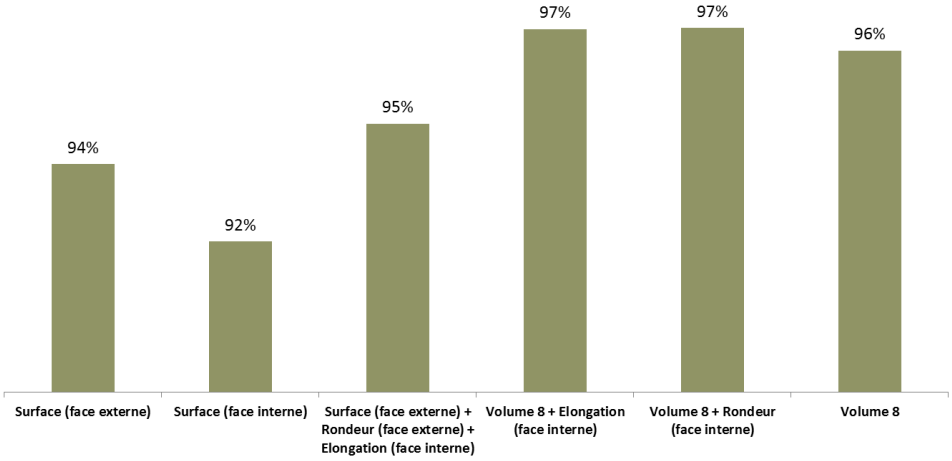


Figure 34 : Comparaison des coefficients de détermination des différents modèles de régression linéaire multiple pour l'évaluation du volume.

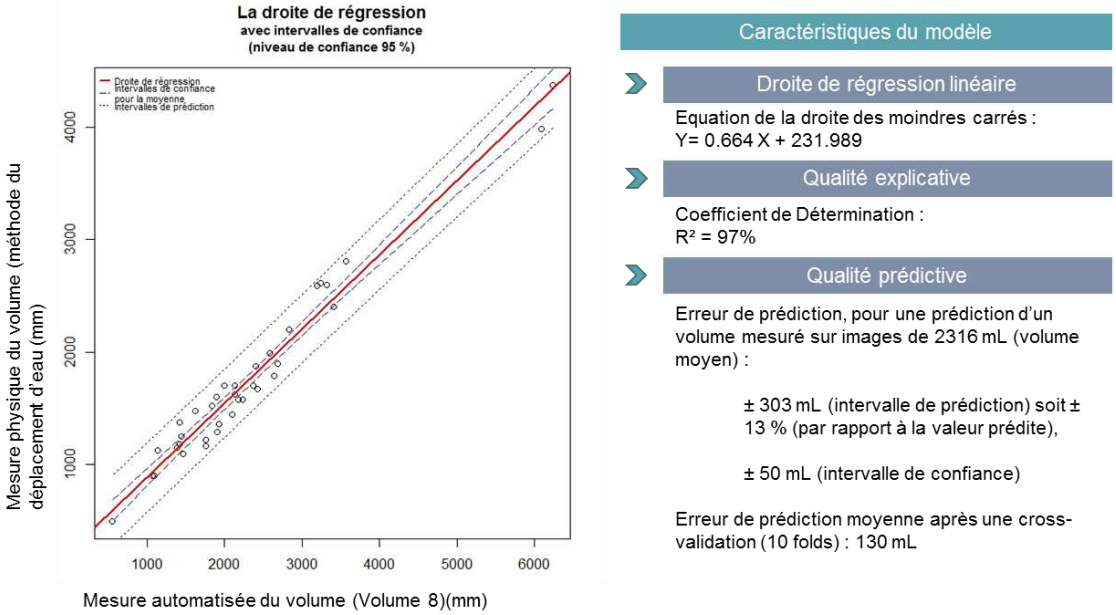


Figure 35 : Modèle de régression linéaire pour l'évaluation du volume par l'algorithme.

Pour les deux validations statistiques, des modèles de régression simples et multiples sont étudiés.

La première étude a été menée sur les formules de volume calculé sur images afin d'identifier la meilleure d'entre elles. Une comparaison du coefficient de détermination des modèles de régression linéaire correspondant aux différentes formules de volume a permis de retenir la formule relative au Volume 8, dont le modèle de régression linéaire a pour coefficient de détermination la valeur de 96% (Figure 33). Une deuxième étude a été menée pour étudier plusieurs variables explicatives, notamment les autres variables correspondant aux mesures de surface, de diamètre etc. Afin de pouvoir mener à bien cette étude, une première vérification est faite concernant l'indépendance linéaire deux à deux des variables étudiées. Un tableau est donné en Annexe IX. Cette première étude permet d'orienter la construction des différents modèles linéaires multiples. Les volumes et les surfaces sont tous corrélés entre eux. Aucun modèle ne sera construit avec plusieurs volumes, plusieurs surfaces ou avec des surfaces et des volumes. Seuls, la surface et le volume avec le coefficient de détermination le plus important seront retenus, soit le Volume 8 et la surface de la face externe d'une coupe longitudinale de chou-fleur. Les variables Rondeur, Circularité et Elongation sont ajoutés au modèle. Les modèles retenus sont illustrés Figure 34 et ont été sélectionnés pas à pas par la méthode de sélection progressive et rétrogressive (backward/forward, en anglais). La surface de la face externe d'une coupe longitudinale à elle seule explique 93 % de la variation du volume. L'ajout des variables Elongation et Rondeur ne permettent pas d'améliorer de manière très significative la qualité explicative et prédictive du modèle avec le Volume 8. Finalement, seule la formule du Volume 8 sera retenue pour la mesure automatisée et prédite du volume sur images.

La deuxième validation constitue la validation de l'algorithme. Parmi les 53 pommes étudiées, 17 pommes ont été retirées pour diverses raisons, soit la segmentation ne permettrait pas de sélectionner toute la surface du chou-fleur, soit les pommes évaluées étaient de trop mauvaise qualité pour être prises en compte dans le modèle. Le volume est compris entre 500 et 4370 mL. Les résultats du modèle de régression linéaire basée sur la formule de Volume 8 sont donnés Figure 35 et le volume mesuré par l'algorithme est compris entre 546 et 6235 mL. La qualité prédictive de ce modèle est moyenne, l'intervalle de prédiction est assez large (± 330 mL) mais l'erreur moyenne de prédiction de la validation croisée est de 6% (130mL) par rapport au volume moyen mesuré sur images (2316 mL). Ainsi, ce modèle est retenu pour la prédiction du volume.

Le volume est prédit automatiquement par analyse d'images. La mesure physique fastidieuse du volume est désormais remplacée par une mesure rapide sur images.

Prédiction de données quantitatives, exemple de la lissitude

La lissitude en surface de la pomme est définie par la variation en trois dimensions des fleurettes. La difficulté de l'évaluation phénotypique de ce caractère justifie son étude par imagerie.

Les mesures réalisées sur images sont des ratios de surfaces comme par exemple la variable Solidity (surface de la sélection/surface convexe), des ratios de périmètre (ratio de périmètre 1 : périmètre de la sélection/périmètre convexe, ratio de périmètre 2 : périmètre de la sélection/périmètre de l'ellipse ajustée), la circularité (voir Annexe III). Les vues utilisées sont la vue du dessus et la vue externe d'un chou-fleur en coupe longitudinale. Pour la vue externe en coupe longitudinale, la partie supérieure de la surface est délimitée de la même façon que pour le caractère pointu. La partie supérieure de la pomme de chou-fleur est sélectionnée puis dupliquée (A, Figure 36). Puis, les deux parties sont assemblées (B, Figure 36).

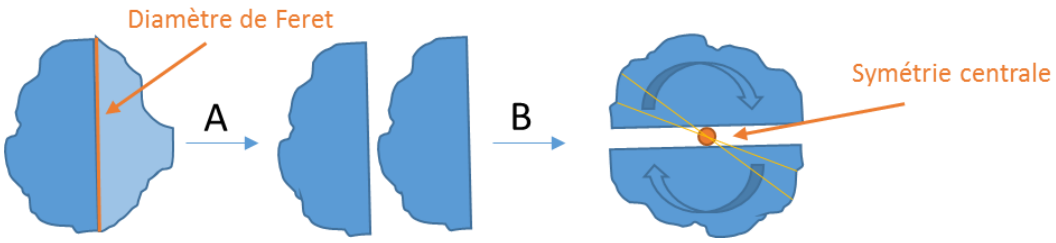


Figure 36 : Sélection de la partie supérieure d'un chou-fleur en coupe longitudinale, duplication de cette sélection (A) et construction d'une forme ronde par l'assemblage des deux parties (rotation de 180° de l'une des deux parties, puis accolement des deux parties au niveau du segment de droite, B).

Solidity	Circularité	Ratio de périmètre 1	Ratio de périmètre 2	Solidity	Ratio de périmètre 1	Solidity
18%	20%	15%	17%	22%	8%	36%

Figure 37 : Coefficient de détermination des modèles d'analyse de variance à un facteur avec en bleu les variables expliquées relatives aux mesures des images de pommes de chou-fleur observées vue du dessus, en jaune les variables expliquées relatives aux mesures des images de pommes de chou-fleur observées en coupe longitudinale (face externe) et en orange, la moyenne de la variable mesurée sur les deux vues.

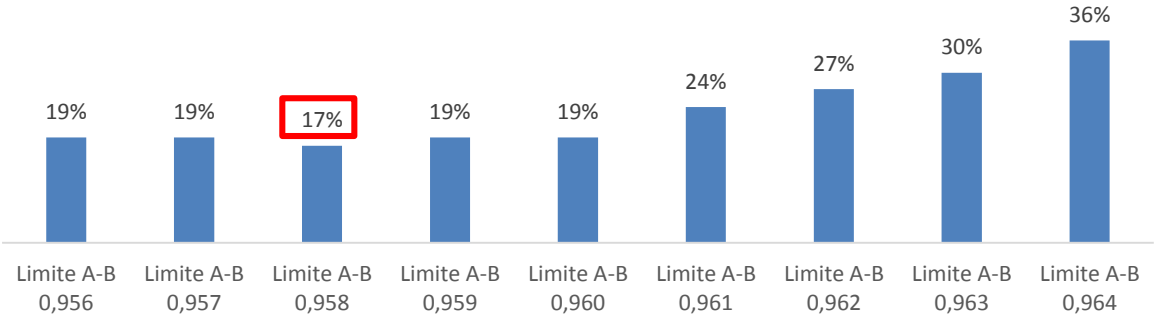


Figure 38 : Taux d'erreur en fonction de la valeur de seuil pour la limite A-B (le taux d'erreur minimal est encadré en rouge)

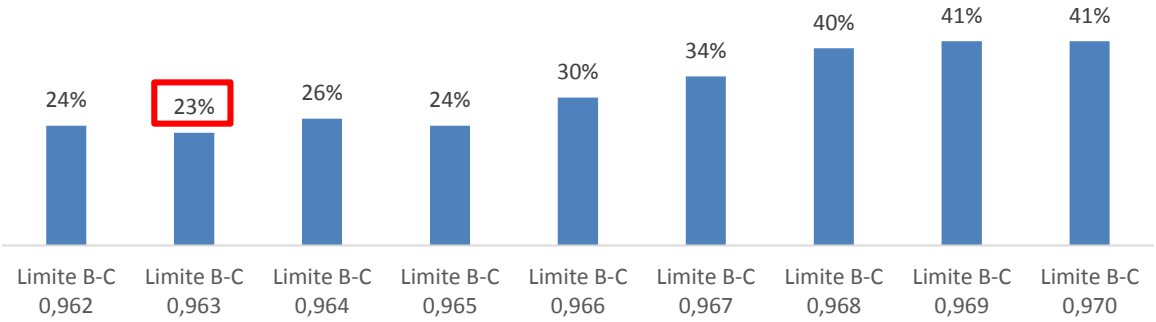


Figure 39 : Taux d'erreur en fonction de la valeur de seuil pour la limite B-C (le taux d'erreur minimal est encadré en rouge)

Une partie est tournée de 180 degrés pour être accolée à la deuxième partie au niveau de la coupe du diamètre maximum de Feret (B, Figure 36). La deuxième partie accolée est obtenue par une symétrie centrale.

Les méthodes d'analyse d'images et les algorithmes sont validés grâce aux notations réalisées de façon concertée par les sélectionneurs, sur les images de chou-fleur. La base d'apprentissage utilisée est de construction identique à la base de test. Une première base de 54 pommes de chou-fleur est utilisée pour tester les variables une à une. Les catégories représentées sont : A (4), A-B (6), B (11), B-C (25), B-C(+) (7), C (1). La catégorie A se réfère aux pommes bosselés tandis que la catégorie C se réfère aux pommes lisses. L'analyse de variance à un facteur a permis d'identifier la variable la plus expliquée par les notations, soit la moyenne de la variable Solidity calculée sur la pomme de chou-fleur observée vue du dessus et en coupe longitudinale sur la face externe, avec un coefficient de détermination de 36% (Figure 37). Les valeurs de seuil des limites entre les catégories A et B et entre les catégories B et C sont choisies en minimisant le taux d'erreur. Pour cela, les notations sur images sont transformées : les catégories A et A-B forment la catégorie A, la catégorie B reste la catégorie B et les catégories B-C(-), B-C, B-C (+) et C seront attribuées à la catégorie C.

Les valeurs de seuils sont validées lorsque le taux d'erreur est minimisé. Les taux d'erreur en fonction des seuils testés sont donnés Figure 38 et Figure 39. Les pommes de chou-fleur appartiennent à la catégorie A lorsque la valeur de la variable Solidity est inférieure à 0.958, à la catégorie B lorsque la valeur est supérieure ou égale à 0.958 et inférieure à 0.965 et à la catégorie C lorsque la valeur est supérieure ou égale à 0.965. Avec les seuils ainsi définis, le taux d'erreur est de 34%. D'autres méthodes ont également été testées, comme par exemple une Analyse en Composantes Principales, une classification par la méthode du « k-means » et une division proportionnelle en trois catégories en utilisant le minimum et maximum de la variable Solidity de la base d'apprentissage. Pour les deux dernières méthodes, les taux d'erreur sont respectivement de 37% et 39%. Pour l'Analyse en Composantes Principales, le pourcentage de variation expliquée par la composante principale qui explique le plus de variation est de 20%. Ce qui reste inférieur au coefficient de détermination pour l'ANOVA à un facteur où seule la variable Solidity (moyenne sur les deux vues) est considérée soit un R^2 égal à 36%.

L'étude de la variabilité inter-expert sur 96 pommes de chou-fleur a mis en avant un fort pourcentage d'erreur, compris entre 56% et 61%. Comparé à ces résultats, l'évaluation de la lissitude sur images est moins erronée, avec un pourcentage d'erreur de 34%. Toutefois, cette erreur reste importante pour une classification en trois catégories. La notation très subjective du caractère lisse est remplacée par une mesure par imagerie plus objective et un peu plus fiable de ce caractère.

V. Les sources de variabilité

A. L'étude de la variabilité inter-experts

L'enquête réalisée à partir des notations de trois sélectionneurs, a montré qu'il existe une variabilité importante dans l'expertise de la forme, le caractère pointu et le caractère lisse. La variabilité a été mise en avant entre experts mais également entre les évaluations au champ et sur images. En dehors du notateur, il y a d'autres facteurs qui expliquent probablement cette variabilité.

Influence de la préparation du chou-fleur

La préparation du chou-fleur aux prises de vues constitue un facteur de variation. Pour l'évaluation du caractère pointu par exemple, les choux-fleurs sont coupés longitudinalement. Si la coupe n'est pas réalisée à l'endroit précis de la pointe, le caractère pointu peut ne pas apparaitre sur les images des faces internes et externes du chou-fleur coupé longitudinalement. Ainsi, le pourcentage d'erreur calculé à partir de pommes notées pointues au champ et non pointues sur images serait en partie expliqué par ce facteur. La préparation du chou-fleur conditionne l'information présente sur les images et par conséquent les résultats issus de l'analyse des images. Les coupes sont destructives et non automatisées, le biais ainsi occasionné ne peut donc pas être corrigé. De plus, les coupes réalisées sur le chou-fleur semblent difficilement automatisables car elles sont précises et nécessitent un jugement de l'opérateur. Toutefois, la préparation est décrite dans un protocole qui vise à homogénéiser les coupes réalisées sur le chou-fleur. De plus, cette préparation permet de préparer et de sélectionner les éléments d'intérêt à analyser sur images.

L'importance de la description des caractères

Si la définition du caractère pointu semble admise et être la même pour les trois experts, la définition de la forme peut porter à confusion. Pour la forme, trois catégories sont données : oval, plat et rond. La forme plate d'une pomme de chou-fleur fait référence au rapport hauteur/largeur, c'est à dire à l'épaisseur de la pomme en coupe longitudinale. La pomme est également désignée comme une pomme ovale à l'horizontal. Or, il existe déjà une catégorie qui fait référence au caractère oval de la pomme mais cette fois-ci à la verticale. Les notations au champ n'ont pas été menées par les sélectionneurs mais par des personnes moins expérimentées. Les deux catégories ont pu être confondues au cours des notations, expliquant ainsi les différences importantes entre les notes au champ et sur images. Pour brouiller encore plus la définition de ce caractère, les sélectionneurs considèrent qu'il existe aussi des pommes rondes aplaties. Par conséquent, les pommes considérées rondes sont des pommes dont le rapport hauteur/largeur est plus important que les pommes plates et que ce critère suffit à priori à distinguer ces deux catégories. La forme ne semble ici ne pas faire référence à la courbure ni au contour de la pomme de chou-fleur mais plutôt aux dimensions de la pomme de chou-fleur. La description du caractère ainsi que les termes choisis pour différencier les pommes en catégories sont très importants, surtout lorsque des mesures sont fondées sur ces explications.

Interprétation et décomposition des caractères

Il a été montré, au travers de la répartition des notes du caractère plus ou moins lisse de la surface de la pomme de chou-fleur, que la variabilité des caractères n'était pas interprétée de la même façon par les experts. Ce caractère évalué en trois dimensions, est un caractère complexe qui peut intuitivement être décomposé en sous-caractères. En effet, il y a le nombre d'irrégularités observées en surface de la pomme et l'intensité des irrégularités. Toutefois, ce niveau de précision n'est pas forcément intégré ni pris en compte de façon consciente par les experts, ce qui rend la notation encore plus subjective et difficilement comparable d'un expert à l'autre.

La difficulté est de pouvoir décrypter ces caractères complexes, de trouver les bonnes variables et la bonne combinaison de variables pour pouvoir être au plus proche de leur notation. Bien évidemment, il y a une limite dans cette exercice, puisque l'évaluation humaine sur images ou au champ est également source d'erreur, comme l'étude de la variabilité inter-experts a pu le démontrer.

La notation sur images

Les sélectionneurs ont dus notés sur images certains caractères du chou-fleur afin d'harmoniser leur notation. Cependant, ils ne sont pas habitués à noter sur des images et ce manque d'expérience a sans doute influencer l'évaluation. De plus, l'information apportée par les images est en deux dimensions et chaque vue du chou-fleur apporte une information différente. Même si pour des caractères qui font appel aux trois dimensions du chou-fleur, plusieurs vues sont données afin de reconstituer en quelque sorte ces trois dimensions, l'information est figée et décomposée. De ce fait, les sélectionneurs ont été influencés dans l'évaluation et les notations par conséquent ont été assez différentes de ce qui a été noté au champ. De plus, les écrans sur lesquels les images ont été projetées n'ont probablement pas été paramétrés et il y a de forte chance que la luminosité ait été différente d'un écran à l'autre. Toutefois, grâce au paramétrage de la caméra et au contrôle des conditions lumineuses dans la boite, les images sont paramétrées, ce qui les rend comparables. Enfin, la notation est relative car les images de choux-fleurs peuvent être comparées à l'écran. De ce fait, les conditions dans lesquelles ont été notées les images n'ont à priori pas du trop influencé la notation.

B. Le matériel végétal et la variabilité des caractères

La maturité des pommes de chou-fleur

L'évaluation de la maturité des plantes est déterminante puisque c'est grâce à ce critère que les plantes sont récoltées et analysées par imagerie. Elle est visuelle et établie essentiellement par les dimensions de la pomme de chou-fleur qui sont déterminées par le marché. La taille de la pomme de chou-fleur évolue au cours du temps car l'inflorescence hypertrophiée est un stade intermédiaire, mais elle se stabilise avant que les fleurettes « débranchent ». La pomme de chou-fleur est considérée hors maturité lorsque les dimensions ne sont plus respectées et lorsque les fleurettes commencent à se séparer du fait de la croissance des primordiaux floraux (Studman, 2001). Toutefois, dans certains cas, la bonne taille de pomme ne coïncide pas toujours avec le stade physiologique où la pomme de chou-fleur est stable. Lorsque ce stade physiologique est dépassé, la qualité du chou-fleur se dégrade assez rapidement, ce qui rend le chou-fleur non évaluable par imagerie. De grosses irrégularités se forment autour de la pomme de chou-fleur, les fleurettes ne sont plus compactes et d'autres caractères sont accentués comme par exemple les colorations roses et jaunes qui se répandent en surface de la pomme de chou-fleur.

Les bases d'apprentissage et les données relatives

Les bases d'apprentissages sont constituées d'un échantillon de pommes de choux-fleurs à priori représentatives de la variabilité des caractères. Toutefois, la répartition des pommes est telle que toute l'échelle de variation n'est pas forcément représentée. Grâce à ces bases d'apprentissage, des ajustements ou des transformations sont réalisés sur les résultats bruts renvoyés par les algorithmes. En effet, dans certains cas, les mesures sont surestimées par rapport aux mesures de vérité terrain (liée à la surélévation progressive du chou-fleur en fonction de son volume) ou encore des données quantitatives sont transformées en données qualitatives. Ces modèles d'ajustement sont construits sur les bases d'apprentissage et sont valables pour une variabilité donnée. Or, il s'agit d'une référence et non de la référence.

C'est pourquoi, dans certains cas, les données brutes sont tout de même fournies avec les données ajustées/transformées, pour que l'ajustement ou la transformation puisse être modifiée si possible par les utilisateurs, en fonction des besoins. Cependant, un système d'apprentissage devra également être utilisé pour modifier celui utilisé par défaut dans le logiciel. La flexibilité du système d'apprentissage permet d'adapter l'outil aux besoins de son utilisateur. C'est ce que propose l'entreprise VISAR au travers de leur trieuse optique de carottes VACS (Visar's Automatic Carrot Sorter) où le système d'apprentissage est également flexible (VISAR, 2014). La sévérité du tri peut être ajustée par l'opérateur.

C. Les réglages de la caméra et l'utilisation de chartes couleurs

Réglages de l'ouverture et du focus

Les réglages de l'ouverture et du focus sont des réglages difficiles et non automatisées. Pourtant, ces deux réglages conditionnent les résultats donnés par les algorithmes. Un mauvais réglage du focus aboutit à une image floue dont la segmentation peut être affectée. Même si les mesures colorimétriques se font relativement aux patchs d'une QPcard, les mesures absolues de couleur par exemple ne pourront pas être exploitées si l'image est surexposée par exemple. De plus, la surexposition est problématique pour le chou-fleur puisque les tons de blanc sont mesurés. Une pomme surexposée peut être confondue à une pomme très blanche. La surexposition retire de l'information au chou-fleur puisque les zones saturées ne donnent pas d'information par définition. D'ailleurs, la luminosité a dû être spécifiquement ajustée pour le chou-fleur de façon à ce que les images soient légèrement sous-exposées mais que le chou-fleur à hauteur ne soit pas surexposé. Pour conserver les bons réglages d'une manipulation à l'autre, les molettes de réglage pourraient être fixées à l'aide des vis, en prenant soin de garder la même hauteur de caméra pour chaque série de prise de vues. La sensibilité des mollettes de réglage ainsi que l'absence de graduation sur l'appareil rend la manipulation délicate.

Les chartes couleurs

La charte couleur QPcard utilisée pour les prises de vues est périssable. La couleur des patchs se dégradent au cours du temps notamment si les patchs sont frottés ou laissés constamment à la lumière du jour. La QPcard est également sensible à l'humidité, ce qui la déforme et la gondole. La charte couleur développée par le CTIFL est également soumise aux mêmes dégradations que la QPcard. Ce sont des références colorimétriques et par conséquent, elles sont très importantes pour l'évaluation des caractères qui attraient à la colorimétrie. Les chartes couleurs doivent être donc renouvelées régulièrement (une fois/an).

VI. L'outil en tant qu'instrument de mesure

A. Le coût et le débit de l'outil

Grâce aux coupes réalisées sur le chou-fleur et aux différentes prises de vues, le système de vision artificielle utilisé pour ce projet de recherche est très peu mécanisé. Par conséquent, à l'achat, ce système est à faible coût par comparaison à des machines qui intègrent un lavage ou un tri. Toutefois, la main d'œuvre requise pour utiliser le système de vision artificielle développé par l'équipe Vision Artificielle est très importante, ce qui rend in fine, ce système couteux par rapport à un phénotypage traditionnel.

L'outil développé permet de phénotyper une trentaine de pommes de chou-fleur par jour, par personne et par outil, ce qui est assez faible par comparaison à des machines qui peuvent débiter plusieurs milliers d'objets à la seconde. Pour ces raisons, cette machine ne sera probablement pas utilisée en routine par les sélectionneurs.

Toutefois, si toutes les mesures physiques devaient être faites manuellement, il y aurait encore moins de choux-fleurs traités par jour. Pour pallier à cette contrainte, plusieurs machines tourneront en même temps pour l'essai à venir. Cet outil a été conçu pour un objectif bien défini, celui de pouvoir phénotyper du chou-fleur de façon objective, précise, reproductible et répétable. De plus, le nombre de critères analysés par imagerie fait de cette machine, une machine unique et performante.

B. Les mesures par imagerie

Les mesures par imagerie sont reproductibles et répétables grâce aux méthodes d'analyse d'images. Toutefois, les algorithmes ne peuvent pas détecter toutes les anomalies présentes sur les images qui pourraient être facilement identifiées à l'œil. Par exemple, si des fleurettes de chou-fleur présentes sur le fond noir s'accolent à la pomme de chou-fleur, la segmentation à elle seule ne permettra pas de les retirer de la sélection. Par conséquent, la mesure de la surface risque d'être surestimée. Cet exemple illustre les limites de l'outil, c'est-à-dire lorsque l'on se place dans des conditions différentes de celles qui ont été utilisés pour développer l'outil.

Cet outil génère un grand nombre de données et d'images et présente l'avantage de sauvegarder toutes ces informations. Ainsi, les données sauvegardées sont potentiellement utilisables pour l'analyse future de caractères non étudiés dans le cadre du projet, comme par exemple la couleur du feuillage.

C. La précision de l'évaluation

Pour les caractères pointus et lisses, la classification est faite respectivement en deux et trois groupes par un simple seuillage. Pour le caractère pointu, le taux d'exactitude est de 83% et pour le caractère lisse, le taux est de 67%. Pour le caractère pointu, la classification est plutôt fiable et proche de la classification visuelle. En revanche, pour le caractère lisse, la classification est moins fiable. La mesure sur images n'est pas fortement corrélée aux notations visuelles du caractère lisse (30 %), d'une part parce que c'est un caractère complexe et d'autre part parce qu'il est difficilement évalué par les experts. Par comparaison à des méthodes de classification par simple seuillage, le taux d'exactitude de la classification de la fraise sur la base de la conicité, du champignon sur la base de la rondeur, du tamarin sur la base de la courbure et de la tomate dont les caractéristiques de forme utilisés ne sont pas spécifiés par imagerie sont respectivement de 98.6 % pour trois classes, 78.0% pour deux classes, 96.0% pour trois classes et 97.9% pour deux classes (Moreda et al., 2012). En ce qui concerne la mesure du diamètre maximum (utilisée pour le tri automatisé du chou-fleur, Moreda et al., 2012) et du volume, les coefficients de détermination sont importants et sont respectivement de 98% et 97%, soit des coefficients de corrélation respectivement égaux à 0.99 et 0.98, autour des mêmes valeurs que du coefficient de corrélation du modèle linéaire construit à partir des mesures manuelles et par analyse d'image de la longueur du cornichon (J. W. van Eck et al., 1998). La précision du phénotypage est déterminante pour les analyses génétiques, puisqu'elle conditionne les résultats de l'analyse de détection de QTL. Par exemple, en passant d'un phénotypage traditionnel de type qualitatif à une évaluation quantitative, on ouvre le champ des possibles en accédant aux analyses quantitatives pour la détection de QTL.

D. Quelques préconisations d'emploi de l'outil à des fins d'analyse génétique

L'évaluation phénotypique doit être interprétée de manière plus large, c'est à dire en fonction des conditions expérimentales. Le phénotype est l'expression du génotype mais également de l'environnement et de l'interaction génotype-environnement. L'essai de phénotypage des hybrides et des lignées pour l'analyse génétique aura lieu dans un seul environnement. Les résultats de phénotypage et d'analyse génétique sont donc à considérer sur cet environnement.

C'est un risque pris puisque le chou-fleur est une sous-espèce particulièrement dépendante des conditions environnementales dont certains caractères comme la couleur du feuillage par exemple, dépendent de l'interaction génotype-environnement, et notamment de la condition azotée (M. Rogolini, 2006). De plus, la sélection du matériel végétal pour l'étude en génétique d'association a été essentiellement portée sur la précocité. D'ailleurs, la population étudiée est essentiellement structurée par ce critère. Tous les caractères corrélés à la précocité ne pourront donc pas être étudiés.

VII. L'avenir de tels outils pour le phénotypage des espèces potagères

A. Utilisation de technologies encore plus innovantes

L'évaluation de la qualité externe des fruits et légumes par des systèmes de vision artificielle est assez répandu. Aujourd'hui, la qualité interne des fruits et légumes est évaluée par des méthodes non destructives et différentes de celles utilisés dans le visible et en 2D (proche infrarouge, imagerie multispectrale et hyperspectrale, fluorescence, rayons X, imagerie thermique, la résonnance magnétique) comme par exemple pour l'étude des dommages internes (Opara and Pathare, 2014), de la maturité et de la teneur en sucre, des maladies, des désordres physiologiques, etc. (Ruiz-Altisent et al., 2010).

B. Les outils de phénotypage à haut débit, des outils d'avenir

Pour les espèces potagères, la productivité n'a pas autant d'importance que pour les espèces de grande culture, où le rendement constitue un des critères les plus sélectionnés notamment pour les trois principales céréales que sont le blé, le riz et le maïs. Les critères recherchés pour les fruits et légumes sont plutôt axés sur la qualité du produit, ne nécessitant pas forcément d'intégrer un phénotypage à l'échelle de la plante. Alors que pour les espèces de grandes cultures, la mauvaise caractérisation des milieux et aussi le manque de caractérisation au cours du temps ont poussé le phénotypage à se développer à l'échelle de la plante, de façon non destructive, au champ ou en conditions contrôlées (J. Cobb et al., 2013). Le développement de la plante est désormais prédit par des modèles qui tiennent compte des paramètres de l'environnement comme par exemple pour le chou-fleur (Olesen and Grevsen, 2000). Il pourrait être envisagé par exemple d'étudier la maturité des pommes de chou-fleur sur plantes entières, avec des systèmes de vision artificielle mobiles et tractés au champ, comme ce qui existe déjà pour le phénotypage du blé avec la « Phénomobile », un projet de collaboration entre l'INRA et Arvalis – Institut du Végétal. Cette machine permet « d'acquérir des données en continu de l'état des cultures et ainsi d'en évaluer la cinétique tout au long de la croissance» (ARVALIS, 2014).

Que ce soit pour les espèces potagères ou des espèces de grande culture, des idéotypes sont définis. La sélection assistée par marqueurs et la sélection génomique permettront d'atteindre un jour ces objectifs : sélectionner des plantes en contrôlant la variation allélique de tous les gènes d'importance agronomique, tel est le concept « Breeding by design », (J. D. Peleman and J. R. van der Voort, 2003).

Conclusion

Un nouvel outil unique a été conçu pour les sélectionneurs et pour la sélection du chou-fleur dans le cadre d'un projet de génétique d'association. L'outil de phénotypage assisté par imagerie et développé dans le cadre de ce projet est capable de phénotyper 14 caractères du chou-fleur et a pour objectif de fournir d'autres outils, des marqueurs moléculaires.

Le développement de l'outil de phénotypage a nécessité une collecte importante d'images et de données (notations, mesures physiques et mesures sur images). Cette collecte, dépendante de la période de production et de la disponibilité du matériel végétal, s'est étalée sur deux ans. La deuxième phase de développement a été essentiellement consacrée au développement et à la validation des algorithmes. La recherche des méthodes d'analyse d'images pour la construction des algorithmes est menée en parallèle des autres projets du pôle de Vision Artificielle. Certaines méthodes peuvent ainsi être exploitées pour le phénotypage par imagerie d'autres espèces potagères comme la carotte par exemple, mais également pour le phénotypage d'autre matériel végétal (pollen, graines...).

Cet outil unique réalise des mesures objectives et précises pour un grand nombre de caractères. Au vu de ce nombre, seuls quatre caractères morphologiques ont été traités dans les résultats. Le diamètre et le volume sont des caractéristiques qui font appel aux dimensions de la pomme du chou-fleur. Ils sont évalués par imagerie avec précision avec un coefficient de détermination de 98% et 97% respectivement. Pour les caractère pointu et lisse de la pomme de chou-fleur, la classification par simple seuillage aboutit à une mesure plus objective et un peu plus précise. La mesure du caractère lisse par imagerie permet une analyse du caractère pointu avec un taux d'erreur de 15% et une analyse du caractère lisse avec un taux d'erreur de 34%. Comparée à la variabilité inter-expert (plus de 50%), ces mesures sont plus fiables.

Malgré un faible débit, cet outil unique réalise des mesures objectives et précises pour un grand nombre de caractères. L'intérêt d'utiliser l'imagerie pour phénotyper le chou-fleur dans le cadre de ce projet est de permettre une analyse plus fine des caractères en génétique. Pour certains caractères, la détection de QTL peut ainsi être menée sur des données quantitatives.

Grâce aux marqueurs moléculaires qui seront développés dans le cadre de ce projet et grâce à l'expertise du sélectionneur, le matériel végétal pourra être sélectionné de façon plus rapide et plus précise. Ils permettront également de répondre aux objectifs de Vilmorin, celui de sélectionner plus précocement les têtes de lignées, de sélectionner du matériel végétal pour le marché de l'industrie mais également de sélectionner du matériel végétal pour le marché du cycle court.

Bibliographie

Alonso-Blanco C. and Méndez-Vigo B. (2014). Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Current Opinion in Plant Biology*, 18, pp. 37–43.

Bachelor B. G. (1985). Lighting and viewing techniques in automated visual inspection. IFS Publication Ltd, Bedford, UK.

Ballard D.A. and Brown C.M. (1982). Computer vision. Prentice-Hall, Englewood Cliffs, NJ, USA.

Baxes G. A. (1994). Digital Image Processing: Principles and Applications. John Wiley & Sons Inc, New York, USA.

Beecher C. W. (1994). Cancer preventive properties of varieties of Brassica oleracea: a review. *The American journal of clinical nutrition*, 59, pp. 1166–1170.

Bizais E. (2013). Business Plan chou-fleur 2012-2013.

Bizais E. (2014). Communication personelle.

Brosnan T. and Sun D.-W. (2004). Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering*, 61, pp. 3–16.

Cobb J., Declerck G., Greenberg A., Clark R. and McCouch S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and applied genetics*, 126, pp. 867–887.

Crisp P. and Tapsell C. R. (1993). 9 - Cauliflower: Brassica oleracea L. In: Genetic Improvement of Vegetable Crops (ed. by G. Kalloo and B. O. Bergh), Pergamon, Amsterdam. pp. 157–178.

Doré C. and Varoquaux F. (2006). Le chou Brassica oleracea L. partim (2n=2x=18) (Brassicacées ou Crucifères). In: Histoire et amélioration de cinquante plantes cultivées Editions Quae, pp. 229–253.

Drapper S. R. and Keefe P. D. (1989). Machine vision for the characterization and identification of cultivars. *Plant Varieties and Seeds*, 2, pp. 53–62.

Drouet D. (2014). Communication personelle.

Fox J. (2005). The R Commander: A Basic Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14, pp. 1–42.

Furbank R.T. and Tester M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16, pp. 635–644.

Gautier L. (2014). Communication personnelle.

Hayashi S., Kanuma T., Ganno K., and Sakaue O. (1998). Cabbage head recognition and size estimation for development of a selective harvester.

Jacob Y. and Lefort M. (2013). Le chou-fleur en 2012-2013. pp. 8.

Kapoor K. S. (1999). Fungal and Bacterial Diseases of Crucifers. In: Diseases of Horticultural Crops: Vegetables, Ornamentals, and Mushrooms Indus Publishing, p. 210.

Koike S.T., Gladders P. and Paulus A.O. (2007). Diseases of vegetable crops - Brassicaceae. In: Vegetable Diseases: A Color Handbook Gulf Professional Publishing, pp. 155–198.

Kolopp J., Desvignes A., Tinturier R., and Vergne F. (2013). F1 - Limagrain et le chou-fleur. pp. 10–13.

Kondo N. (2010). Automation on fruit and vegetable grading system and food traceability. *Trends in Food Science & Technology*, 21, pp. 145–152.

Mac Leod P. and Strigler F. (2009). Métrologie sensorielle. In: Évaluation sensorielle – Manuel méthodologique (3e éd.) Lavoisier, pp. 35–48.

Moreda G.P., Muñoz M.A., Ruiz-Altisent M., and Perdigones A. (2012). Shape determination of horticultural produce using two-dimensional computer vision – A review. *Journal of Food Engineering*, 108, pp. 245–261.

Olesen J.E. and Grevsen K. (2000). A simulation model of climate effects on plant productivity and variability in cauliflower (Brassica oleracea L. botrytis). *Scientia Horticulturae*, 83, pp. 83–107.

Opara U.L. and Pathare P.B. (2014). Bruise damage measurement and analysis of fresh horticultural produce—A review. *Postharvest Biology and Technology*, 91, pp. 9–24.

Pandit R. B., Tang J., Liu F. and Pitts M. (2007). Development of a novel approach to determine heating pattern using computer vision and chemical marker (M-2) yield. *Journal of food engineering*, 78, pp. 522–528.

Patel K.K., Kar A., Jha S.N. and Khan M.A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. *Journal of Food Science and Technology*, 49, pp. 123–141.

Peleman J. D. and van der Voort J. R. (2003). Breeding by Design. *Trends in Plant Science*, 8, pp. 330–334.

Polder G., Blokker G., and van der Heijden G. W. A. M. (2012). An ImageJ plugin for plant variety testing. In: Proceedings of the ImageJ User and Developer Conference 2012, pp.168–173.

Rasband W. (2012). ImageJ.

Rogolini M. (2006). Caractérisation de la réponse écophysiologique du chou-fleur (Brassica oleracea var. botrytis L.) à une carence temporaire en azote. Application à la mise en évidence de différences variétales. Agrocampus Ouest, Rennes.

Ruiz-Altisent M., Ruiz-Garcia L., Moreda G.P., Lu R., Hernandez-Sanchez N., Correa E.C., Diezma B., Nicolaï B., and García-Ramos J. (2010). Sensors for product characterization and quality of specialty crops—A review. *Computers and Electronics in Agriculture*, 74, pp. 176–194.

Schmidt R. and Bancroft I. (2011). Genetics and Genomics of the Brassicaceae - Springer.

Séguin-Swartz G.,Warwick S. I., and Scarth R. (1997). Cruciferae : Compendium of trait genetics. Agriculture and Agri-Food Canada.

Studman C.J. (2001). Computers and electronics in postharvest technology — a review. *Computers and Electronics in Agriculture*, 30, pp. 109–124.

Tao Y. and Wen Z. (1999). An adaptive spherical image transform for high-speed fruit defect detection. *Transaction of the ASAE*, 42, pp. 241–246.

The MathWorks Inc. (2012). MATLAB and Statistics Toolbox Release. United States.

Timmermans A.J.M. (1998). Computer vision system for online sorting of pot plants based on learning techniques. *Acta Horticulturae*, 421, pp. 91–98.

Tjeertes P. (2004a). Brassica oleracea L. (chou-fleur et brocoli). In: Ressources végétales de l'Afrique Tropicale 2. Legumes (ed. by G. J. H. Grubben and O.A. Denton), Fondation PROTA, Backhuys Publishers, CTA, Pays-Bas. pp. 154–159.

Torres C. (2010). Phénotypage de végétaux par classification spatio-colorimétrique d'images couleur. Université d'Angers.

van Eck J. W., van der Heijden G. W. A. M. and Polder G. (1998). Accurate Measurement of Size and Shape of Cucumber Fruits with Image Analysis. *Journal of Agricultural Engineering Research*, 70, pp. 335–343.

Vilmorin (2013). Dossier de presse 2013.

Wang H.-H. and Sun D.-W. (2002). Correlation between Cheese Meltability Determined with a Computer Vision Method and with Arnott and Schreiber Tests. *Journal of Food Science*, 67, pp. 745–749.

Watada A.E. (1980). Quality evaluation of horticultural crops: the problem. *HortScience*, 15, pp. 47–48.

Zhang B., Huang W., Li J., Zhao C., Fan S., Wu J., and Liu C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International*, 62, pp. 326–343.

Sitographie

Agreste (2009). Alimentation - Consommation alimentaire. http://agreste.agriculture.gouv.fr/IMG/file/Gaf09p075-079(1).pdf (consulté le 21 mai 2014).

Aprifel Aprifel - Agence fruits et légumes frais : fiches nutritionnelles par produits - Le tableau de composition moyenne. http://www.aprifel.com/fiches,produits.php?p=73&c=3 (consulté le 21 mai 2014).

ARVALIS (2014). La Phénomobile : un véhicule pour caractériser l'état des cultures en continu. http://www.arvalis-infos.fr:80/view-14586-arvarticle.html (consulté le 21 août 2014).

CTIFL (1991). Code couleur chou-fleur. http://www.ctifl.fr/Pages/Kiosque/DetailsOuvrage.aspx?idouvrage=630 (consulté le 22 mai 2014).

Ellips B. V. (2014). Our products. http://www.ellips.nl/our-products (consulté le 23 mai 2014).

Fachaux (2013). Calibreuses Fachaux | Fabricant de calibreuse de qualité et de précision. http://www.cal-fachaux.com/fr (consulté le 23 mai 2014).

Ferreira T. et Rasband W. (2012). ImageJ User Guide - IJ 1.46r. http://imagej.nih.gov/ij/docs/guide/146.html (consulté le 30 aout 2014).

Food and Agriculture Organization of the United Nations (2014). FAOSTAT Domains - Production - Cultures. http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/F (consulté le 21 mai 2014).

INTERFEL, Règles de qualité spécifiques aux fruits, légumes et pommes de terre. http://www.interfelbio.com/documents/gpem_da.pdf (consulté le 23 mai 2014).

Pacific Northwest Plant Disease Handbook (2014). Cabbage and Cauliflower (Brassica sp.)-Clubroot. http://pnwhandbooks.org/plantdisease/sites/default/files/imagecache/image_lightbox/images/253.jpg (consulté le 21 mai 2014).

R Development Core Team (2008). R: A language and environment for statistical computing. URL http://www.R-project.org (consulté le 28 aout 2014).

Tjeertes P. (2004b). Brassica oleracea L. (chou-fleur et brocoli), PROTA 2: Vegetables/Légumes. http://database.prota.org/PROTAhtml/Brassica%20oleracea%20(cauliflower%20and%20broccoli)_Fr.htm (consulté le 21 mai 2014).

VISAR (2014). Tri des carottes : VACS 2.0. http://www.visar.be/FR/carottes.html (consulté le 28 aout 2014).

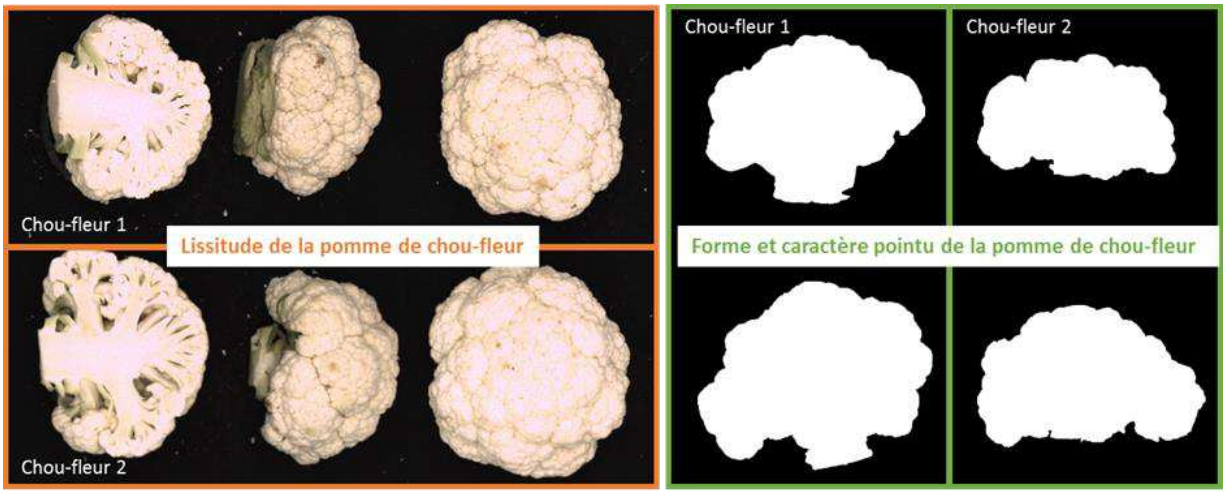


Figure 40 : Images de chou-fleur utilisées pour la notation de la lissitude, la forme et le caractère pointu de la pomme de chou-fleur (source Vilmorin).

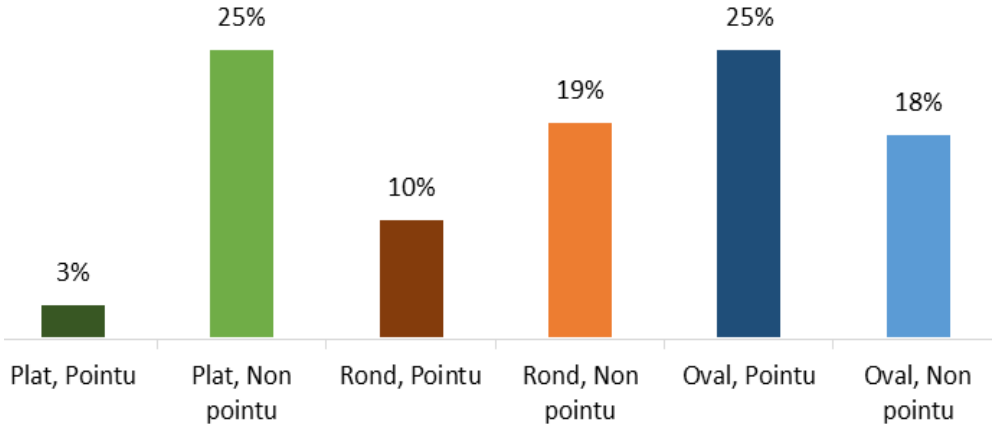


Figure 41 : Répartition des pommes prélevées en fonction des notations au champ de la forme et du caractère pointu.

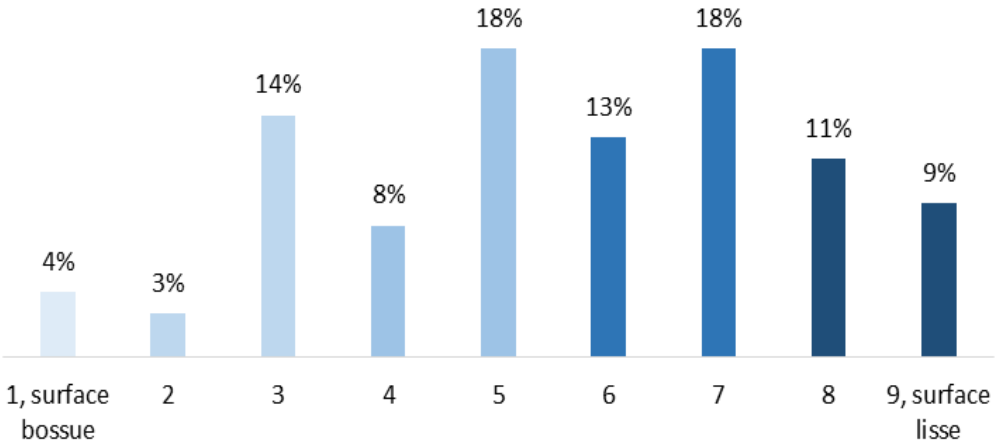


Figure 42 : Répartition des pommes prélevées en fonction des notations au champ de la lissitude.

Annexe I : Les notations sur images et les bases d'apprentissage pour la forme et les caractères pointu et lisse.

Les images utilisées pour la notation du caractère pointu, de la forme et du caractère lisse sont illustrées Figure 40.

Le même lot d'images est utilisé pour la notation de la forme et du caractère pointu. La répartition des pommes est donnée en fonction des notes au champ (Figure 41 et Figure 42).

Deux vues sont données pour la notation du caractère lisse de la pomme. La vue du dessus permet de mettre en avant la variation de la dernière couronne de fleurettes et les vues en coupe longitudinale permettent de visualiser la variation en surface des fleurettes coupées depuis la dernière couronne jusqu'au sommet de la pomme de chou-fleur. Pour la forme et le caractère pointu, des masques binaires sont utilisés. Ils permettent aux notateurs de s'affranchir de la structure interne des pommes de chou-fleur et de ne prendre en compte que la surface et les contours des faces interne et externe.

Figure 43 : Colorimétre « CR-400 » Konica Minolta Sensing, INC (Vilmorin)

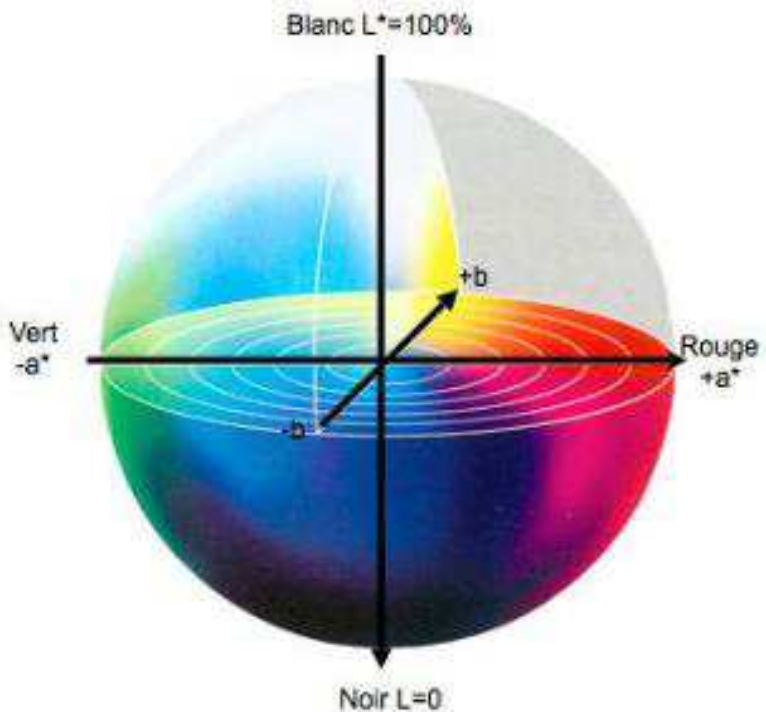


Figure 44 : Espace couleur L*a*b* ou CIELAB

Annexe II : Description du colorimètre utilisé dans le cadre du projet.

Le colorimètre « CR-400 » de marque Konica Minolta Sensing, INC est un instrument de mesure de la couleur (Figure 43).

La mesure de la couleur est réalisée sous conditions contrôlées. Une source lumineuse est incorporée dans l'appareil et permet à l'utilisateur de se placer dans les conditions standards de l'illuminant D65 (défini par la Commission Internationale de l'Eclairage) et qui correspond à la moyenne de la lumière du jour (spectre du visible et de l'ultra-violet) pour une température de couleur de 6504 K. La couleur peut être décomposée en trois paramètres : la teinte, la clarté et la saturation. Grâce à cette décomposition, la CIE a développée différents espaces couleurs, c'est-à-dire des méthodes pour exprimer la couleur d'un objet ou d'une source lumineuse au moyen d'une notation spécifique, telle que des chiffres. L'espace colorimétrique utilisé par le « CR-400 » est l'espace CIELAB (L*a*b*). L* représente « l'axe des niveaux de gris » ou encore la clarté d'une couleur. Les composante a et b sont des composantes de chromacité (Figure 44). La composante a varie entre le vert et le rouge tandis que la composante b représente les variations entre le bleu et le jaune. Dans la bibliographie, des différences de couleur du chou-fleur ont été étudiées grâce à cet espace colorimétrique (Volden J. et al., 2009).

Volden J., Bengtsson G. B. et Wicklund T. (2012). Glucosinolates, L-ascorbic acid, total phenols, anthocyanins, antioxidant capacities and colour in cauliflower (Brassica oleracea L. ssp. botrytis); effects of long-term freezer storage. *Food Chemistry*, 112, pp. 967-976.

Annexe III : <u>Formules des descripteurs de forme sous ImageJ.</u>

Caractéristiques géométriques :

- ***Circularité ou Circularity***

$$= 4 \times \pi \times \frac{Surface}{Périmètre^{2}}$$

- ***Elongation ou Aspect Ratio***

$$\frac{\text{Axe majeur de l'ellipse ajustée}}{\text{Axe mineur de l'ellipse ajustée}}$$

- ***Rondeur ou Roundness***

$$= 4 \times \frac{\text{Surface}}{\pi \times \text{Axe majeur ellipse ajustée}^{2}}$$

- ***Solidity***

$$= \frac{\text{Surface de la sélection}}{\text{Surface convexe de la sélection}}$$

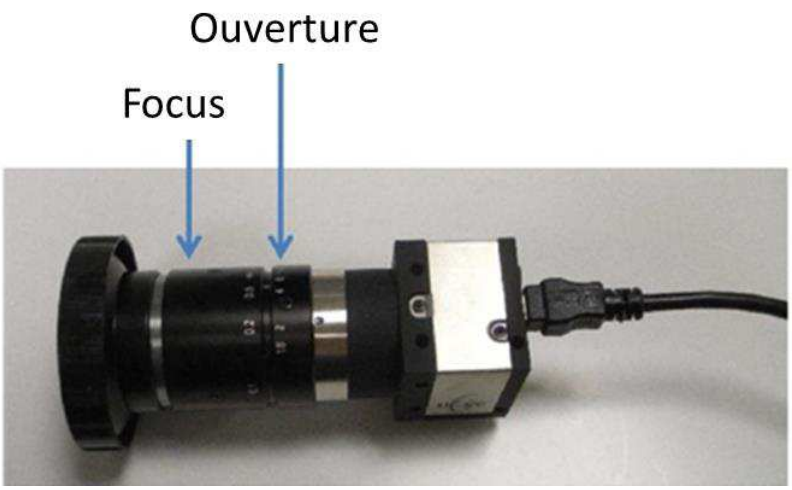


Figure 45 : Bagues de réglage de la camera : le focus et l'ouverture.

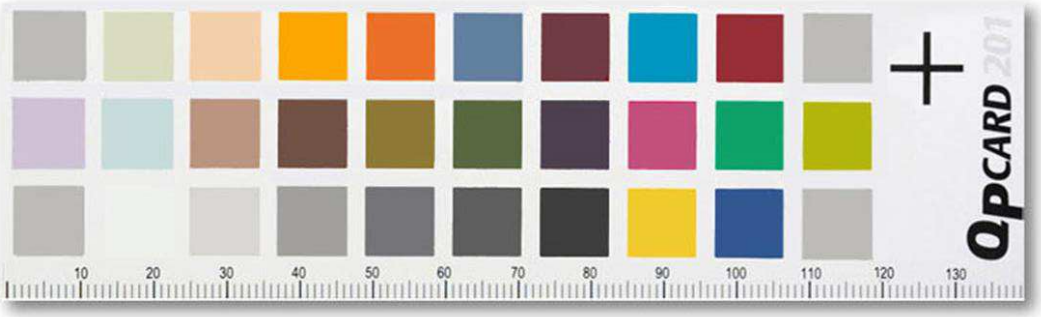


Figure 46: Image de QPcard avec un bon ajustement de l'ouverture.

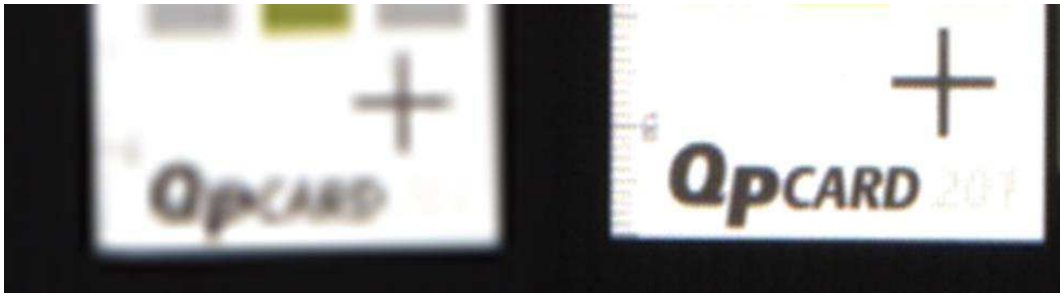


Figure 47: Ajustement correct et incorrect de la netteté de l'image (respectivement à droite et à gauche)

Annexe IV : Ajustement de l'ouverture et du focus de la caméra.

Les molettes de réglages du focus et de l'ouverture

L'étape la plus critique et la plus longue est celle qui correspond à l'ajustement du focus et de l'ouverture de la caméra, qui permettent d'ajuster respectivement la netteté et la luminosité de l'image. Deux mollettes de réglages sont localisées sur la caméra, une molette pour l'ouverture et une mollette pour le focus (Figure 45). La localisation de ces deux mollettes dépend de la caméra. Sur la bague de réglage du focus, des valeurs sont inscrites et comprises entre 0.1 et ∞. Pour la bague de réglage de l'ouverture, des valeurs sont inscrites et comprises entre 1 et 12, pour une caméra avec une ouverture de 12 mm par exemple.

Réglage du focus et de la netteté des images

1. Positionner une QPCard (Figure 46) centrée sur le champ de vision de la caméra.
2. Zoomer grâce au logiciel de la caméra sur la croix de la QPcard. Utiliser les barres de déroulement pour se positionner sur la croix de la QPcard.
3. Dans un second temps et si besoin est, utiliser un zoom supplémentaire. Pour zoomer, cliquez sur l'icône de la barre d'outils du programme uEye Demo qui correspond au Magnifying glass tool. Maintenir la touche « Ctrl » et faire un clic droit sur la croix noire de la QPcard. Une croix verte apparait à l'écran et une fenêtre apparaît également avec une image pixélisée de cette zone zoomée.
4. Tourner généreusement la mollette de réglage du focus de droite à gauche et de gauche à droite, la netteté de l'image varie à l'écran (fenêtre uEye Demo). Une plage de netteté est observée et l'objectif étant de s'en approcher au maximum. Pour cela, ajuster la netteté grâce à la croix noire zoomée de la QPcard. Les pixels de la croix noire doivent être le plus foncé possible. La Figure 47 donne un exemple d'ajustement correct et incorrect.

Réglage de l'ouverture et de la luminosité des images

L'ajustement est mené également avec la QPcard centrée sur l'image.

1. Ajuster la luminosité de l'image grâce à la bague de réglage de l'ouverture en regardant à l'écran la QPcard zoomée. Les patchs clairs doivent être visibles à l'écran.
2. Une fois ce premier ajustement fait, enregistrer l'image avec la QPcard**.**
3. Utiliser le logiciel d'analyse d'images ImageJ pour contrôler la luminosité de l'image en mesurant les valeurs RVB des patchs gris placés aux extrémités de la QPcard (Valeurs RVB de référence : 250, 250, 250).

Pour le réglage de l'ouverture, il y a deux choses à vérifier par rapport aux valeurs RVB de référence :

1. L'égalité des valeurs entre le Rouge, le Vert et le Bleu (écart maximum de 5 toléré). L'écart de valeurs est ajusté grâce aux gains RVB, réglables par l'intermédiaire du logiciel de la caméra.
2. Les valeurs en tant que telles qui doivent être toutes proches des valeurs de référence du patch gris clair soit une valeur de 180. Le réglage ici est réalisé soit manuellement en ajustant la bague de l'ouverture, soit en réglant le temps d'exposition depuis le logiciel de la caméra.

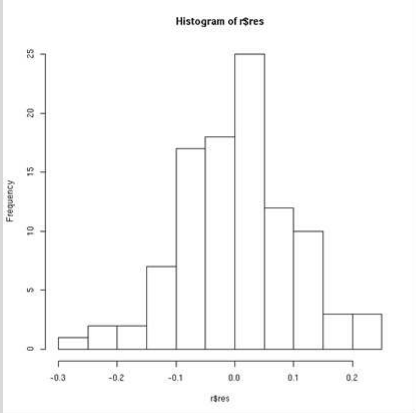
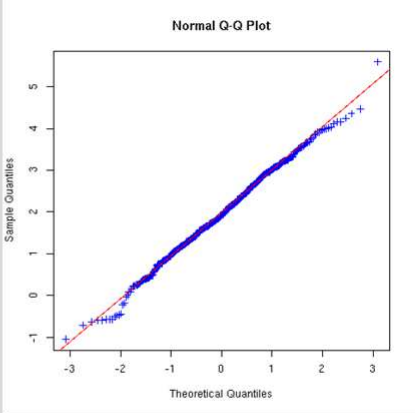
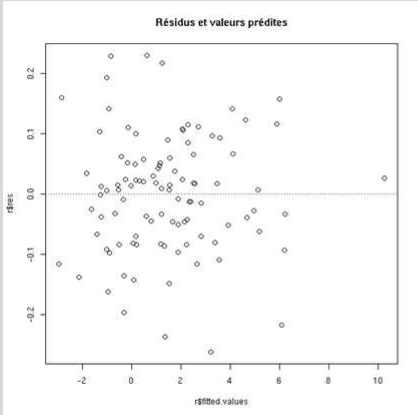
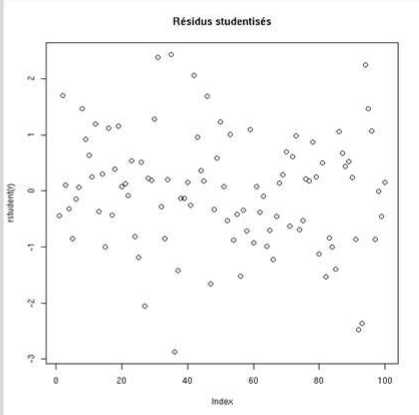
Conditions à vérifier	Tests statistiques et analyses graphiques	Illustrations
Normalité des résidus	Histogramme des résidus (Graphe 1) Normal Quantile-Quantile plot (quantiles des résidus estimés comparés à l'espérance des mêmes quantiles sous l'hypothèse de la normalité, Graphe 2) Test de Shapiro-Wilk et test de Kolmogorov-Smirnov	 Graphe 1  Graphe 2
Homoscédasticité (égalité des variances)	Test de Levene	
Indépencance des résidus	Graphe des résidus en fonction des observations (Graphe 3) Test de Durbin Watson (auto-corrélation des résidus)	 Graphe 3  Graphe 4
Indépendance des observations	Déterminé par la façon dont est menée l'expérimentation	
Valeurs extrêmes	Résidus normalisés en fonction des valeurs prédites (Graphe 4)	
Indépendance des variables explicatives	Test de corrélation	

Figure 48 : Tests statistiques et analyses graphiques pour la vérification des conditions d'applications des modèles statistiques.

Annexe V : Vérification des conditions d'applications des modèles statistiques.

La normalité des résidus

La normalité des résidus doit être testée pour tout modèle linéaire. L'histogramme des résidus doit avoir une allure gaussienne et pour le diagramme Q-Q, les points doivent être proches de la droite de Henry. La normalité des résidus est également vérifier au travers de tests statistiques. Une p-valeur est renvoyée pour les tests de Shapiro-Wilk et de Kolmogorov-Smirnov et pour valider l'hypothèse de la normalité des résidus, cette valeur doit être supérieure à 0.05, pour un risque de première espèce de 5%. Il est admis que l'analyse de variance à un facteur est assez robuste à la non-normalité des résidus (Acutis M. et al., 2012).

Acutis M. Scaglia B. et Confalonieri R. (2012). Perfunctory analysis of variance in agronomy, and its consequences in experimental results interpretation. *Europ. J. Agronomy*, 43, pp. 129-135.

L'homoscédasticité des résidus

L'homoscédasticité c'est-à-dire l'égalité des variances doit être vérifiée dans le cas d'une analyse de variance à un facteur. La variance des résidus par modalité du facteur doit être égale. Un des tests statistiques qui permet de mettre en avant l'hypothèse d'égalité des variances est le test de Levene. L'hypothèse est acceptée quand la p-valeur est supérieure à 0.05, pour un risque de première espèce de 5%.

L'indépendance des résidus et auto-corrélation des résidus

L'indépendance des résidus et l'auto-corrélation des résidus sont testées pour tous les modèles linéaires. Les résidus sont considérés indépendants lorsqu'ils sont répartis de façon homogène en fonction des valeurs prédites. Les résidus doivent être répartis et contenus dans un rectangle et ne doivent pas présenter de structure graphique particulière comme une forme d'entonnoir par exemple. L'auto-corrélation des résidus est testée au travers du test de Durbin Watson. Les résidus ne doivent pas être autocorrélés et cette hypothèse est validée lorsque la p-valeur est supérieure à 0.05, pour un risque de première espèce de 5%.

L'indépendance des observations

Egalement testé dans le cas de modèles linéaires, les observations doivent être indépendantes les unes des autres. Cette hypothèse ne peut être vérifiée que par les conditions expérimentales.

Identification des valeurs extrêmes

Les valeurs extrêmes sont identifiées au travers du graphe des résidus normalisé en fonction des valeurs prédites. Si la valeur du résidu n'est pas comprise dans l'intervalle de valeurs [-1.5 ; 1.5], il s'agit d'une valeur extrême, c'est-à-dire qu'il s'agit d'une donnée dont la valeur est assez éloignée de la valeur ajustée, prédite.

L'indépendance linéaire des variables explicatives, dans le cas de modèles linéaires multiples

Un test de corrélation permet de mettre en avant les relations linéaires d'ordre 1 entre les variables explicatives.

Un résumé est donné Figure 48.

Annexe VI : Description des critères d'évaluation de la qualité des modèles de prédiction.

- *Coefficient de détermination (R^2)*

Le coefficient de détermination est le coefficient de corrélation élevé au carré et correspond au rapport de la variance expliquée sur la variance totale. Soit la formule suivante :

$R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$, avec $\hat{y}_i$ la valeur prédite, $\bar{y}$ la moyenne des Y et y_i la valeur de Y.

- *Pourcentage d'erreur, pourcentage d'inexactitude*

Le pourcentage d'erreur ou d'inexactitude correspond au nombre de réponses discordantes sur le nombre de réponses totales pour la comparaison de deux variables qualitatives.

- *Critères de vraisemblance pénalisée : Le critère AIC d'Akaïké et le critère BIC de Shwartz*

Le critère AIC correspond à la formule suivante : AIC = $-2\times \ln\left(L(\hat{\theta})\right) + 2k$, où k est le nombre de paramètres du modèle, $L(\hat{\theta})$ correspond à la vraisemblance calculé en $(\hat{\theta})$ (elle permet de calculer l'adéquation du modèle aux données c'est-à-dire la probabilité d'avoir observé l'échantillon sous le modèle). Ici, est utilisée la log-vraisemblance, soit la fonction logarithme de la fonction de vraisemblance $L(\hat{\theta})$. La log-vraisemblance est maximale lorsque le nombre de paramètres est maximal. Les critères AIC et BIC vont pénaliser la log-vraisemblance pour tenir compte du nombre de paramètres. Les valeurs minimales des critères AIC et BIC définissent les meilleurs modèles.

Le critère BIC correspond à la formule suivante : BIC= $-2\times \ln L(\hat{\theta}) + \ln(n)k$, avec n le nombre d'observations. Ces critères permettent de comparer des modèles avec un nombre de descripteurs différents. AIC est plutôt un critère explicatif alors que le critère BIC est plutôt un critère prédictif.

- *Intervalles de confiance et de prédiction d'une valeur prédite*

<u>Droite de régression pour une valeur x^* donnée :</u>

$\hat{y}_i = \bar{y} + \widehat{\beta_1}(x^* - \bar{x})$,avec $\widehat{\beta_1}$ la pente du modèle linéaire (droite des moindres carrés), x^* la valeur à prédire, $\bar{x}$ la moyenne des valeurs de la variable X.

<u>Intervalle de confiance pour une valeur x^* donnée :</u>

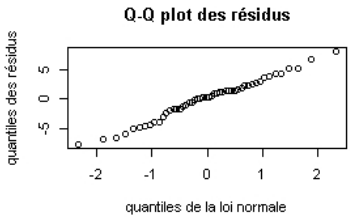
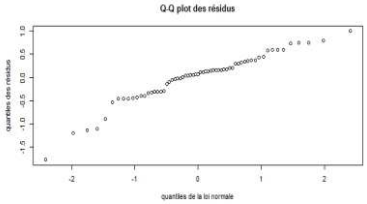
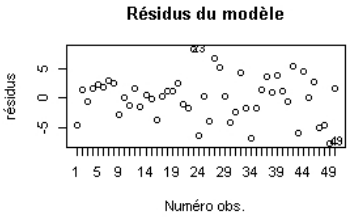
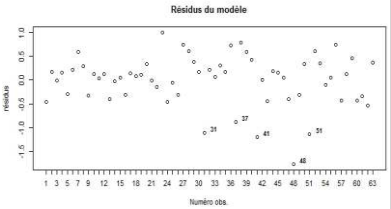
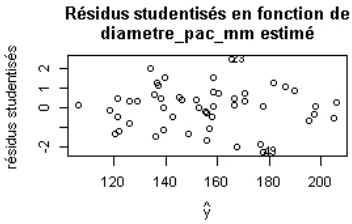
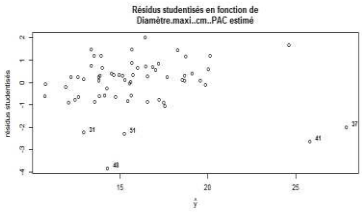
$$IC \pm t_{n-2,1-(\frac{\alpha}{2})} \times \left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}\right)^{0.5} \times \left[\frac{1}{n} + \frac{\left((x^* - \bar{x})^2\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^{0.5}$$

<u>Intervalle de prédiction pour une valeur x^* donnée :</u>

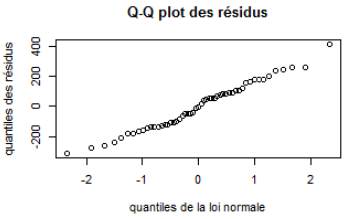
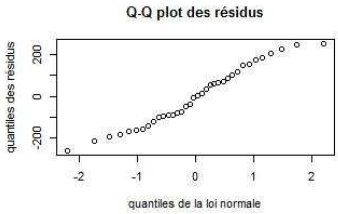
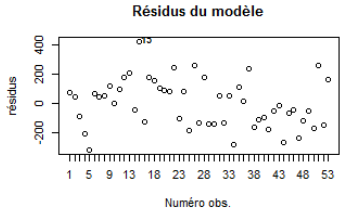
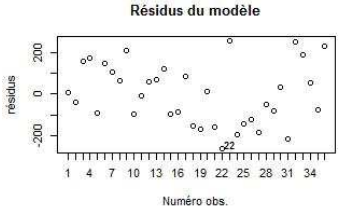
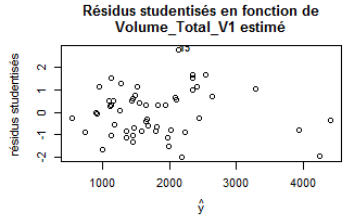
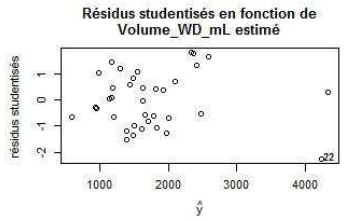
$$IP \pm t_{n-2,1-(\frac{\alpha}{2})} \times \left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}\right)^{0.5} \times \left[1 + \frac{1}{n} + \frac{\left((x^* - \bar{x})^2\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^{0.5}$$

- *La technique de validation croisée (cross-validation)*

Cette technique vise à partager des données généralement en 10 sous-ensembles disjoints de même taille et à calculer l'erreur de prédiction moyenne sur chacun des sous-ensembles, en utilisant les 9 autres sous-ensembles pour l'apprentissage.

Conditions à vérifier	Tests associés et analyses graphiques associées	**Diamètre** Régression linéaire simple (Validation 1)	**Diamètre** Régression linéaire simple (Validation 2)
Normalité des résidus	Histogramme des résidus, Normal Quantile-Quantile plot, Test de Shapiro-Wilk, Test de Kolmogorov-Smirnov	 Test de Shapiro-Wilk p-valeur : 0.607 Test de Kolmogorov-Smirnov p-valeur : 0.8	 Test de Shapiro-Wilk p-valeur : 0.00463 Test de Kolmogorov-Smirnov p-valeur : 0.271
Homoscédasticité	Test de Levene		
Indépendance des résidus	Graphe des résidus en fonction des observations, Test de Durbin-Watson	 Test de Durbin-Watson p-valeur : 0.659	 Test de Durbin-Watson p-valeur : 0.504
Indépendance des observations		Pas de choux-fleurs identiques mesurés plusieurs fois physiquement ou informatiquement	Pas de choux-fleurs identiques mesurés plusieurs fois physiquement ou informatiquement
Valeurs extrêmes	Graphe des résidus normalisés en fonction des valeurs prédites		
Indépendance des variables explicatives	Test de corrélation		

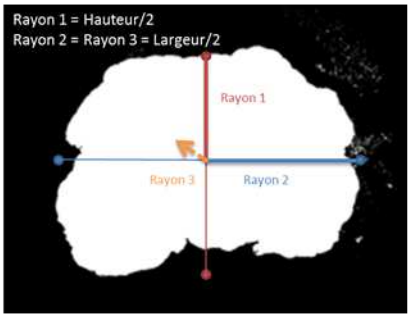
Annexe VII : Vérification des conditions d'applications des modèles statistiques pour les différents caractères traités dans la partie Résultats.

Conditions à vérifier	Tests associés et analyses graphiques associées	**Volume** Régression linéaire simple (Validation 1)	**Volume** Régression linéaire simple (Validation 2)
Normalité des résidus	Histogramme des résidus, Normal Quantile-Quantile plot, Test de Shapiro-Wilk, Test de Kolmogorov-Smirnov	 Test de Shapiro-Wilk p-valeur : 0.568 Test de Kolmogorov-Smirnov p-valeur : 0.796	 Test de Shapiro-Wilk p-valeur : 0.271 Test de Kolmogorov-Smirnov p-valeur : 0.711
Homoscédasticité	Test de Levene		
Indépendance des résidus	Graphe des résidus en fonction des observations, Test de Durbin-Watson	 Test de Durbin-Watson p-valeur : 0.561	 Test de Durbin-Watson p-valeur : 0.823
Indépendance des observations		Pas de choux-fleurs identiques mesurés plusieurs fois physiquement ou informatiquement	Pas de choux-fleurs identiques mesurés plusieurs fois physiquement ou informatiquement
Valeurs extrêmes	Graphe des résidus normalisés en fonction des valeurs prédites		
Indépendance des variables explicatives	Test de corrélation		

Annexe VIII : Formules de volume testées

Volume 1

Formule	4/3*pi*(Largeur/2)^2*(Hauteur/2)
Rayon 2, Rayon 3	Largeur (Feret), face interne
Rayon 1	Hauteur (MinFeret), face interne

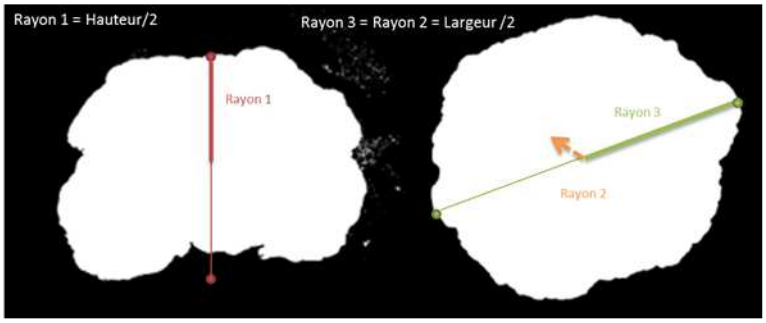


Volume 2

Formule	4/3*pi*(Largeur/2)^2*(Hauteur/2)
Rayon 2, Rayon 3	Largeur (Feret), face externe
Rayon 1	Hauteur (MinFeret), face externe

Volume 3

Formule	4/3*pi*(Largeur/2)^2*(Hauteur/2)
Rayon 3, Rayon 2	Largeur (Feret), vue du dessus
Rayon 1	Hauteur (MinFeret), face interne

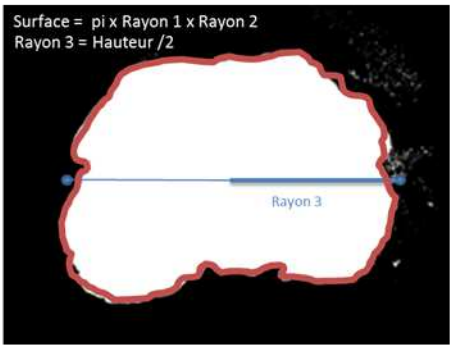


Volume 4

Formule	4/3*pi*(Largeur/2)^2*(Hauteur/2)
Rayon 3, Rayon 2	Largeur (Feret), vue du dessus
Rayon 1	Hauteur (MinFeret), face externe

Volume 5

Formule	4/3*Surface*(Largeur/2)
Surface (Rayon 1, Rayon 2)	Surface (Segmentation par seuillage de l'histogramme), face interne
Rayon 3	Largeur (Feret), face interne



Volume 6

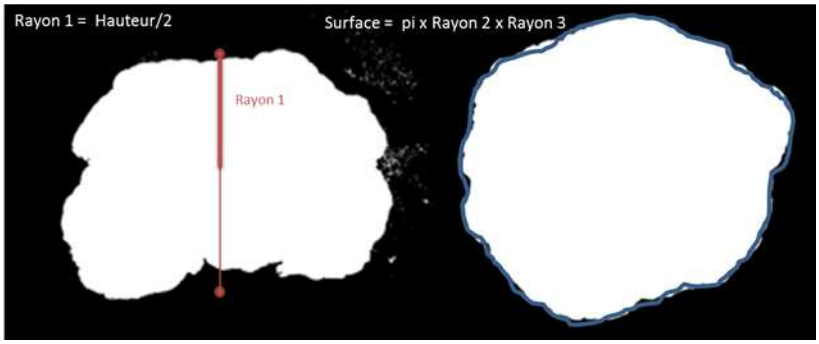
Formule	4/3*Surface*(Largeur/2)
Surface (Rayon 1, Rayon 2)	Surface (Segmentation par seuillage de l'histogramme), face externe
Rayon 3	Largeur (Feret), face externe

Volume 7

Formule	4/3*Surface*(Hauteur/2)
Surface (Rayon 2, Rayon 3)	Surface (Segmentation par seuillage de l'histogramme), vue du dessus
Rayon 1	Hauteur (MinFeret), face interne

Volume 8

Formule	4/3*Surface*(Hauteur/2)
Surface (Rayon 2, Rayon 3)	Surface (Segmentation par seuillage de l'histogramme), vue du dessus
Rayon 1	Hauteur (MinFeret), face externe



Annexe IX : Matrice de corrélation des variables testées pour l'évaluation du volume

Variables	AspectRatio_LC_externe	AspectRatio_LC_interne	Circularity_LC_externe	Circularity_LC_interne	Circularity_T	Diametre_Feret_T	Hauteur_MinFeret_LC_externe	Hauteur_MinFeret_LC_interne	Largeur.maximale_Feret_LC_interne	LargeurMaximale_Feret_LC_externe	Round_LC_externe	Round_LC_interne	Round_T
AspectRatio_LC_externe	1,00	0,65	0,55	0,32	0,30	0,12	0,50	0,13	0,20	0,13	0,98	0,62	0,18
AspectRatio_LC_interne	0,65	1,00	0,41	0,62	0,19	0,32	0,17	0,19	0,26	0,27	0,61	0,99	0,29
Circularity_LC_externe	0,55	0,41	1,00	0,50	0,14	0,34	0,02	0,28	0,39	0,38	0,53	0,40	0,12
Circularity_LC_interne	0,32	0,62	0,50	1,00	0,10	0,49	0,29	0,22	0,40	0,56	0,32	0,64	0,00
Circularity_T	0,30	0,19	0,14	0,10	1,00	0,46	0,06	0,21	0,51	0,29	0,33	0,20	0,13
Diametre_Feret_T	0,12	0,32	0,34	0,49	0,46	1,00	0,62	0,64	0,75	0,81	0,18	0,35	0,02
Hauteur_MinFeret_LC_externe	0,50	0,17	0,02	0,29	0,06	0,62	1,00	0,80	0,59	0,78	0,45	0,12	0,10
Hauteur_MinFeret_LC_interne	0,13	0,19	0,28	0,22	0,21	0,64	0,80	1,00	0,80	0,83	0,04	0,17	0,15
Largeur.maximale_Feret_LC_interne	0,20	0,26	0,39	0,40	0,51	0,75	0,59	0,80	1,00	0,84	0,27	0,29	0,02
LargeurMaximale_Feret_LC_externe	0,13	0,27	0,38	0,56	0,29	0,81	0,78	0,83	0,84	1,00	0,20	0,31	0,00
Round_LC_externe	0,98	0,61	0,53	0,32	0,33	0,18	0,45	0,04	0,27	0,20	1,00	0,59	0,13
Round_LC_interne	0,62	0,99	0,40	0,64	0,20	0,35	0,12	0,17	0,29	0,31	0,59	1,00	0,28
Round_T	0,18	0,29	0,12	0,00	0,13	0,02	0,10	0,15	0,02	0,00	0,13	0,28	1,00
Surface_LC_externe	0,19	0,07	0,16	0,45	0,21	0,75	0,93	0,84	0,76	0,94	0,12	0,11	0,05
Surface_LC_interne	0,05	0,02	0,27	0,40	0,26	0,77	0,84	0,93	0,81	0,94	0,04	0,05	0,10
Surface_T	0,01	0,18	0,29	0,53	0,30	0,80	0,84	0,85	0,82	0,97	0,07	0,22	0,05
Volume1_mL	0,10	0,10	0,38	0,38	0,41	0,68	0,67	0,90	0,95	0,86	0,18	0,13	0,07
Volume2_mL	0,05	0,15	0,30	0,55	0,25	0,76	0,86	0,83	0,77	0,96	0,02	0,19	0,09
Volume3_mL	0,04	0,13	0,35	0,47	0,43	0,94	0,74	0,82	0,81	0,89	0,11	0,16	0,09
Volume4_mL	0,07	0,16	0,26	0,51	0,39	0,94	0,80	0,73	0,73	0,87	0,01	0,20	0,07
Volume5_mL	0,05	0,10	0,33	0,44	0,37	0,77	0,78	0,92	0,90	0,94	0,13	0,14	0,08
Volume6_mL	0,06	0,15	0,25	0,53	0,25	0,76	0,87	0,83	0,78	0,96	0,01	0,19	0,07
Volume7_mL	0,02	0,06	0,32	0,46	0,28	0,74	0,83	0,92	0,83	0,94	0,06	0,10	0,12
Volume8_mL	0,13	0,09	0,24	0,51	0,25	0,76	0,90	0,84	0,76	0,94	0,06	0,14	0,10
Volume_Total_V1	0,15	0,03	0,20	0,44	0,23	0,74	0,89	0,86	0,75	0,92	0,08	0,07	0,15

Code couleur	0,0-0,5
	0,5-0,7
	0,7-1

Variables	Surface_LC_externe	Surface_LC_interne	Surface_T	Volume1_mL	Volume2_mL	Volume3_mL	Volume4_mL	Volume5_mL	Volume6_mL	Volume7_mL	Volume8_mL	Volume_Total_V1
AspectRatio_LC_externe	0,19	0,05	0,01	0,10	0,05	0,04	0,07	0,05	0,06	0,02	0,13	0,15
AspectRatio_LC_interne	0,07	0,02	0,18	0,10	0,15	0,13	0,16	0,10	0,15	0,06	0,09	0,03
Circularity_LC_externe	0,16	0,27	0,29	0,38	0,30	0,35	0,26	0,33	0,25	0,32	0,24	0,20
Circularity_LC_interne	0,45	0,40	0,53	0,38	0,55	0,47	0,51	0,44	0,53	0,46	0,51	0,44
Circularity_T	0,21	0,26	0,30	0,41	0,25	0,43	0,39	0,37	0,25	0,28	0,25	0,23
Diametre_Feret_T	0,75	0,77	0,80	0,68	0,76	0,94	0,94	0,77	0,76	0,74	0,76	0,74
Hauteur_MinFeret_LC_externe	0,93	0,84	0,84	0,67	0,86	0,74	0,80	0,78	0,87	0,83	0,90	0,89
Hauteur_MinFeret_LC_interne	0,84	0,93	0,85	0,90	0,83	0,82	0,73	0,92	0,83	0,92	0,84	0,86
Largeur.maximale_Feret_LC_interne	0,76	0,81	0,82	0,95	0,77	0,81	0,73	0,90	0,78	0,83	0,76	0,75
LargeurMaximale_Feret_LC_externe	0,94	0,94	0,97	0,86	0,96	0,89	0,87	0,94	0,96	0,94	0,94	0,92
Round_LC_externe	0,12	0,04	0,07	0,18	0,02	0,11	0,01	0,13	0,01	0,06	0,06	0,08
Round_LC_interne	0,11	0,05	0,22	0,13	0,19	0,16	0,20	0,14	0,19	0,10	0,14	0,07
Round_T	0,05	0,10	0,05	0,07	0,09	0,09	0,07	0,08	0,07	0,12	0,10	0,15
Surface_LC_externe	1,00	0,94	0,96	0,82	0,97	0,86	0,89	0,92	0,98	0,94	0,98	0,97
Surface_LC_interne	0,94	1,00	0,96	0,88	0,94	0,90	0,86	0,97	0,94	0,97	0,94	0,96
Surface_T	0,96	0,96	1,00	0,86	0,97	0,90	0,89	0,96	0,98	0,97	0,98	0,97
Volume1_mL	0,82	0,88	0,86	1,00	0,84	0,84	0,74	0,96	0,84	0,92	0,84	0,84
Volume2_mL	0,97	0,94	0,97	0,84	1,00	0,89	0,90	0,93	0,99	0,97	0,99	0,97
Volume3_mL	0,86	0,90	0,90	0,84	0,89	1,00	0,97	0,91	0,88	0,90	0,88	0,88
Volume4_mL	0,89	0,86	0,89	0,74	0,90	0,97	1,00	0,85	0,90	0,86	0,90	0,88
Volume5_mL	0,92	0,97	0,96	0,96	0,93	0,91	0,85	1,00	0,94	0,98	0,93	0,94
Volume6_mL	0,98	0,94	0,98	0,84	0,99	0,88	0,90	0,94	1,00	0,96	0,99	0,97
Volume7_mL	0,94	0,97	0,97	0,92	0,97	0,90	0,86	0,98	0,96	1,00	0,97	0,97
Volume8_mL	0,98	0,94	0,98	0,84	0,99	0,88	0,90	0,93	0,99	0,97	1,00	0,98
Volume_Total_V1	0,97	0,96	0,97	0,84	0,97	0,88	0,88	0,94	0,97	0,97	0,98	1,00

Code couleur	0,0-0,5
	0,5-0,7
	0,7-1